David Heyd · Uzi Segal

Democratically elected aristocracies

Received: 16 June 2004 / Accepted: 24 January 2005
© Springer-Verlag 2006

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 to have 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 formal and philosophical reasons that support the model.

1 Introduction

Voting is one among many procedures of making social choices, where the aim is the translation of individual preferences into social policies while 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.

D. Heyd (✉)
Department of Philosophy, The Hebrew University, Jerusalem 91905, Israel
E-mail: david.heyd@huji.ac.il

U. Segal
Department of Economics, Boston College, Chestnut Hill MA 02467, USA
E-mail: segalu@bc.edu
The principal attraction of voting 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. 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 \(A \succ_j^\alpha B\) means that person \(\alpha\) believes that the opinion of group \(A\) of individuals regarding issue \(j\) is more relevant than the opinion of group \(B\). Under some assumptions we conclude that this ranking can be represented by a function \(V_\alpha\) in the following way. Each member of society is assigned a certain weight, and \(V_\alpha(A)\) is obtained by taking the sum of these weights over all members of the set \(A\) (see Theorem 1 in Sect. 3). Although (assuming that all weights are non-negative) the most relevant 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 rankings. In Sect. 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. 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 Sect. 5 we analyze the case where the weights one person wishes to 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 Sect. 6 we discuss some possible objections to the model, and in Sect. 7 we offer some remarks on the literature. All proofs appear in the Appendix.

2 The two-tier voting model

To be able to obtain a social welfare function, Harsanyi (1955) 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 v\(\text{N&M}\) utilities.

\(^1\) 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.
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 a 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 production 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 of 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 \( "(a, \text{not } b, \text{not } c) \text{ with probability } p \text{ and (not } a, \text{not } b, c) \text{ with probability } 1 - p" \) and \( "(\text{not } a, \text{not } b, \text{not } c)." \) 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, 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 may enhance or reduce the relevance of their opinion in the social choice.\(^2\) 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 1990; Wolff 1994; Karni 1996; Karni and Safra 2002; Segal 2000).

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 different weights to particular categories of people on the basis of these people’s alleged privileged position, authority, moral standing, particular sensitivity to the outcome of the substantive decision, or in general, because they find the views of different individuals to bear different degrees of relevance to the issue.\(^3\) The procedure relates to issues about which there is not only first-order

\(^2\) 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 issue or the stronger impact any policy would have on their lives and interests.

\(^3\) 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. Also, we ignore here the issue of strategic behavior – see Sect. 6 below.
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

Typical issues to which our model applies are those involving two possible out-
comes (“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 1997; Samet and Schmeidler 1998). Our
model does not require that members of society assign different weights to differ-
ent individuals, and therefore standard referendum-voting procedures of “one man
one vote” are a special case of our model. Our approach is useful in situations like
abortion rights, where utilities are hard to be assigned. 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.5

Take, for instance, abortions. Some may find women’s view on this issue to be
more relevant than men’s, and would therefore wish to give women more weight
than men, because of the particular position of the pregnant woman with regard
to her own body, while some may find women’s views less relevant and conse-
quently may wish to give them less weight because being preoccupied with their
own interests, women may not give enough thought to the interests of the fetuses.
Others may give everybody an equal vote on that matter on the basis 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 as
their (learned) views are more relevant. 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 never-
theless have strong views on the question of who should decide what to do with
the money. In a democratic procedure, we claim, this second-order disagreement
should also be democratically settled.

4 By characterizing the suitable issues as “ideologically charged” we mean to exclude cases
which are hard to decide due to purely 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.

5 Frankfurt (1988) 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.
There is some tension between the two kinds of considerations in assigning weights to prospective voters on a particular policy: those based on the interests of the people who are going to be affected by the policy (both positively and negatively), and those based on the cognitive position of the voters making the decision on that policy (their knowledge, experience, skill or authority in matters pertaining to the issue at hand). Sometimes there is a measure of overlap between the two categories, i.e. when the stronger interests of some subjects make them better able to judge the desirability of the policy in question. But there are many more cases in which the two sets of reasons may stand in conflict, e.g. when strong interests cast a shadow on the cognitive ability to make a sound judgment, or when a privileged epistemic position suppresses the sensitivity to the price paid by the party which stands to lose from the policy. In any case, since the resolution of this tension is in itself value-laden, the relative merits of the two categories of considerations should be left to the voters in the primary, weight-assigning stage of the voting method proposed here.

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 endowed 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, since they are held as more knowledgeable and able to form judgment, or in general, because their views on the subject matter are deemed more relevant. 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, and then in the second, 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 fail to acknowledge that some individuals may have a stronger say on some matters, while others have more authority on other matters. Our model is sensitive to such differences, and permits society to evaluate them.

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 them. 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. One might object that the voting system suggested here might be seen as legitimizing segregational voting (e.g. the allocation of particular, even exclusive, weight to voters of a particular race or ethnic group). However, this danger is by no means unique to the two-tier method. In direct, one-tier voting systems the law can ban racist political parties, but it cannot stop individual citizens from casting their vote on a racist basis. The whole point of democratic voting is the separation of the issue of the faithfulness or fairness of representation of citizens’ opinions from their correctness or moral soundness. Society may put limits to the representation of unacceptable opinions and biases, but these are imposed as external constraints on the issues put to vote.

It should be emphasized that the voting system suggested here is an idealized model of democratic representation rather than a politically realistic suggestion for electoral reform. Therefore, the fundamental rationality of the voters and their basic conscientiousness are taken as background assumptions. It is indeed true that real-life systems of democratic voting must take into consideration the possibility of narrowly self-interested, insincere, manipulative or even straightforwardly immoral use of one’s power to allocate differential weights to voters (e.g. deciding arbitrarily to grant privileged power to people of a particular race just because of their color). In our idealized model, individuals are taken to be “disinterested,” which is in line with the idealized conditions of a Kantian Kingdom of Ends, Rawlsian Original Position, or the assumption of rationality of players in game-theoretic models.

---

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).
3 Individual rankings

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 axiomatic foundation for these weights. Given two groups $A$ and $B$ of individuals, we ask each member $\alpha$ of society the opinion of which of these two groups he finds to be more relevant to the issue. Under some assumptions discussed below, we show that person $\alpha$’s ranking of groups can be represented in the following way. A certain weight is attached to each (type) of voters, and $A$ is ranked higher than $B$ if its aggregate weight is higher.

We assume that society is composed of a continuum set $S$ of agents, say $S = [0, 1]$. Consider a question that fits our domain, for example, abortion rights. Each member $\alpha$ of society has complete and transitive ranking $\succeq_\alpha$ over measurable subsets of $S$, where $A \succeq_\alpha B$ means “person $\alpha$ finds the opinion of group $A$ of individuals to be more relevant to the issue than the opinion of group $B$.”

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 $S = \{S_1, \ldots, S_N\}$ and assume that each $S_j$ 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 $S$.

To make use of our structure, we need to rule out the possibility that for some $\alpha \in [0, 1]$, the partition will be $S_1 = \{\alpha\}$ and $S_2 = \text{everyone else}$. Of course, individuals may have the feeling they should be appointed dictators, and therefore their ranking of 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. Moreover, we do not require that everyone in society agrees that all elements of the partition are relevant to his own ranking, as long as everyone is willing to agree that it is legitimate to differentiate between individuals who belong to different elements of the partition. For example, person $\alpha$ may feel that the views of the clergy on abortion rights are as relevant as those of doctors, but may still accept the fact that other people want to treat these views differently. (For a further discussion, see “Preferences versus values” in Sect. 6 below).

Next we analyze the preferences of a given individual $\alpha$. To simplify notation, we will omit, for the present discussion, the subscript $\alpha$.

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

---

7 For example, on the issue of abortion rights, one person may find the opinion of a group of one hundred women and fifty men to be more relevant than that of a group of fifty women and two hundred men, while someone else may have an opposite view. Both may agree that the opinion of women is more relevant to the issue, but the first person pays more attention to it than the second person.
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 $\alpha$ finds the opinion of members of subset $A$ to be more relevant to the issue than that of subset $B$, and if $A'$ and $B'$ are $\lambda$-replicas of $A$ and $B$ with respect to the partition $S$ of $[0,1]$, then he should also find the opinion of members of $A'$ to be more relevant than that of members of $B'$. 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 $S$ that counts, and not the identity of its members.

To say that the opinion of group $A$ is more relevant to the issue than that of group $B$ is the same as saying that the opinion of individuals who are not in $A$ is less relevant to the issue than that of individuals who are not in $B$. We can therefore define a relation $\succeq'$ of “less relevant than” over subsets of $S$ by $C \succeq' D$ iff $A \succeq B$ where $C = S \setminus A$ and $D = S \setminus B$.

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

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

In other words, for all sets $A$, $B$, and $\lambda > 0$, $A \succeq B$ iff $\lambda A \succeq \lambda B$. For a further discussion of this axiom, and possible alternatives, see Sect. 6 below.

The next axiom\(^8\) compares two sets $A$ and $B$ that have the same number of individuals of type $i^*$. We suggest that the ranking of 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.

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 these three axioms, 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 NY State,” and “residents of other states,” and suppose indifference between the sets $A = (500, 100, 200)$ and $B = (500, 150, 100)$. The scaling axiom implies indifference between (250, 50, 100) and (250, 75, 50). The argument is that the relative representation of the three groups in $A'$ and $B'$ is the same as it was in $A$ and $B$, therefore their relative rankings should be the same. Likewise, if the overall sizes of $S_1$, $S_2$, and $S_3$ are (1,000, 500, 500), then residual scaling suggests that $A'' = (750, 300, 350) \sim B'' = (750, 325, 300)$. The rationale here is that $A \sim B$ implies that society is indifferent between excluding (500, 400, 300) and (500, 350, 400), and should therefore also be indifferent between excluding (250, 200, 150) and (250, 175, 200).

\(^8\) For a discussion of the complete separability axiom see, among others, Blackorby et al. (1978) and Wakker (1989).
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 below 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 below needs non-triviality, but there are situations where the stronger monotonicity axioms should be imposed).

**Continuity** If for all \( i, \mu(A^{k_i}_i) \rightarrow \mu(A_i), \mu(B^{k_i}_i) \rightarrow \mu(B_i), \) and for all \( k, A^k \succeq B^k, \) then \( A \succeq B. \)

**Strict monotonicity** \( A \subset B \) implies \( B \succ A. \)

**Monotonicity** \( S \succ \emptyset \) and \( A \subset B \) implies \( B \succeq A. \)

**Non-triviality** \( S \not\sim \emptyset. \)

**Theorem 1** Assume \( N \geq 3. \) The following two conditions are equivalent.

1. The ranking \( \succeq \) over the subsets of \( S \) satisfies 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 \( \beta, \) 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).

There are of course many axiomatizations of linear functions in the literature, the most famous of them being the expected utility theorem. And of course, homotheticity (called “scaling” above) and complete separability are not new, but these two together don’t imply linearity (see Footnote 13 below). Theorem 1 differs from standard axiomatizations of linear functions in utilizing the residual scaling assumption. As is evident from the proof of the theorem, this assumption creates another point with respect to which rankings are homothetic. Other models did not use this assumption because usually there is only one natural “zero” point with respect to which rankings are homothetic.

Suppose that on the issue of abortion rights society recognizes three groups: clergy, lay men, and lay women, and that we find out that person \( \alpha \) assigns these groups the weights \((1, 2/3, 4/3)\), respectively. Given a set \( A \), he is indifferent, in his “relevance” relation, 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, the opinion of one lay woman is as relevant as that of two lay men. A possible way for him to express this view is by giving women twice the voting power of men. In other words, we can imagine person \( \alpha \) assigning voting coupons to members of society, where person \( x \) is assigned by him \( f^\alpha(x) \) coupons. (Here \( k(x) = 1, 2/3, 4/3 \) for \( x = \text{clergy, lay men, and lay women} \). If people now vote on the issue itself (that is,
whether or not to have abortion rights) while using these coupons, then effectively the opinion of each lay woman will count twice as that of every lay man.\footnote{Note that coupons here represent voting power rather than the means of acquiring resources as is the case in Dworkin’s Dworkin (2000, 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).}

\section{4 Aggregation}

The analysis of the previous section yields the conclusion that each member $\alpha$ of society would like to assign the voting weights $k^{\alpha} = (k^{\alpha}_1, \ldots, k^{\alpha}_N)$ to society’s $N$ subgroups. Given these individual rankings, society too, we suggest, should assign voting 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 rankings (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.\footnote{We may in fact use a weaker restriction, namely that all individual weights are uniformly bounded (with respect to $\alpha$).}

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 $\alpha$’s weighing functions $k^{\alpha} = 1, 2/3, 4/3$ 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 extra voting power. Formally, we give each person a set of coupons in the size of society (that is, 1), and impose the restriction

$$\sum k^{\alpha}_i \mu(S_i) = 1 \quad (1)$$

(For some recent discussion of possible normalizations in Harsanyi’s model, see Dhillon 1999; Karni 1998; Segal 2000).

Denote by $\mathcal{K}$ the set $\{ k \in \mathbb{R}_+^N : \sum_{i=1}^N k_i \mu(S_i) = 1 \}$. For every $\alpha \in [0, 1]$, person $\alpha$’s ranking leads to an element of $\mathcal{K}$. Denote this function $f = (f_1, \ldots, f_N) : [0, 1] \to \mathcal{K}$, where $f(\alpha) = k^{\alpha} = (k^{\alpha}_1, \ldots, k^{\alpha}_N)$ and $f_1(\alpha) = k^{\alpha}_1$. 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 $f \in F$, 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 $\varphi = (\varphi_1, \ldots, \varphi_N) : F \to K$. 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 $K$, 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 that the clergy are 10% of the population, 45% of the population 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 clergy or 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 Sect. 3).

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 $\lambda$, 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 $N \geq 3$. If the social aggregation rule satisfies the anonymity and the unanimity axioms, then the social coefficients $k_1^s, \ldots, k_N^s$ 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 $\varphi_i$ 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 definition 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 definition of sexual harassment, some may feel that women should be given more voice on both issues. Suppose person $\alpha$ 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 $\alpha$ will be willing to compensate women by offering them more weight on the other issue. We do not suggest that person $\alpha$ 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 ranks subsets of $S$ for issue $m$, $m = 1, \ldots, M$. These rankings satisfy the axioms of Sect. 3, but they may now depend on the weights each of the $N$ categories receive on other issues. Thus, for issue $m$, person $\alpha$ has the ranking $\preceq^m_{\alpha}(k^s_m)$, where $k^s_m$ are the social weights to all groups in all other issues. Since, by Theorem 1, these rankings are representable by the linear weights $k^S_\alpha(m, k^s_m)$, we express the following continuity axiom in terms of these weights, but the translation into continuity of the rankings themselves in $k^s_m$ (via measurable subsets of $2^S \times 2^S$) is simple.

**Continuity** The rankings $\succeq^m_{\alpha}(k^s_{m-1}, k^s_{m+1})$ person $\alpha$ has over decisive sets for issue $m$ are uniformly continuous in $k^s_{m-1}$ and in $\alpha$. That is, $\forall \delta \exists \varepsilon$ such that $\| k^s_{m-1} - k^s_{m} \| < \varepsilon$ implies, for all $\alpha$, $\| k^S_\alpha(m, k^s_{m-1}) - k^S_\alpha(m, k^s_{m+1}) \| < \delta$.

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 $\mathcal{K}^m = \{ k^m \in \mathfrak{M}_+ : \sum_{i=1}^N k^m_i \mu(S_i) = 1 \}$, $m = 1, \ldots, M$. There are $M$ continuous functions, $g^m : \prod_{\ell \neq m} \mathcal{K}^\ell \rightarrow \mathcal{K}^m$, such that given the social weights $k^s, k^x, l$ for issue $\ell$, $\ell = 1, \ldots, M$, $\ell \neq m$, 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 $g : \prod_m \mathcal{K}^m \rightarrow \prod_m \mathcal{K}^m$ by
Democratically elected aristocracies

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

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 \( \bar{k}^{s,m} \).

### 6 Q & A

**Why should people report true preferences?** Strategic behavior cannot be denied in many social interactions, including voting. Our structure is of course vulnerable to such manipulations, for example, a person may pretend that in his view his own group should receive special attention. But our aim is not to design a strategy-proof mechanism. Rather, like many other models of social choice, we want to find out what society should do assuming true utilities are known. In other words, assuming that we know how people sincerely feel about different issues and what they think about the relevance of other people’s views on these issues, our model offers a way to aggregate these different opinions.

**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 \( 1/4 \) 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 \( 3/4 \) of \( S_2 \) (that is, \( 60\% \) of
D. Heyd, U. Segal

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.

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 (1955) 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 Sect. 2). 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.\(^\text{11}\)  

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.

**IS THE RESIDUAL SCALING AXIOM REASONABLE?** Consider the following alternative approach, which is mathematically equivalent to the residual scaling axiom. Suppose, someone finds the opinion of group \(A\) to be more relevant to the issue of abortion rights than that of \(B\). For example, \(A\) is the group of all doctors, and \(B\) is the group of all the clergy in society. Let \(C = S \setminus A\) be the set of all non-doctors and \(D = S \setminus B\) be the set of all the non-clergy. Then it is reasonable to assume that this person also ranks the relevance of \(D\) higher than \(C\). Formally:

**Complement reversal** For all \(A\) and \(B\), \(A \succ B\) iff \(S \setminus B \succ S \setminus A\).

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

The difference between the two approaches is that in Sect. 3, the ranking \(\succ\)' does not assume anything new – it is just a restatement of the original ranking \(\succeq\).\(^\text{12}\) Residual scaling makes an assumption about this ranking. We claim that if one is willing to accept the rationale for 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

\(\text{11}\) 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.

\(\text{12}\) 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.”
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 the possible intuition that the opinion of balanced sets of types is more relevant than that of unbalanced sets. 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 orderings.\(^{13}\) It is the combination of the three that implies indifference to balanced sets, and not any of the axioms by itself.

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. Mill (1861, 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 1998), 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 (1990), 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

\[ V(A) = \prod \mu(A_i) \] 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 \( \pi \) be a permutation of \( \{1, \ldots, N\} \) such that \( \mu(A_{\pi(1)})/\mu(S_{\pi(1)}) \leq \ldots \leq \mu(A_{\pi(N)})/\mu(S_{\pi(N)}) \), and define \( V(A) = \sum \mu(A_{\pi(i)})/\mu(S_{\pi(i)}) \left[ \sqrt{i/N} - \sqrt{i-1/N} \right] \). This function satisfies scaling and residual scaling, but not complete separability. All three functions are quasi concave, hence indicate preferences for balanced sets.
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 (1994) “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 (1997) and Samet and Schmeidler (1998).

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. In each of these lemmas we assume that preferences satisfy the axioms of scaling, residual scaling, complete separability, continuity, and non-triviality.

The ranking $\succeq$ induces a ranking over $X = \prod_{i=1}^{N} [0, \mu(S_i)]$, where $(a_1, \ldots, a_N)$ is ranked higher than $(b_1, \ldots, b_N)$ 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 $\mu(B_i) = b_i$. With a little abuse of notation, we will also denote this induced ranking by $\succeq$. 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 ranking $\succeq$ is strictly (positively or negatively) monotonic along $L$. That is, either for all $b \in L$, $b \neq 0$, and $\lambda \in [0, 1)$, $b \succ \lambda b$, or for all such $b$ and $\lambda$, $\lambda b \succ b$.

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 $p > 0$ (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 plane containing $L$. Then $H \cap X$ is a parallelogram in $\mathbb{R}^N$.

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

**Proof** Since $\succeq$ is strictly monotonic along $L$, it follows by continuity that there are $a \in L$ and $b \in H \setminus L$ such that $a \sim b$. Let $c \in [a, b]$. The points 0, $p$, $a$, $b$, and $c$ are all in $H$. Following Lemma 2, $H \cap X$ is depicted in Fig. 1 by the parallelogram $0 \text{gph}$. 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, w.l.g., that there is in $H$ a sequence $b_n \to b$ such that for all $n$, $b \in [a, b_n]$, and $b_n \sim a$, 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).

\[14\] Otherwise, pick $a \in (0, p)$, and assume w.l.g that for all $b \in H \setminus L, b \succ a$. Pick $a' \in (0, a)$ and obtain by continuity that $a' \succeq a$, a contradiction to monotonicity.
Fig. 1. Clearly, \(a_{n+1} \parallel ab\). By the scaling axiom it follows that there is a point \(d_{n+1} \in L\), strictly between \(a\) and \(a_{n+1}\), such that \(b \sim d_{n+1}\), a contradiction to Lemma 1.

The scaling axiom implies that if for \(a \in L\) and some \(b \notin L\), the chord \([a, b]\) is part of an indifference set, then for all \(\lambda < 1\), the chord \([\lambda a, \lambda b]\) is part of (another) indifference set. By the above argument, all indifference curves below \(a\) in the triangle \(\Delta 0gp\) are parallel to \([a, b]\). Similarly, by the residual scaling axiom, all indifference curves above \(a\) are parallel to \([a, b]\). It thus follows that on the triangle \(\Delta 0gp\), \(a\) can be represented by a linear function.

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

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 \(a, a^* \in X\) and \(I \subset \{1, \ldots, n\}\), define \(a(I, a^*) = (a_i(I, a^*))_{i=1}^N\) 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, \(p = (\mu(S_1), \ldots, \mu(S_N))\)).

\[(I, a^*)\text{–Scaling}\] Let \(a, b \in X(I, a^*)\). For all \(\lambda > 0\), \(a \succeq b\) iff \(\lambda a(I, a^*) + (1 - \lambda)0(I, a^*) \geq \lambda b(I, a^*) + (1 - \lambda)0(I, a^*)\).

\[(I, a^*)\text{–Residual Scaling}\] Let \(a, b \in X(I, a^*)\). For all \(\lambda > 0\), \(a \succeq b\) iff \(\lambda a(I, a^*) + (1 - \lambda)p(I, a^*) \geq \lambda b(I, a^*) + (1 - \lambda)p(I, a^*)\).

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 from scaling, residual scaling, and complete separability.\(^{15}\)

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

**Proof** Let \(a, b \in X(I, a^*)\). Then \(\lambda a, \lambda b \in X(I, \lambda a^*)\). Suppose \(I = \{i_0 + 1, \ldots, N\}\). If \(i_0 = 0\), then \((I, a^*)\text{–scaling}\) and \((I, a^*)\text{–residual scaling}\) trivially hold, so assume \(i_0 \geq 1\). Define \(a^i_0 = \lambda a\), and for \(i = i_0 + 1, \ldots, N\), 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 \(a^*_i\). Define \(b^{i_0}, \ldots, b^N\) in a similar way. By scaling, \(a \succeq b\) iff \(a^{i_0} \succeq b^{i_0}\), and by complete separability, for \(i = i_0 + 1, \ldots, N\), \(a^i \succeq b^i\) iff \(a^{i-1} \succeq b^{i-1}\). But \(a^N = \lambda a(I, a^*) + (1 - \lambda)0(I, a^*)\), hence \((I, a^*)\text{–scaling}\).

The proof of \((I, a^*)\text{–residual scaling}\) is similar. \(\square\)

**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 \(i = i_0 + 2, \ldots, N\),

\(^{15}\) The geometric difference between complete separability and the last two axioms is clear. The former imposes no restrictions on the ranking \(\succeq\) when one coordinate is fixed, but connects together such rankings 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.
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 \( a^* \). By scaling, \( a \geq b \) iff \( a_{i0}^{i+1} \geq b_{i0}^{i+1} \), and by complete separability, for \( i = i_0 + 2, \ldots, N \), \( a^i \geq b^i \) iff \( a^{i-1} \geq b^{i-1} \). 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. \( \square \)

Similarly to the above analysis, it follows that for every \((I, a^*)\) and for every \( b \in X(I, a^*) \), on the plane \( H \) through \( b \) and \( L(I, a^*) \) (the line through \( 0(I, a^*) \), and \( p(I, a^*) \)), the ranking \( \geq \) can be represented by a function that is linear on each of the two sides of \( L(I, a^*) \) in \( H \).

The ranking \( \geq \) over the product set \( X \) is 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 1960; Gorman 1995). Consider now the set \( X \{3, \ldots, N\}, a^* \), where all but the first two variables are fixed. On this set, the ranking can be represented by \( v_1(a_1) + v_2(a_2) \), but also by

\[
W(a_1, a_2) = \begin{cases} 
    k_1a_1 + k_2a_2 & a_2 < \frac{p_2}{p_1}a_1 \\
    k'_1a_1 + k'_2a_2 & a_2 \geq \frac{p_2}{p_1}a_1,
\end{cases} \tag{3}
\]

where \( k_1p_1 + k_2p_2 = k'_1p_1 + k'_2p_2 \).

**Lemma 5** \( v_1 \) is linear.

**Proof** Consider the range \( a_2 < (p_2/p_1)a_1 \). 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

\[
v'_1(a_1^0) = k_1h'(k_1a_1^0 + k_2a_2^0). \tag{4}
\]

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 \( k_1a_1 + k_2a_2 = k_1a_1^0 + k_2a_2^0 \). (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 \( v_2 \). \( \square \)

By similar arguments all the functions \( v_i \) are linear, hence the theorem. Proving that strict monotonicity implies positive coefficients is trivial. \( \square \)
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 \( I \not\subset \not\in \{1, \ldots, N\} \) and \( i^* \not\in I \), and is given by \( \{a \in X : a_i = 0 \text{ for } i \in I, a_i = 1 \text{ for } i \not\in I \text{ and } a_{i^*} \in [0, 1]\} \). Pick a plane \( H \) such that \( I \subset H \), let \( a^* \not\in H \), \( p \) 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 \( \gamma \) and \( \delta \) 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 \( I' = [1, \ldots, N] \setminus \{i'\} \) 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 \( j \not\equiv i' \) 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 \( j \not\equiv i' \), \( a_j^* = 1 \). 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 \( j \not\equiv i' \) such that \( a_j^* = 0 \), then \( b_j = \gamma \), hence \( \gamma = 1 \), \( \delta = 0 \), and \( b = p \). Otherwise, for all \( j \not\equiv i' \), \( a_j^* = b_j = 1 \). 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 \( j \not\equiv i' \) 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 \( j \not\equiv i' \), \( b_j = 1 \), then either \( \exists j \not\equiv i' \) such that \( a_j^* = 1 \), hence \( \delta = 0 \) and \( b = p \), or for all \( j \not\equiv i' \), \( a_j^* = 0 \). 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). \( \square \)

Proof of Theorem 2 Unanimity applies to all \( \lambda \), and in particular to \( \lambda = 1 \). Let \( f \) and \( g \) agree on the \( i \)-th group, that is, for all \( \alpha, f_i(\alpha) = g_i(\alpha) \). By unanimity, \( \varphi_i(f) = \varphi_i(g) \). 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 \( \varphi_i \) depends only on \( f_i \), hence \( \varphi(f) = (\varphi_1(f), \ldots, \varphi_N(f)) \).

To simplify notation, we assume wlg that \( \mu(S_1) = \cdots = \mu(S_N) = 1/N \).\(^{16}\) Note that for every \( \alpha \), \( \sum f_i(\alpha) = N \), and for every \( f = (f_1, \ldots, f_N) \)

\[
\sum \varphi_i(f) = \sum \varphi_i(f_i) = N \quad \text{(5)}
\]

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. (5), \( \sum \varphi_i(f_i) = N \). Therefore, if \( f_i \equiv N \) and for all \( j \not\equiv i \), \( f_j \equiv 0 \), then \( \varphi_i(f_i) = N \). Unanimity now implies the lemma. \( \square \)

\(^{16}\) Alternatively, we can define \( f_i(\alpha) = N \mu(S_i) f_i(\alpha) \) and work with these functions instead of the functions \( f_i \).
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-allocation to $S_i$.

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

$$\varphi_i(\xi f_i^1 + (1-\xi) f_i^2) = \xi \varphi_i(f_i^1) + (1-\xi) \varphi_i(f_i^2).$$  

**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 $\xi > 0$, $h_1^1 = (1+\xi)(1+\xi)f_1^1 - \xi f_1^1$ and $h_1^2 = (1+\xi)(1-\xi)f_1^1 - \xi f_1^2$ are bounded away from 0 and $N$ and satisfy $h_1^1, h_1^2 \in \mathcal{F}_1$. Let $H = \text{Conv}\{h_1^0, h_1^1, h_1^2\}$ be the minimal convex set containing $h_1^0, h_1^1, h_1^2$, 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 $\inf g_1^i(\alpha), \inf (N - g_1^i(\alpha)) > \delta$, $j = 1, 2$. Obviously, for all $\alpha$, $0 < g_1^j(\alpha)N/(N-\delta) < N$. Define

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

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 $\delta$, so are $N - g_1^1$ and $N - g_1^2$. If $g_1^1$ and $g_1^2$ are sufficiently close to each other, (the exact requirement is that for every $\alpha$,

$$g_1^2(\alpha)(1-\frac{\delta}{N}) < g_1^1(\alpha) < g_1^2(\alpha)/(1-\frac{\delta}{N}),$$

then for all $\alpha$

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

and likewise, $N - g_3^j(\alpha) > g_1^j(\alpha)$. By Eq. (7), for all $\alpha$, $N - g_3^j(\alpha) > g_1^j(\alpha)$. 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.

Figure 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 $\alpha_1$ and $\alpha_2$ 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).

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, \( \varphi_1 \) is linear on \( H^j \) on both sides of the chords \( [0, \frac{g_j^N}{N-\delta}] \), \( j = 1, 2 \).

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 \( \varphi_1 \) 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 1995) that \( \varphi_1 \) is linear on the interior of \( H \), in particular, eq. (6) 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. (5), for $j = 1, 2$,

$$\varphi_1(f^1_j) = N - \frac{\delta}{2} - \varphi_2(f^2_j).$$

(8)

By Case 1, $\varphi_2$ is linear on the chord connecting $f^1_2$ and $f^2_2$, hence, by Eqs. (8), (6) of the lemma is satisfied for $f^1_1$ and $f^2_1$.

We similarly handle the case where the functions $f^1_1$ and $f^2_1$ are bounded away from zero, but not necessarily from $N$.

**Case 3**

The functions $f^1_1$ and $f^2_1$ are not necessarily bounded away from $N$, nor from zero. Define

- $f^1_2 = f^3_1 = \frac{1}{2}(N - f^1_1), \ j = 1, 2$;
- $f^1_i = 0, \ j = 1, 2, \ i = 4, \ldots, N$.

The functions $f^1_2$ and $f^3_1$ are bounded away from $N$, hence, by Case 2, $\varphi_i$ is linear on the chord connecting $f^1_i$ and $f^2_i$, $i = 2, 3$. As before, it follows by Eq. (8) that Eq. (6) of the lemma is satisfied for $f^1_1$ and $f^2_1$.  

Lemma 7 implies that $\varphi_i$ satisfies betweenness: $\varphi_i(f^1_1) = \varphi_i(f^2_1)$ implies for all $\zeta \in [0, 1], \quad \varphi_i(f^1_1) = \varphi_i(\zeta f^1_1 + (1 - \zeta) f^2_1)$. Indifference sets of $\varphi_i$ are planar, and parallel on any plane, hence $\varphi_i$ can be represented by a linear function. By unanimity, $\varphi_i$ is linear, and by the anonymity axiom, it is the average of $f_i$.  

Acknowledgements

We wish to thank Brian Barry, Kim Border, Eddie Dekel, Larry Epstein, Edi Karni, Joe Ostroy, Ariel Rubinstein, Joel Sobel, John Weymark, and two anonymous referees for helpful comments.

References


Democratically elected aristocracies

Segal U (2000) Let’s agree that all dictatorships are equally bad. J Polit Econ 108:569–589