Soc Choice Welfare (2007) 29:581–608
DOI 10.1007/s00355-007-0246-z

ORIGINAL PAPER

# Welfarist evaluations of decision rules for boards of representatives

**Claus Beisbart · Luc Bovens**

Received: 12 August 2006 / Accepted: 4 May 2007 / Published online: 2 August 2007


**Abstract** We consider a decision board with representatives who vote on proposals on behalf of their constituencies. We look for decision rules that realize utilitarian and (welfarist) egalitarian ideals. We set up a simple model and obtain roughly the following results. If the interests of people from the same constituency are uncorrelated, then a weighted rule with square root weights does best in terms of both ideals. If there are perfect correlations, then the utilitarian ideal requires proportional weights, whereas the egalitarian ideal requires equal weights. We investigate correlations that are in between these extremes and provide analytic arguments to connect our results to Barberà and Jackson (J Polit Econ 114(2):317–339, 2006) and to Banzhaf voting power.

## 1 Introduction

Consider a group of people (a *society*) with a decision board that votes on proposals. The society is partitioned into *constituencies*. Every constituency has a *representative* on the decision board. What decision rule should we institute for amalgamating the votes of the decision board? This problem has many applications, such as the Council of Ministers in the EU (e.g. Felsenthal and Machover 2002), the electoral college in the US (e.g. Grofman and Feld 2005), or management boards of companies.

C. Beisbart
Institute of Philosophy, Faculty 14, University of Dortmund,
D-44221 Dortmund, Germany
e-mail: Claus.Beisbart@udo.edu

L. Bovens (✉)
Department of Philosophy, Logic and Scientific Method,
London School of Economics and Political Science,
Houghton Street, London, WC2A 2AE, UK
e-mail: L.Bovens@lse.ac.uk

One way of determining the best decision rule is to look for a decision rule that serves the interests of the people. Call this approach *welfarist*. There are two views of serving the interests of the people through democratic decision-making. Bentham writes that "the dictates [of the Public Opinion Tribunal]... coincide, on most points, with those of the *greatest happiness principle*." (Bentham 1980, p. 36; quoted in Harrison 1993, p. 104). This passage indicates that for Bentham, democratic decision making is justified in so far as it maximizes utility in society. Let us call this the utilitarian ideal of democracy. For Rousseau, sovereignty is in the hands of the people and "every act of sovereignty, i.e. every authentic act of the general will, binds or favors all the citizens equally." (Rousseau 1762, II, 4). This is not the place to discuss Rousseau's General Will, but this passage indicates that for Rousseau, democratic decision making is justified in so far as it imposes equal costs and benefits on all citizens, or, in other words, equalizes utility throughout society. Let us call this the egalitarian ideal of democracy. In this paper we will examine which decision rules do best, where "best" is understood in terms of these welfarist ideals. To do so, we will take the distribution of constituencies as given and we will assume that the representatives vote in line with the net interest of their constituencies.

There are different types of constituencies. On the one hand, a constituency may be like an aggregate of citizens who do not have many interests in common. Learning that a proposal is beneficial for one member of the aggregate does not teach us anything about whether it will be beneficial for other members of the aggregate. On the other hand, a constituency may be like an interest group. For any given proposal, the interests of the group members substantially overlap. If we learn that a proposal is beneficial for one member of the interest group, then it is likely to be beneficial for the other members of the group as well. Clearly there is a continuum of possibilities here, but let us first focus on the extreme points. To fix terms, let us stipulate that within an *aggregate* the interests of the people are independently distributed, whereas within an *interest group*, the interests of the people fully overlap.

We have two welfarist ideals for democracy and two types of constituencies. This yields the following four questions:

1. Which decision rule satisfies the utilitarian ideal of democracy for aggregates?
2. Which decision rule satisfies the egalitarian ideal of democracy for aggregates?
3. Which decision rule satisfies the utilitarian ideal of democracy for interest groups?
4. Which decision rule satisfies the egalitarian ideal of democracy for interest groups?

In this work, we present a general procedure to answer these questions. We assume a probability distribution over the utility that an arbitrary proposal yields for each citizen. In this paper, the utilities for people from *different* constituencies are taken to be independent. Also, the marginal distributions for the utilities of individual people are taken to be identical. Under these assumptions, the best decision rule is determined by the type of constituency and the sizes of the constituencies only. For the special case in which we assume a standard normal distribution for each citizen, we obtain the following elegant results. The representatives of aggregates should have weights proportional to the square roots of the sizes of their respective aggregates both on the egalitarian and the utilitarian ideals. The representatives of interest groups should have equal weights on the egalitarian ideal and weights proportional to the sizes of their respective interest groups on the utilitarian ideal.

The relation to the existing literature is as follows. The question of how the legislative body should decide is most often approached in terms of influence (Felsenthal and Machover 1998; FM 1998, henceforth). The normative ideal to be met is supposed to be an equal distribution of influence for the different political agents. Influence is then quantified in terms of voting power measures. The most popular measure of influence is *Banzhaf* voting power (Penrose 1952; Banzhaf 1965; see also FM 1998). Another power measure is the Shapley–Shubik index (Shapley and Shubik 1954; for an application to the EU Council of Ministers see, e.g., Widgrén 1994). Power measures can be interpreted in terms of expected utilities.[1] Our approach is different. We will first address our questions in welfarist terms and then consider whether our results can be mapped onto analytical results in the theory of voting power.

There has recently been growing interest in welfarist assessments of decision rules. Beisbart et al. (2005) rank decision rules for the Council of Ministers according to their expected utility under the interest group model. Inspired by Rawls (1971), Coelho (2005) maximizes the minimal utility of a decision rule. Barberà and Jackson (2006) investigate what decision rule maximizes expected utility under very general conditions. Finally, Bovens and Hartmann (2007) investigate what decision rules satisfy the Rawlsian and utilitarian ideals for constituencies as interest groups. We extend Barberà and Jackson's results by investigating the egalitarian ideal. We extend Bovens and Hartmann's results by investigating constituencies as aggregates. Furthermore, we develop an analytical framework to connect these results and we investigate analytical connections between the utility-based approach and the power-measure approach. Our presentation of the theoretical framework and simple model below builds on Beisbart et al. (2005). There are similarities between our framework and Schweizer (1990).

## 2 The theoretical framework

Consider a *society* of $N$ people. We will label them consecutively from 1 to $N$. We use the uppercase letters $I$ and $J$ as labels for persons. Assume a partition of the people into $n$ *constituencies*. The constituencies are consecutively numbered from 1 to $n$. Labels for constituencies are written in lowercase letters. Constituency $i$ consists of $n_i$ people. Of course, $N = \sum_i n_i$.

The society has a *board* with representatives from the different constituencies. We assume that every constituency has one representative. If a proposal comes to the vote, every representative casts one block vote. We assume that abstention is not possible and that every representative votes either in favor of or against the proposal. Let us say that $\lambda_i = -1$, if the representative from constituency $i$ votes against a proposal, and that $\lambda_i = 1$, if the representative votes in favor of the proposal. Put the $\lambda_i$ s into an $n$-dimensional vector $\lambda$. Different values of this vector correspond to different voting profiles.

[1] It is straight forward to interpret the Shapley–Shubik index in terms of expected utility [e.g., Sketch 6.1.1 on pp. 171 f., Theorem 6.2.6 on p. 181 and Remark 6.3.2 (ii) on pp. 196 f. in FM 1998]. For an interpretation of Banzhaf voting power in terms of expected utility see Beisbart et al. (2005), Appendix.

A decision rule for the board determines for which voting profiles a proposal is accepted. Mathematically, each decision rule $R$ is a function $D^R$ that maps $\lambda$-vectors into the set $\{1, 0\}$. Our convention is that $D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)) = 1$, if the proposal is accepted under $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, and zero otherwise.

We will only consider monotonic decision rules. They are characterized by the following property: If a proposal is accepted under a voting profile, then it will also be accepted under a voting profile with one or more additional yes-votes. In mathematical terms, the condition is as follows: If $D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)) = 1$ and if for each $i = 1, \ldots, n$, $\lambda'_i \geq \lambda_i$, then $D^R((\lambda'_1, \lambda'_2, \ldots, \lambda'_n)) = 1$ (see FM 1998, Definition 2.1.1 on p. 11 for a definition in set-theoretic terms). We will also confine ourselves to rules under which a proposal is accepted if every representative votes yes (see FM 1998, *ibid.*). These are very natural restrictions.

So far, we have set up the structure of a simple voting game ( cf. FM 1998, Definition 2.1.1 on p. 11). For a welfarist assessment of decision rules we have to add the following. Each proposal is described in terms of a utility vector, $(v_1, v_2, \ldots, v_N) \in \mathrm{IR}^N$. The real number $v_I$ is the utility that person $I$ will receive in case the proposal is accepted. It can be assumed without loss of generality that each person gets zero utility, if the proposal is rejected.

It is useful to consider the average utility $w_i$ that a proposal will yield to people from constituency $i$ if it is accepted. $w_i$ is well-defined only if we assume interpersonal comparability of utilities. Clearly,

$$w_i = \sum_{I \text{ in } i} v_I / n_i. \tag{1}$$

Call $u_i$ the average utility that a decision affords to a person from constituency $i$. Clearly, $u_i = w_i D^R((\lambda_1, \lambda_2, \ldots, \lambda_n))$. Accordingly, the net utility $u$ that the whole society receives from the vote on a proposal is

$$u = \sum_i u_i n_i = \sum_i w_i n_i D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)). \tag{2}$$

Here each utility $w_i$ has to be weighed by the number of people $n_i$, since $w_i$ is the *average* utility. Note also that the $D^R$ factor is identical for each addend.

A vote with a given utility vector $(w_1, w_2, \ldots, w_N)$ and under a given decision rule can be conceived of as a game with the representatives as players—suppose that the utilities $w_i$ are known and that it is the representatives (and not the people represented by them) who receive the utilities $w_i$ if the proposal is accepted. Each representative has two pure strategies, viz. voting in favor of or voting against the proposal.

Let us have a closer look at this game (cf. Schweizer 1990, Sects. 3 and 4). Suppose you are representative $i$. The best strategy for you depends on the vector $(w_1, w_2, \ldots, w_n)$. If $w_i > 0$, then voting in favor of the proposal weakly dominates every other strategy (whether pure or mixed). Likewise, if $w_i < 0$, then voting against the proposal weakly dominates every other strategy. If $w_i = 0$, you will not be affected by the vote at all, so both voting for and voting against are weakly dominant strategies. Altogether, there is at least one Nash solution for every game: All players, for whom

$w_i > 0$, vote in favor of the proposal, whereas all other players vote against it. Call the resulting equilibrium $\lambda$-vector $\lambda^*$. The components of the $\lambda^*$-vector are given as $\lambda_i^* = \text{sign}(w_i)$, where the sign function maps negative numbers into $-1$ and positive numbers into 1. We stipulate that sign(0) = $-1$.[2]

This prompts the following question: Given a fixed $(w_1, w_2, \ldots, w_N)$, which decision rule does best under $\lambda^*$, where goodness is specified in terms of our welfarist ideals? The answer is as follows: 1. Given an arbitrary utility vector $(w_1, w_2, \ldots, w_N)$, any decision rule with the following condition maximizes utility under $\lambda^*$: A proposal is accepted under $\lambda^*$, if and only if $\sum_i w_i n_i \geq 0$; 2. Given an arbitrary utility vector $(w_1, w_2, \ldots, w_N)$, any decision rule with the following condition equalizes the $w_i$s under $\lambda^*$: A proposal is rejected under $\lambda^*$, unless the $w_i$s are identical.[3]

Now of course we cannot tune a decision rule to a single proposal. It is not possible either to find a rule that accommodates one of our welfarist ideals *for each and every proposal*. There is no decision rule that maximizes utility for each and every proposal, since two proposals could induce the same voting profile under $\lambda^*$, but the sum total of utility of the former is positive and of the latter is negative. Neither is there a decision rule that could equalize utility for each and every proposal, since, two proposals could induce the same voting profile under $\lambda^*$, but one affords equal utilities whereas the other does not. So every proposal would need to be rejected, which contradicts the requirement that there be acceptance under unanimity.

Instead we need to determine decision rules that fulfill our welfarist ideals in a *probabilistic* way. We will assume a probability distribution over proposals. For fulfilling the utilitarian ideal, we will look for a decision rule that maximizes *expected* utility. For fulfilling the egalitarian ideal, we will look for a decision rule that equalizes *expected* utilities across the constituencies. For this probabilistic approach, we think of the $w_i$s as values of random variables $W_i$. A joint probability distribution over the $W_i$s is assumed. Likewise, $u_i = w_i D^R((\lambda_1, \lambda_2, \ldots, \lambda_n))$ and $u = \sum_i u_i n_i$ become values of random variables, call them $U_i$ and $U$, respectively. Note that we distinguish between the adjectives "average" and "expected". An average utility is a mean with respect to a group of people. An expected utility is a mean with respect to a probability distribution. Of course, there are also expected average utilities.

Let us conclude this section by clearly laying out our framework.

We conceptualize proposals in terms of utility vectors. If a proposal is accepted, it will confer *utility* $v_I$ to person $I$ for each $I = 1, \ldots, N$. The average utility from a proposal for the people in constituency $i$ is called $w_i$. As a proposal is being drafted, its utilities become known, and we have a game with a Nash solution $\lambda^*$ on which each representative $i$ votes for a motion if and only if $w_i > 0$. $\lambda^*$ is taken as input for the decision rule. If the proposal is accepted, the utilities are awarded to the people so that a person from constituency $i$ receives $w_i$ on average. If it is rejected, each person gets zero utility. The average utility of constituency $i$ after the decision is called $u_i$.

[2] The Nash equilibrium $\lambda^*$ is not necessarily unique, e.g. some representative $i$ may vote for a proposal if $w_i = 0$. Working with different Nash equilibria, however, will not affect our results, unless the probability distribution over some $w_i$ has a Dirac delta peak at $w_i = 0$.

[3] Note that, for aggregates, there may still be inequality within constituencies, even if the $w_i$s are identical, since the $v_I$s may be different within constituencies.

We assume an exogenous probability distribution over the utilities $v_I$ from proposals. For simplicity we take it that the $v_I$s follow the same marginal distribution. The utilities for people from different constituencies are independent. Within the constituencies, there is independence under the aggregate model and full dependence under the interest group model. The probability distribution over the $v_I$s implies a probability distribution over the $w_i$s and induces a probability distribution over the $u_i$s via $u_i = w_i D^R((\lambda_1, \lambda_2, \ldots, \lambda_n))$.

Welfarist assessments of a decision rule look at the expected utilities for single persons. Since every person has the same marginal probability distribution for proposals and since there is no way in which persons within a constituency $i$ differ otherwise in our model, the expected utility for a single person equals the expectation value over the average utility $u_i$ in her constituency $i$, $\mathrm{E}[U_i]$. We can therefore concentrate on the $\mathrm{E}[U_i]$s. The challenge is to find a decision rule under which the expected utility for the society $\mathrm{E}[U]$ is maximal and a decision rule under which the expected utilities $\mathrm{E}[U_i]$ are equal, with constituencies as aggregates and as interest groups.[4]

## 3 A simple model

We will now set up a concrete model (cf. Beisbart et al. 2005). Call $p(w_1, w_2, \ldots, w_n)$ the probability density for a proposal with utility vector $(w_1, w_2, \ldots, w_n)$. As we said before, we assume that the probability density $p$ factorizes, as far as the different constituencies are concerned:

$$p(w_1, w_2, \ldots, w_n) = p_1(w_1) p_2(w_2) \ldots p_n(w_n). \tag{3}$$

Let us now calculate expected utilities $\mathrm{E}[U_i]$. We start from their definitions:

$$E[U_i] = \int \mathrm{d}u_1 \mathrm{d}u_2 \ldots \mathrm{d}u_n u_i \, p(u_1, u_2, \ldots, u_n),$$

where $p(u_1, u_2, \ldots, u_n)$ is the probability density over the $(u_1, u_2, \ldots, u_n)$ vectors. Under our model, the values of the $(u_1, u_2, \ldots, u_n)$ vectors are completely determined by the values of the $(w_1, w_2, \ldots, w_n)$ vectors via $u_i = w_i D^R((\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*))$, where the $\lambda_i^*$s denote the votes under the Nash equilibrium: $\lambda_i^*$ = sign(wi).

After inserting the factorization from Eq. (3), this yields

$$\mathrm{E}[U_i] = \int \mathrm{d}w_1 \mathrm{d}w_2 \ldots \mathrm{d}w_n w_i \, p_1(w_1) p_2(w_2) \ldots p_n(w_n) D^R(\mathrm{sign}(w_1), \ldots, \mathrm{sign}(w_n)),$$

The sign($w_i$)s are constant at $-1$ and 1 in the halfspaces $w_i < 0$ and $w_i > 0$, respectively. We can therefore perform the integration over each $w_i$ for each halfspace. We obtain the following sum of $2^n$ addends:

[4] Our calculations in the next section are also compatible with the following weaker assumptions: We could remain silent on the distributions of single persons' utilities from proposals and just work with the average utilities $w_i$ and $u_i$. Then we maximize expected utility across society under the utilitarian ideal, and we equalize the expected *average* utilities under the egalitarian ideal (which does not necessarily imply equality for the people).

$$\mathrm{E}[U_i] = \sum_{\lambda_1} \cdots \sum_{\lambda_n} p^1_{\lambda_1} \cdots p^n_{\lambda_n} w^i_{\lambda_i} D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)); \tag{4}$$

Here the sums extend over $\lambda_i = -1$ and $\lambda_i = 1$. The $p^i_{\lambda_i}$ s and $w^i_{\lambda_i}$ s are defined as follows:

$$p^i_{\lambda i} = \int_{-\infty}^{\infty} \mathrm{d}w_i \, \theta(\lambda_i w_i) \, p_i(w_i) \tag{5}$$

and

$$w^i_{\lambda i} = \int_{-\infty}^{\infty} \mathrm{d}w_i \, \theta(\lambda_i w_i) \; w_i \; p_i(w_i) \Big/ p^i_{\lambda i}. \tag{6}$$

Here $\theta$ is the Heaviside step function[5] $\theta(\lambda_i w_i)$ picks those values of $w_i$ that have the same sign as $\lambda_i$. $p^i_{\lambda i}$ is the probability that the vote of constituency $i$ is $\lambda_i$. $w^i_{\lambda i}$ is the mean utility that is proposed for constituency $i$, given that its vote is $\lambda_i$ (or, equivalently, given that the proposed utility has the same sign as $\lambda_i$).

We learn from Eq. (4) that, for a given society, the only free parameters in our model are the $p^i_{\lambda i}$s and the $w^i_{\lambda i}$s. Other characteristics of the probability densities $p_i(w_i)$ are irrelevant for our purposes. Nevertheless, in order to fix the $p^i_{\lambda i}$s and $w^i_{\lambda i}$s, we have to say something about the probability distributions $p_i(w_i)$. The idea is to constrain the $p_i(w_i)$s by considering the probability distributions over the personal utilities $v_I$.

As we said before, we assume that the probability distribution over $v_I$ is the same for every person $I$ in the society. This means that, at the level of the proposals, there is no bias for or against certain people in the society. We furthermore assume that the probability density over $v_I$ is normal with mean $\mu$ and standard deviation $\sigma$ for every person $I = 1, \ldots, N$.

Following Eq. (1), we can now calculate the probability densities over the $W_i$s, i.e. the $p_i(w_i)$s. We obtain Gaussians with a mean $\mu$ for each $i$. The standard deviation depends on the type of constituency. In order to distinguish between aggregates and interest groups, let us introduce a parameter $\rho$. We set $\rho = 0$ for an interest group and $\rho = 1$ for an aggregate. The meaning of $\rho$ will become clearer later in the paper. Let $s(n_i, \rho)$ denote the standard deviation of $W_i$. For an aggregate, we obtain $s(n_i, 0) = \sigma/\sqrt{n_i}$.[6] For an interest group, the utilities of the people are

[5] The Heaviside step function $\theta(x)$ equals 1, if $x > 0$, and zero otherwise.

[6] This result for an aggregate holds more generally. It does not depend on the assumption that the probability distribution for an individual person is normal. To see this, assume that every person has the same utility distribution with mean $\mu$ and standard deviation $\sigma$. If the number of people per constituency is large, then, by the central limit theorem, the average utility for a person in constituency $i$, $w_i$, is approximately normally distributed with mean $\mu$ and with standard deviation $\sigma/\sqrt{n_i}$. Since there are other more general versions of the central limit theorem (e.g. Feller (1968), pp. 253–256) our result can even be generalized, but we will not discuss the details here.

completely dependent, thus the probability distribution of one person exactly equals the probability distribution of the average utility. So, $s(n_i, 1) = \sigma$.

Our models for aggregates and interest groups are special cases of the block model proposed by Barberà and Jackson (2006). Our aggregate corresponds to a constituency under the fixed-size-block model, with each person forming a block. Our interest group corresponds to a constituency with one block under the fixed-number-of-blocks model.

To obtain results we have to fix $\mu$ and $\sigma$ (i.e. the mean and the standard deviation of the probability density function over the utility $v_I$ for a single person). Without loss of generality, $\sigma$ can be set at 1. This means that utility is measured in units of $\sigma$. We construct a *default model* by setting $\mu$ at 0. It follows from Eq. (5) that the probability density is symmetric and that $p_1^i = p_0^i = 1/2$ and from Eq. (6) that $w_1^i = -w_0^i$. So, under the default model, proposed gains and losses balance each other out on average. Performing the integration in Eq. (6) we obtain

$$w_1^i(n_i, \rho) = s(n_i, \rho)\sqrt{(2/\pi)} \tag{7}$$

This holds in general for any normal distribution with zero mean and standard deviation $s(n_i, \rho)$. Thus for aggregates, both $w_1^i$ and $w_{-1}^i$ are proportional to $1/\sqrt{n_i}$. If the default model is applied to an interest group, the $w_1^i$s are identical for all constituencies $i$ and similarly for the $w_{-1}^i$s.

## 4 Results for the default model

Which decision rules do best in fulfilling our ideals for the different types of constituencies?

There are a very large number of decision rules. In order to simplify matters, we restrict ourselves to weighted rules of a special type. Under weighted decision rules, every constituency $i$ has a weight, call it $x_i$. A proposal is accepted if and only if the sum of the weights for the constituencies voting yes exceeds a certain threshold $t$. In mathematical terms, $D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)) = 1$, if and only if

$$\sum_i \theta(\lambda_i)x_i > t.$$

Lemma 1 in Appendix 1 provides an alternative condition of acceptance under weighted rules: A weighted rule yields acceptance, if and only if

$$\sum_i \lambda_i x_i > 2t - 1.$$

There are many assignments of weights possible. We restrict ourselves to weights $x_i$ that follow the formula

$$x_i \propto n_i^{\alpha}$$

(cf. Felsenthal and Machover 1998, p. 73; Bovens and Hartmann 2007, Eq. 1). Here $\alpha$ is a real number. If $\alpha = 0$, every constituency has the same weight, and we have equal representation. If $\alpha = 1$, the weights are proportional to population sizes, and we have proportional representation. Decision rules with an $\alpha$ strictly between 0 and 1 are called degressively proportional. $\alpha$ is a measure of proportionality. Degressive proportionality is a compromise between equal and proportional representation. We normalize the $x_i$s such that the sum of weights is one. This leaves us with two free parameters, viz $\alpha$ and $t$. Although some pairs of $\alpha$ and $t$ do not yield different decision rules, it is useful to think in terms of $\alpha$ and $t$.

There is no a priori justification for our restriction to weighted decision rules. Later in our paper, however, it will become clear why we restrict ourselves in this way.

In the following we consider the European Union (EU) as an example. Presently, the EU consists of 27 countries. In terms of our model, the EU countries are thought of as constituencies. We calculate the E[$U_i$]s for different values of $\alpha$ and $t$. For this we use a computer program to evaluate the sum in Eq. (4).

We first consider the ideal of *maximizing expected utility*. For *aggregates*, results can be seen in Fig. 1. We show the expected utility for an arbitrary person in the society as a function of $\alpha$ for a series of thresholds ranging from $t = 0.5$ to 0.8. We do not consider thresholds smaller than 0.5. The reason for not considering such thresholds is a symmetry invariance: We obtain almost the same curve, if we switch from threshold $t$ to $(1 - t)$. An analytic argument to this effect is given in Appendix 1.

The figure only contains results for certain thresholds $t$ and does not contain enough detail, but our numerical results indicate that the voting rule ($t = 0.5$, $\alpha = 0.5$) yields

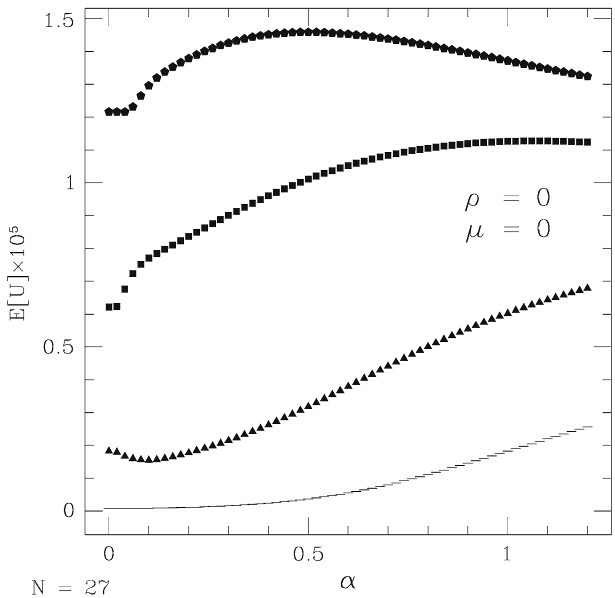


**Fig. 1** Expected utilities under the default model for aggregates as a function of the $\alpha$-parameter. *Filled pentagons* threshold $t = 0.5$; *filled squares* $t = 0.6$; *filled triangles* $t = 0.7$; *bars* $t = 0.8$

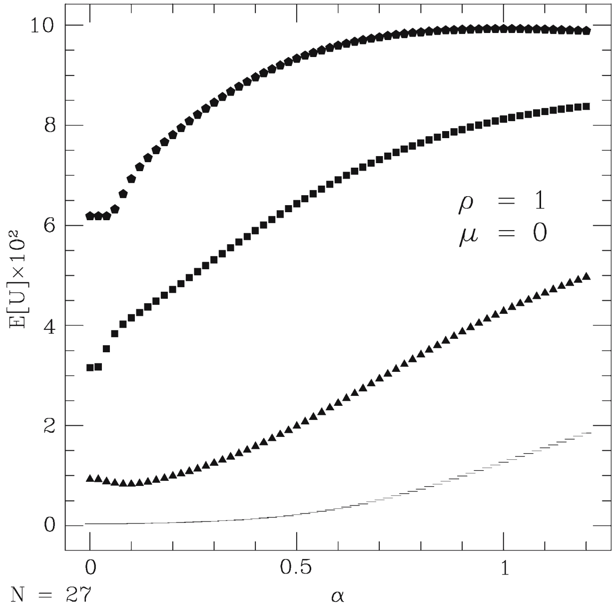


**Fig. 2** Expected utilities under the default model for interest groups. Point styles as in Fig. 1

the highest expected utility for aggregates. We will explain this result in more detail in Sect. 5.

Let us now look at interest groups. Results for expected utility are shown in Fig. 2. The figure only contains results for certain thresholds and does not contain sufficient detail, but our numerical result indicate that the voting rule ($t = 0.5, \alpha = 1$) yields the highest expected utility for interest groups. Again, we will explain this result in more detail in Sect. 5.

Let us now consider the ideal of equalizing expected utilities. To do so it is useful to quantify the degree to which equality is not fulfilled. This yields a measure of inequality. We measure inequality by taking the standard deviation of the expected utilities of a proposal across the people in the society. Under this measure, there is more inequality, if a large constituency is an outlier with respect to the $\mathrm{E}[U_i]$s than if a small constituency is an outlier with respect to the $\mathrm{E}[U_i]$s. Note, however, that our choice is not critical here—other measures that quantify the spread in the $\mathrm{E}[U_i]$s produce similar results.

For aggregates, our measure of inequality is shown in Fig. 3. Again we consider four different thresholds, $t = 0.5, 0.6, 0.7$ and 0.8, so that four curves are obtained. Since our concern is to minimize inequality, we are now looking for a minimum. We observe that for $t = 0.5, 0.6$ and 0.7 every curve has a clear minimum. The minima of these curves are very close to each other, both in terms of $\alpha$ and in terms of our measure of inequality. The $\alpha$-value at which inequality is minimized, is close to 0.5. At this point, the measure of inequality is small compared to the plot range, but not exactly zero.

Note that the minima become less pronounced, if we start with $t = 0.5$ and move to higher thresholds. Since thresholds of $t$ and $(1 - t)$ yield almost always the same

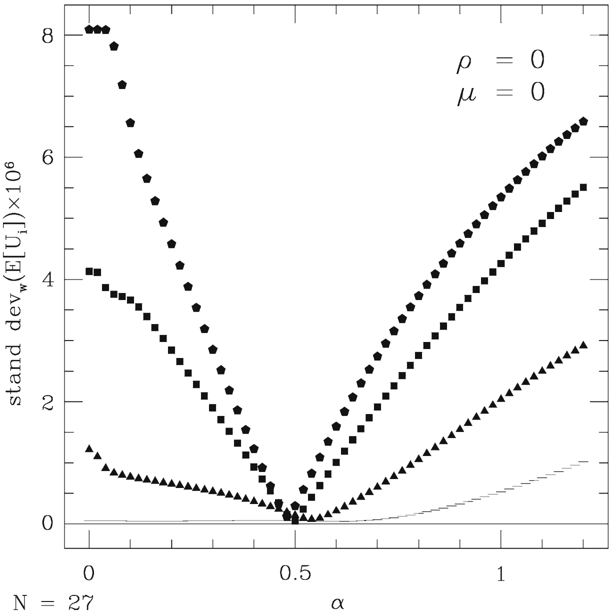


**Fig. 3** Our measure for inequality for the different thresholds under aggregates. Point styles as in Fig. 1

measure of inequality, the same is true for lower thresholds: The minima become less pronounced, as we start from $t = 0.5$ and move towards lower thresholds. For $t = 0.8$ and 0.2, the curves stay almost constant at some low value of our measure of inequality, until $\alpha$ becomes larger than about 0.63. For a threshold of 0.99 or 0.01, say, the whole curves are almost flat at zero. The reason is as follows: Under high thresholds almost no proposal is accepted, so the $E[U_i]$s stay close to zero independently of the value of $\alpha$. Under low thresholds, almost every proposal is accepted, so the $E[U_i]$s approach the mean of the proposals, $\mu$, which is set at zero.

It might well be that none of the minima in Fig. 3 is a *global* minimum, i.e. yields a measure of inequality that is at least as small as anywhere else. A systematic scan of the $(t, \alpha)$ space indicates that we obtain minimum inequality at very low (or at very high thresholds) together with low $\alpha$-values. However, very high (or very low) thresholds are not plausible choices for a decision rule. Thus the minima in Fig. 3 for $t = 0.5$, 0.6 and 0.7 remain most interesting.

Let us now look at interest groups. Results are shown in Fig. 4. Perfect equality is achieved at $\alpha = 0$ and in the neighborhood of $\alpha = 0$, regardless of the threshold. The reason is that, under the interest group model, the probability distributions over the $w_i$s are completely identical for all constituencies $i$. If the constituencies have the same weights, the $E[U_i]$s must be identical for symmetry reasons.

Let us now put together our results so far. Both under the aggregate and the interest group models, expected utility is maximized for a threshold of $t = 0.5$. However, the value of $\alpha$, under which expected utility is maximized, is different for both models. Whereas, under the aggregate model, square root weights ($\alpha = 0.5$) are favored, under the interest group model, proportional weights ($\alpha = 1$) score best. Under the aggregate

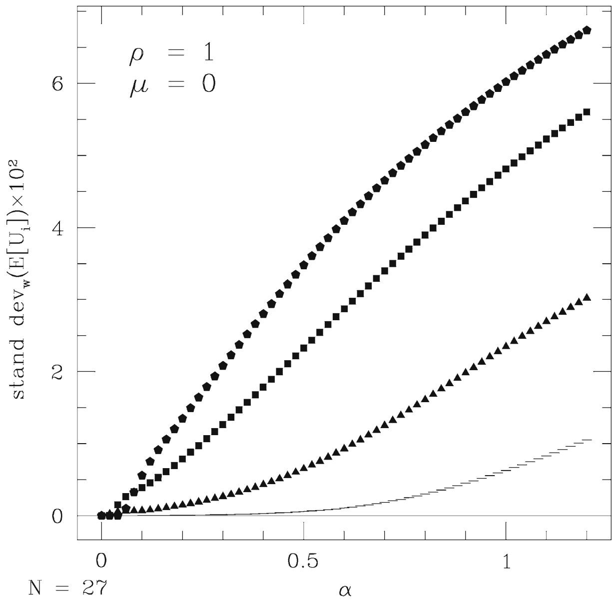


**Fig. 4** Our measure for inequality for the different thresholds under interest groups. Point styles as in Fig. 1

model, there is the following coincidence: If we constrain ourselves to decision rules with a threshold that is not too far away from 0.5, both the egalitarian and utilitarian ideals recommend that $\alpha$ equal about 0.5. This coincidence is lost in the interest group model. For a threshold of $t = 0.5$, maximizing expected utility requires us to set $\alpha$ at 1, whereas equalizing expected utilities requires us to set $\alpha$ at 0.

## 5 Analytic arguments

So far we have produced our results by running a computer program. But is there a way of understanding our results in terms of analytic arguments? Indeed, concerning the utilitarian ideal, our results can be accounted for by a recent result by Barberà and Jackson (Sect. 2.1), and concerning both ideals, our results can be mapped to results known in the voting power literature (Sect. 2.2).

### 5.1 Arguments based upon the Barberà and Jackson result

Let us first turn to expected utilities. Recently, Barberà and Jackson (2006) have provided a very general characterization of decision rules that maximize expected utility. Our findings on the utilitarian ideal for the interest group model and the aggregate model in Sect. 4 are special cases of their result. Drawing on their proof, we can design a simple analytical argument for our results.

We start with an expression for the quantity we would like to maximize, viz E[$U$]. From Eqs. (2) and (4), we obtain

$$\mathrm{E}[U] = \sum_{\lambda 1} \dots \sum_{\lambda n} p^1_{\lambda 1} \dots p^n_{\lambda n} \sum_i n_i w^i_{\lambda i} D^R((\lambda_1, \lambda_2, \dots, \lambda_n)).$$

This is a sum of $2^n$ addends. Every addend represents a voting profile. $\mathrm{E}[U]$ is maximal when every addend is as large as possible. Now every addend is a product of three factors, viz a product of probabilities $p^1_{\lambda 1} \dots p^n_{\lambda n}$, of the sum $\sum_i n_i w^i_{\lambda i}$, and of $D^R((\lambda_1, \lambda_2, \dots, \lambda_n))$. In this product the first and the second factor are determined by our model—they cannot be varied in order to maximize $\mathrm{E}[U]$. Thus, for every addend, i.e. for every voting profile, there is only one choice left: We can make the decision rule accept the proposal with that voting profile or not. That is, we can either set $D^R((\lambda_1, \lambda_2, \dots, \lambda_n))$ to 1 or to zero.

We will maximize $\mathrm{E}[U]$, if we proceed as follows: Set $D^R((\lambda_1, \lambda_2, \dots, \lambda_n))$ to 1 if and only if the product of the two other factors in the addend is positive. Since the first factor, i.e. $p^1_{\lambda 1} \dots p^n_{\lambda n}$, is always non-negative, we have to set $D^R((\lambda_1, \lambda_2, \dots, \lambda_n))$ to 1 if and only if

$$\sum_i n_i w^i_{\lambda i} > 0. \tag{8}$$

Under the default model, the condition in Eq. (8) can further be transformed. Because of the symmetry under the default model, we can write

$$w^i_{\lambda i} = \lambda_i w^i_1. \tag{9}$$

To do the calculations for the different types of constituencies in one turn, we consider $w^i_1$ as a function of $n_i$ and $\rho$: $w^i_{\lambda i} = \lambda_i w^i_1(n_i, \rho)$. Inserting Eq. (9) into Eq. (8) yields an equivalent condition:

$$\sum_i n_i \lambda_i w^i_1(n_i, \rho) > 0. \tag{8$'$}$$

Let us now check, whether there is a weighted voting rule, under which a voting profile yields acceptance if and only if the condition in Eq. (8′) is fulfilled. Since a weighted voting rule yields acceptance if and only if $\sum_i \lambda_i x_i > 2t - 1$ (Lemma 1), the challenge is to find weights $x_i$ and a threshold $t$, such that

$$\sum_i \lambda_i x_i > 2t - 1 \quad \text{if and only if} \sum_i n_i \lambda_i w^i_1(n_i, \rho) > 0$$

for every possible voting profile.

Obviously, this condition is fulfilled if

$$2t - 1 = 0 \quad \text{and} \quad x_i = n_i w^i_1(n_i, \rho).$$

That is, the optimal threshold is 0.5 independently of the type of constituency. This is exactly what we found above. Furthermore, for an aggregate, $w^i_1 = \sqrt{(2/\pi)}/\sqrt{n_i}$

(see our argument after Eq. 6 above), and so weights that are proportional to the square root of the size of the constituency maximize expected utility. For an interest group, $w_1^i = \sqrt{(2/\pi)}$, and so weights proportional to the size of the constituency maximize expected utility.

Note that, in this proof, we did not restrict ourselves to weighted voting rules. It just turned out that a weighted voting rule does best in terms of expected utility. So we know now that this restriction is not critical for maximizing expected utility under the default model.

There is a different analytical argument for this result. This argument relies on the fact that, under the default model, there is an analytic connection between expected utility and Banzhaf voting power indices. This connection will also be useful for understanding our results regarding inequality. We will thus proceed as follows: We will first show, how the expected utilities for the constituencies are related to Banzhaf voting power. We will then provide a second argument for the fact that expected utility maximizes at $(t = 0.5, \alpha = 0.5)$ for aggregates and at $(t = 0.5, \alpha = 1)$ for interest groups. We will conclude by considering inequality.

## 5.2 Arguments based on Banzhaf voting power

Let us begin by defining Banzhaf voting power (for details see FM1998, Chap. 3). As we said in the introduction, Banzhaf voting power measures the influence of the players in voting games. The voting power of a player is the probability of her being pivotal. A voter is pivotal within a given voting profile, if and only if, had she cast a different vote, the outcome of the vote would have been different. For calculating the probability of being pivotal, we have to assume a probability distribution over the voting profiles. For *Banzhaf* voting power, it is assumed that every voter can only vote yes or no. Every voting profile that is possible under this restriction is then assigned the same probability. The resulting Banzhaf voting power of a player $i$ is denoted by $\beta_i'$ in the literature.

In order to establish the relation between expected utilities $\mathrm{E}[U_i]$ and voting powers $\beta_i'$, we calculate the $\mathrm{E}[U_i]$s in a different manner. The idea is to conditionalize to a suitably defined random variable.

We construct a random variable $X$ that equals 0 when a proposal is rejected and constituency $i$ votes for the proposal (votes yes, for short); equals 1 when a proposal is rejected and $i$ votes against the proposal (votes no, for short); equals 2 when a proposal is accepted and constituency $i$ votes yes; and equals 3 when a proposal is accepted and constituency $i$ votes no. We can now calculate $\mathrm{E}[U_i]$ by conditioning on $X$:

$$E[U_i] = \sum_{k=0}^{3} E[U_i|X = k] \times P(X = k).$$

This expression can further be simplified. Clearly for rejected proposals, i.e. for $X = 0$ and $X = 1$, $\mathrm{E}[U_i|X] = 0$. Moreover, because of the symmetry of the default model, we have

$$\mathrm{E}[U_i|X = 3] = -\mathrm{E}[U_i|X = 2].$$

That is, the losses that a person in constituency $i$ will suffer on average, if a proposal is accepted against the vote of constituency $i$, exactly balance the gains that a person in constituency $i$ will receive on average, if a proposal is accepted in line with constituency $i'$s vote.

This leaves us with

$$\mathrm{E}[U_i] = \mathrm{E}[U_i|X = 2](P(X = 2) - P(X = 3)). \tag{10}$$

Under the default model, $\mathrm{E}[U_i|X = 2]$ is the expectation value of the utility of the proposal $w_i$ given that $w_i$ is positive:

$$\mathrm{E}[U_i|X = 2] = \int_0^\infty \mathrm{d}w_i\; w_i\; p_i(w_i) \bigg/ \int_0^\infty \mathrm{d}w_i\; p_i(w_i) = w_1^i.$$

Inserting this into Eq. (10) yields

$$\mathrm{E}[U_i] = w_1^i(P(X = 2) - P(X = 3)).$$

Let us now consider $(P(X = 2) - P(X = 3))$. Clearly,

$$P(X=2) - P(X=3) = P(X=2) + P(X=1) - P(X=1) - P(X=3). \tag{11}$$

$P(X = 2) + P(X = 1)$ is the probability that constituency $i$ is on the winning side (i.e. the proposal is accepted and constituency $i$ votes yes, $X = 2$; or the proposal is rejected and constituency $i$ votes no, $X = 1$). Likewise, $P(X = 1) + P(X = 3)$ is the probability that constituency $i$ votes no (i.e. the proposal is rejected and constituency $i$ votes no, $X = 1$; or the proposal is accepted and constituency $i$ votes no, $X = 3$). Since $\mu = 0$, the probability of constituency $i$ voting against the proposal, is 1/2. Thus, Eq. (11) becomes

$$P(X = 2) - P(X = 3) = P(X = 2) + P(X = 1) - 1/2$$

Now Penrose has established a connection between the probability of a voter casting a vote on the winning side and her voting power. If we apply it at the level of the representatives in the decision board and denote the probability that $i$'s representative is pivotal by $\beta_i'$ the connection reads

$$P(X = 2) + P(X = 1) = (\beta_i' + 1)/2$$

Hence, the r.h.s. of Eq. (11) simplifies to

$$P(X = 2) - P(X = 3) = 1/2\; \beta_i'.$$

We conclude that, altogether,

$$\mathrm{E}[U_i] = 1/2\ \beta_i' w_1^i. \tag{12}$$

Thus, *under the default model, the expected average utility for persons in constituency i under some decision rule R is proportional to the voting power of i's representative under R*.

Everything that follows in this section is based on this result. Note, however, that the equality in Eq. (12) has only been established for the default model, where $\mu = 0$. In general, we cannot reinterpret our expected utilities in terms of voting power in this way.[7]

We will now present our results in terms of analytical arguments involving Banzhaf voting power indices.

*i. The utilitarian ideal*

In order to maximize $\mathrm{E}[U]$, we have to maximize the weighted sum of the $\mathrm{E}[U_i]$, viz.

$$\mathrm{E}[U] = \sum_i n_i \mathrm{E}[U_i] = \sum_i 1/2 n_i \beta_i' w_1^i. \tag{13}$$

This is a weighted sum of powers with weights $y_i = 1/2 n_i w_1^i$. For maximizing $\mathrm{E}[U]$, we can assume that the $y_i$s are normalized (i.e. add up to one) without loss of generality—inserting a constant factor in Eq. (13) does not affect maximization. Drawing on the voting power literature (FM1998, Sect. 3.3), one can prove the following result:

**Theorem 1** *Consider the weighted sum $B = \sum_i y_i \beta_i'$ of Banzhaf voting powers with positive and normalized weights $y_i$. Suppose that decision rule R yields a maximum B. Then a voting profile $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is accepted under R, if $\sum_i y_i \theta(\lambda_i) > 0.5$. A voting profile $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is rejected under R, if $\sum_i y_i \theta(\lambda_i) < 0.5$.*

This result states that the best decision rule is a weighted rule with a threshold of 0.5. A proof for this result is provided in Appendix 2.[8]

Let us now apply Theorem 1 to aggregates and interest groups in turn.

After Eq. (6) it was shown that, for an *aggregate*, $w_1^i = \sqrt{(2/(\pi n_i))}$. Hence, $y_i = 1/2 n_i w_1^i$ is proportional to $\sqrt{n_i}$. Thus, according to Theorem 1, square root

[7] Under the default model, the expected average utility for persons in constituency $i$ is also proportional to the satisfaction index of $i'$s representative. Her satisfaction index is the probability that her vote agrees with the outcome (Rae 1969, cf. also Dubey and Shapley 1979, pp. 124 f. and Straffin et al. 1982). Her satisfaction index can be thought of as expected utility, if we assume that people in her constituency receive a unit utility if she agrees with the outcome and zero utility otherwise. Under the default model, the satisfaction index is a linear transform of Banzhaf voting power (Theorem 3.2.16 on p. 45 in FM1998), and thus of our expected utility. In general, however, the satisfaction index and our expected utility are not related in this simple way.

[8] Theorem 1 does not yet determine whether there should be acceptance or rejection, if the threshold 0.5 is exactly met, i.e. if $\sum_i y_i \theta(\lambda_i) = 0.5$. For most sets of weights, however, this case does not occur. Moreover, as one can learn from the proof of Theorem 1, it does not matter whether the rule yields acceptance or rejection – in both cases a maximum of $B$ is reached.

weights and a threshold of 0.5 are optimal for maximizing E[$U$] under the aggregate model. This is what we found before.

After Eq. (6) it was also shown that, for an *interest group*, $w_1^i = \sqrt{}(2/\pi_i)$. Hence, $y_i = 1/2n_i w_1^i$ is proportional to $n_i$. Thus, according to Theorem 1, proportional weights and a threshold of 0.5 are optimal for maximizing E[$U$] under the interest group model. Again, this is what we found before.

The proof of Theorem 1 cannot simply be mapped onto the proof provided by Barberà and Jackson (2006). In this way Theorem 1 yields a substantially different argument for our results in Sect. 4. If one were to characterize the difference between the Barberà and Jackson (2006) argument and the proof of Theorem 1, one could say the following: Whereas Barberà and Jackson (2006) construct the optimal decision rule straightforwardly by considering every possible voting profile, the proof of Theorem 1 uses the fact that any deviation from a maximum, even if it be very small, yields less than the maximum.

Before turning to equality, let us mention that the r.h.s. of Eq. (13) yields a well-known quantity in voting theory, if the aggregate model is taken for granted for an arbitrary rule $R$ and for suitably large constituencies. It yields up to a constant the sensitivity under a two-tier voting system of the following type: The people are partitioned into constituencies. Within each constituency, a representative is elected by a simple majority vote. The representatives form a decision board and vote on some issue according to rule $R$. The sensitivity is the sum of the people's powers (FM 1998, Definition 3.3.1 on p. 52), where the power of a person is the probability of her being doubly pivotal. A citizen $I$ is doubly pivotal, if and only if the decision of the board would have been different, had $I$ cast a different vote. As one can learn from FM 1998 (see particularly the proof of their Theorem 3.4.3 in FM 1998 on pp. 66), the sensitivity in such a two-tier voting system approaches

$$\sum_i n_i \beta_i' \sqrt{}(2/(\pi_i n_i)),$$

as the $n_i$ is approach infinity. Hence the expected utility for the society is proportional to sensitivity and so *maximizing expected utility is equivalent to maximizing sensitivity*. FM 1998, Theorem 3.4.9 (p. 74) show that sensitivity maximizes under square root weights and a threshold of 0.5.[9]

*ii. The egalitarian ideal*

According to Eq. (12), equalizing the E[$U_i$]s means equalizing the expressions

$$1/2\beta_i' w_1^i,$$

where $w_1^i$ depends on $\rho$ and $n_i$. In other words, we have to find a decision rule, under which the $\beta_i'$s are proportional to the inverse $w_1^i$s.

[9] Under the assumptions of Banzhaf voting power, maximizing sensitivity amounts to minimizing the mean majority deficit according to Theorem 3.3.17 on p. 60 in FM 1998.

In general it is not easy to tune the $\beta_i'$s to some function. However, in the limit of a large number of constituencies, $n$, (and under some additional conditions), voting powers become proportional to weights for weighted voting rule. This is known as Penrose's Theorem and holds for any fixed threshold $t \in (0, 1)$ (see Lindner and Machover 2004 and references therein). As a consequence of Penrose's Theorem, we expect to approach equality in the limit of large $n$, if we set the weights $x_i$ proportional to $1/w_1^i$.

For aggregates, $w_1^i \propto 1/\sqrt{n_i}$, and so square root weights yield the smallest possible inequality in this limit. This provides some explanation why our measure of inequality is comparatively low in the neighborhood of $\alpha = 0.5$ for several thresholds $t$ in Fig. 3. Note, however, that this explanation cannot account for the details of Fig. 3. For instance, we observe that there is always a non-zero measure of inequality left at $\alpha = 0.5$. The reason, of course, is that our explanation is based on taking the limit of large $n$, whereas Fig. 3 considers a society with 27 constituencies only.[10]

For interest groups, $w_1^i$ is constant across society, thus equal weights should minimize inequality in the limit of large $n$. It is easy to see that we achieve perfect zero inequality even for any finite $n$—equal weights always yield equal powers, independently from the threshold. This explains why our measure of inequality is 0 at $\alpha = 0$ and in the neighborhood thereof (see Fig. 4) for all values of $t$ that are considered.

So far, our argument has not yet established that there is no different value of $\alpha$, under which there is less inequality for aggregates, or under which there is also zero inequality for interest groups. In the case of aggregates, things are indeed complicated, as we have said in Sect. 4. Nevertheless, we have argued that the minima around $\alpha = 0.5$ and reasonable thresholds are most interesting for aggregates. For interest groups, the measure of inequality is a monotonically increasing function of $\alpha$ (see Fig. 4). Once we leave the neighborhood of $\alpha = 0$, finite values of our measure of inequality are obtained. As a consequence, there is no point outside the neighborhood of $\alpha = 0$, under which the inequality is zero. There is a simple explanation for this (cf. Barberà and Jackson 2006, Proposition 1). $\mathrm{E}[U_i]$ is proportional to $\beta_i'$. As $\alpha$ increases, larger constituencies have larger weights, whereas smaller constituencies have smaller weights. As a consequence, the voting power of larger constituencies goes up, whereas the voting power of smaller constituencies goes down. The effect is that the spread in voting powers increases.

## 6 Constituencies on the continuum between aggregates and interest groups

So far, our results only apply to aggregates and interest groups under the default model. But in practice, constituencies will typically neither be aggregates nor interest

[10] There is a relation to recent work by Życzkowski and Słomczyński (2004) here. They restrict themselves to $\alpha = 0.5$ and look for a threshold under which the spread in voting power is minimum. They find that $t = 0.62$ is optimal. As we have shown, minimizing the spread in voting power is the same as minimizing the spread in the expected utilities. We can reproduce the results by Życzkowski and Słomczyński (2004) in the following way. Using our measure of spread which is different from theirs, we find a local minimum around $t = 0.62$, as we fix $\alpha$ at 0.5 and consider different thresholds. Note, however, that, according to our Lemma 1, there is a second minimum with the same spread around $t = 1 - 0.62 = 0.38$. Also, on our measure, these minima are not global ones—very high thresholds yield a lower spread even for $\alpha = 0.5$.

groups. In this section, we will relax the assumption that there are either perfect or no correlations at all. We will consider a continuum of different types of constituencies. But we will still assume that there is independence between the utilities of a proposal for the different constituencies.

Start with the personal utilities for people from constituency $i$, the $v_I$s (where $I$ is a member of $i$). Think of them as values of random variables $V_I$. A model that allows correlations of different kinds can be defined as follows: the joint probability density over the $V_I$s, call it $p_i$, is a multivariate normal distribution. Such a multivariate normal is uniquely determined, if the means for the different $V_I$s, call them $\mu_I$, and the covariance matrix, call it $C$, are specified. Properly speaking, $C$ should have an index $i$, since it depends on the constituency, but for simplicity we will suppress it.

We assume that, for any $I$, $\mu_I = 0$. The entries of the covariance matrix, call them $C_{IJ}$, are the following covariances

$$C_{IJ} = \int \mathrm{d}v_I \int \mathrm{d}v_J \, p(v_I, v_J)\,(v_I - \mu_I)\,(v_J - \mu_J).$$

Here $p(v_I, v_J)$ is the joint probability density over $V_I$ and $V_J$. It can be obtained by integrating the multivariate normal $p_i$ over all other variables. If $I = J$, $p(v_I, v_J)$ has to be replaced by the marginal probability density over $V_I$. Accordingly, $C_{II}$ is the variance of the random variable $V_I$. As before, we assume that $V_I$ has the same variance for any member $I$ in society. So the diagonal elements of $C$ are identical. It is therefore useful to normalize all of $C$ in terms of this variance. As before we set this variance at 1.

We need now to specify the non-diagonal elements of $C$, i.e. the correlation pattern. We make the following assumption: For each person $I$, there are $k_i$ other persons $J$ of the same constituency, such that $I$ and $J$'s utilities are correlated with strength $\rho_0$. That is, in each row (column) of the matrix $C$, one can find exactly $k_i$ off-diagonal entries $\rho_0$. The other off-diagonal entries in that row (column) are zero. For instance, each constituency consists of intra-constituency subgroups. Each subgroup has $k_i$ members. Utilities for persons are correlated with strength $\rho_0$, if they are part of the same subgroup, and not correlated otherwise. But there are other correlation patterns that are compatible with our assumptions as well. Because of our normalization of $C$ by the variance, $\rho_0$ is in the interval [0, 1]. $k_i$ may depend on the constituency, whereas $\rho_0$ is taken as fixed throughout the whole society.

In order to get quantitative results for our model we need to derive the distribution of the quantity

$$W_i = 1/n_i \sum_{I \text{ in } i} V_I,$$

on which the calculation of the expected utilities $E[U_i]$ is based. For maximizing the expected utility for society and for equalizing the $E[U_i]$s, the quantity $w_1^i$ will be crucial (see Eqs. 6 and 8′).

It can be shown that, since the $V_I$s follow a multivariate normal, $W_i$ is also normally distributed. A proof for this starts from the multivariate probability density $p_i$ and derives the probability density over $W_i$ directly by suitable integrations.

Since a normal is completely specified in terms of its mean and its variance, we need to calculate only these parameters of the density function over $W_i$. Clearly, since the mean of a sum is the sum of the means, the mean of $W_i$ equals $\mu_I$, which was assumed to be zero. The variance is $s(n_i, \cdot)^2$, where we drop the $\rho$—we are neither talking of aggregates nor of interest groups here. It can be obtained by the following calculation

$$s(n_i, \cdot)^2 = E\left[\left({}^1\!/_{n_i} \sum_{I\ in\ i} v_I - \mu_I\right)^2\right] = {}^1\!/_{n_i^2} \sum_{I,J\ in\ i} E[(v_I - \mu_I)(v_J - \mu_J)]. \quad (14)$$

Here $E$ denotes the expectation value with respect to the joint probability distribution over the random variables $V_I$. This resulting sum in Eq. (14) can be split into two parts. The first part comprises addends for which $I = J$. Each addend of this type yields the variance for personal utility, which was set to 1. So the first part yields $n_i$ times the variance. The second part comprises addends with $I \neq J$. These addends are the off-diagonal entries in the matrix $C$. According to our model, in each row of $C$, there are exactly $k_i$ off-diagonal entries $\rho_0$. Since there are $n_i$ rows in $C$, the second part yields $n_i k_i \rho_0$. Hence, the standard deviation of $W_i$ equals

$$s(n_i, \cdot) = \surd((1 + k_i \rho_0)/n_i). \quad (15)$$

We can now obtain the values of the $w_1^i$s by inserting Eq. (15) into Eq. (7):

$$w_1^i = s(n_i, \cdot, )\sqrt{2/\pi} = \sqrt{2(1 + k_i \rho_0)/n_i \pi}.$$

Note, also, that, because $W_i$ has zero mean, $p_{\lambda i}^i = 1/2$ for $\lambda_i = -1, 1$, each.

So far we have not yet said anything about $k_i$ and $\rho_0$. We assume that $\rho_0$ is constant over the whole society, whereas $k_i$ may vary between constituencies. Let us explore two possibilities.

(i) $k_i$ is the same for all constituencies. For instance, regardless of whether you live in a large or a small constituency, you are only correlated with other members of your family, where family size is constant over the whole society. In this case, the $w_1^i$s are proportional to one over the square root of $n_i$. This is exactly like in the aggregate model. We therefore know already how to fulfill our welfarist desiderata. We will obtain maximum utility if and only if the threshold is set at 0.5 and if the weights are proportional to the square root of $n_i$. A relative minimum of the inequality will originate for an $\alpha$-value close to 0.5 under reasonable thresholds not too far from 0.5.

(ii) $k_i$ depends on the size of the constituency—it is larger for larger constituencies. A natural assumption is the following: For every person in constituency $i$, there are exactly $(n_i - 1)$ people from $i$ left with whom she could be correlated. We assume that she is correlated with a constant fraction $k$ of them, such that $k_i = k(n_i - 1)$. Hence,

$$w_1^i = \sqrt{2(1 + k\rho_0(n_i - 1))/n_i\pi}.$$

$k\rho_0$ can be equated with our earlier parameter $\rho$, since, if we set $k\rho_0$ at 0, we recover our aggregates, whereas, if we set $k\rho_0$ at 1, all people in the constituency are perfectly correlated and we recover our interest groups. So we now have a continuum between aggregates and interest groups.

What do the welfarist solutions look like for intermediate cases of $\rho$? If we want to maximize expected utility, we need a threshold of 0.5 and weights proportional to $w_1^i n_i$, that is, proportional to

$$n_i\sqrt{2(1 + \rho(n_i - 1))/n_i\pi} \tag{16}$$

in accordance with our proofs in Sect. 5. Remarkably, such weights are in general not compatible with an $\alpha$-rule.

In order to equalize expected utility under reasonable thresholds around 0.5, we can set the weights proportional to

$$\sqrt{n_i\pi/(2(1 + \rho(n_i - 1)))} \tag{17}$$

in the limit of large $n$, in accordance with our argument in Sect. 5 (ii).

What do the weights that Eqs. (16) and (17) require look like under realistic conditions? Let us consider the example of the European Union again. In the EU, the $n_i$s are of the order of $10^5$ at least. Thus, in the term

$$(1 + \rho(n_i - 1))$$

we can substitute $n_i$ for $(n_i - 1)$ and drop the first addend—unless $\rho$ is significantly smaller than $1/n_i$. In this approximation, (16) and (17) yield the same results as our results for an interest group model, i.e. expected utility is maximized for proportional weights, whereas equality is achieved at equal weights. Now $1/n_i$ is very small for all the constituencies in the EU and hence for values of $\rho$ that are not extremely small we run into a situation that is close to the interest group model.

An illustration can be seen in Fig. 5. We show the weights that score best under the utilitarian ideal (Eq. 16) or that yield a very small measure of inequality (Eq. 17). We consider two values of $\rho$, viz. $\rho = 0$ and $\rho = 5 \times 10^{-6}$. For $\rho = 0$, square root weights are optimal for both the utilitarian and the egalitarian ideal. Already for $\rho = 5 \times 10^{-6}$, the weights that are optimal on the utilitarian ideal are close to proportional weights, whereas the weights that yield a very small measure of inequality are close to equal weights. $\rho = \rho_0 k = 5 \times 10^{-6}$ holds if everyone's utility distribution displays rather weak correlations of strength $\rho_0 = 0.1$ with about $k = 0.005\%$ of the total population of one's constituency, e.g. with about 3,000 people in the UK. This does not seem like an unreasonably high $\rho$-value. We therefore conclude that, for

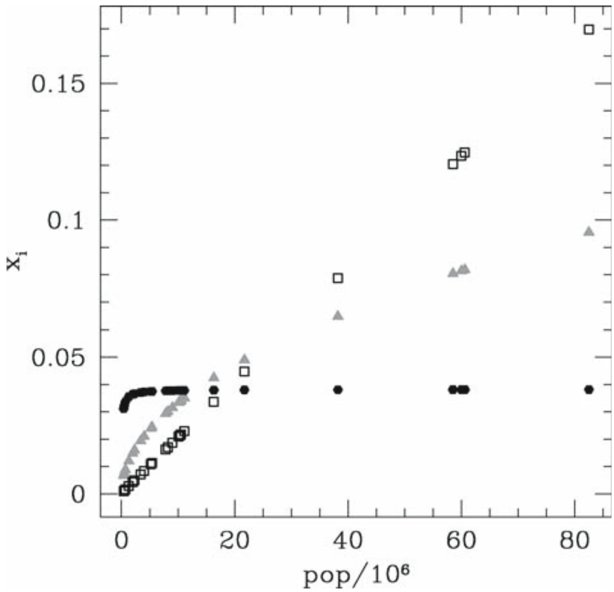


**Fig. 5** Optimal weights for the utilitarian ideal (*grey triangles* $\rho = 0$; *open squares* $\rho = 5 \times 10^{-6}$) and the egalitarian ideal (*grey triangles* $\rho = 0$; *filled hexagons* $\rho = 5 \times 10^{-6}$)

societies with large constituencies and reasonable utility correlations, implementing the welfarist ideals corresponds to implementing these ideals for interest groups rather than for aggregates under correlations of type (ii). The aggregate model is unstable once we start introducing even minimal correlations of type (ii).

## 7 Conclusions

A society is partitioned in constituencies of unequal sizes. Each constituency has a representative on the decision board. The board votes on proposals and these votes are amalgamated by means of a decision rule. Our question is, what constitutes an appropriate decision rule?

We start with two ideals of democracy. For utilitarians, democratic procedures should maximize the expected utility of proposals in society. For (welfare) egalitarians, democratic procedures should equalize the expected utilities of proposals for all persons in society. Both ideals pose optimization problems, viz. what decision rule maximizes the expected utility of an arbitrary proposal or minimizes the distance from equal expected utility. Furthermore, we distinguish between constituencies as aggregates *versus* as interest groups. In an aggregate, the utilities of the members are independent. In an interest group, the utilities of the members are perfectly aligned. To keep things simple, we assume that there are no inter-constituency dependencies.

In our default model, we construct a standard normal as the probability density function over the utility of a proposal for the members of the society. We construct a family of decision rules that can be parameterized by a threshold of acceptance $t$

and a measure of proportionality $\alpha$ ranging from equal to proportional representation. We take the European Union as our example, run a computer program, and scan the $(t, \alpha)$-space to solve our optimization problems.

For aggregates, we found the following surprising coincidence. Both the utilitarian and the egalitarian ideals find optima in the neighborhood of weights proportional to the square roots of the constituency sizes.[11] For interest groups, this coincidence disappears. On the utilitarian ideal, expected utility is maximal for proportional weights and on the egalitarian ideal, the expected utilities for the various constituencies are equal for equal weights. On the utilitarian ideal, thresholds of 0.5 yield the highest expected utility whereas on the egalitarian ideal, the threshold is of no or little consequence.

We then construct analytical arguments to support these computational results and explore connections to the existing literature. For the utilitarian ideal, we draw on the work by Barberà and Jackson. Subsequently, we show how the optimization results on both the egalitarian and the utilitarian ideal can be connected to existing analytical results in the voting power literature, drawing on the work by Felsenthal and Machover.

We consider how stable our results are when moving on a continuum from aggregates to interest groups. Clearly real constituencies are neither interest groups nor aggregates. There are some dependencies between the interests of various parties within a constituency without these interests being fully aligned. Now as long as these dependencies are within interest groups of a fixed size, say, within fixed-size family units, optimization results are the same as for aggregates. But dependencies will also occur between the interests of the members of intra-constituency subgroup (say professional groups) and the sizes of such groups tend to be proportional to constituency size. We construct a parameter $\rho$ that ranges from zero for aggregates to one for interest groups. As we move on this continuum, intra-constituency subgroups come to play a more and more important role in society. We show that even if there are only weak correlations between relatively small intra-constituency subgroups whose sizes are proportional to the constituency size, optima swiftly move away from the optima for aggregates and towards the optima for interest groups.

The welfarist evaluation of decision boards provides for a rich research framework in which a lot of questions still remain unanswered. We conclude by flagging a few of these questions here:

i There is a philosophical problem of what constitutes a reasonable ideal for a decision rule to amalgamate the votes of a decision board for a society consisting of constituencies of different sizes. As long as the constituencies are aggregates or have fixed-size interest-groups, both ideals point in the same direction. But as constituencies come to have more variable-size interest groups, a balance between the utilitarian and egalitarian ideals needs to be struck. But how much weight should these ideals carry? And are there circumstances in which one ideal should carry more or less weight?

ii So far we have assumed that there are no inter-constituency dependencies. But clearly this is not a realistic assumption. Within the European Union, the interests

[11] Whereas the optimum is global for the utilitarian ideal, it is only local for the egalitarian ideal. See Sect. 4 for details.

of Mediterranean countries, of the newly ascending countries, of larger countries etc. are to a certain extent aligned. To account for such dependencies, empirical data and more sophisticated models need to be brought into our framework.

iii So far we have assumed that constituencies strictly vote in their self-interest. But countries tend to vote as factions. A country may support a befriended country on a particular proposal although it is strictly speaking not in its interest, but they know that they can count on reciprocal support for proposals that are important to them. We have investigated faction formation in Bovens and Beisbart (2007).

iv We model the probability distribution over proposals as exogenous. But the probability distribution might depend on the decision rule. For instance, proposals might be drafted in such a way that a certain percentage is likely to pass. We do not take into account this possibility here, but our work provides a theoretical framework to address such a possibility. Once we know how the probability distribution depends on the decision rule, we can compare different rules under different probability distributions, as we compared different rules under the same probability distribution in this paper.

v We also assumed that the marginal probability densities are identical for all members of the society. However, the proposing board might be biased towards some people. For instance, people from certain constituencies might be favored. This would require more sophisticated modeling with more free parameters that have to be fixed in a reasonable way. In order to do so, empirical data may be required.

## Appendix 1. Some results concerning weighted rules

A weighted decision rule can be described in terms of weights $w_i$ and a threshold $t$. A proposal is accepted if and only if

$$\sum_i \theta(\lambda_i)x_i > t.$$

Lemma 1 provides an equivalent condition for acceptance.

**Lemma 1** *Let $x_i$ be a set of non-negative and normalized weights. For any voting profile $(\lambda_1, \ldots, \lambda_n)$ with $\lambda_i = 1$ or $\lambda_i = -1$ for $i = 0, \ldots, n$, we have*

$$\sum_i \theta(\lambda_i)x_i > t \quad \textit{if and only if} \sum_i \lambda_i x_i > 2t - 1.$$

*The same holds with the reverse sign < or the = sign instead.*

*Proof* Because $\lambda_i$ is either 1 or minus 1, we have $\lambda_i = \theta(\lambda_i) - \theta(-\lambda_i)$ and $1 = \theta(\lambda_i) + \theta(-\lambda_i)$. Using this and $\sum_i x_i = 1$, we see that

$$\sum_i \lambda_i x_i > 2t - 1$$

is equivalent to

$$\sum_i \theta(\lambda_i)x_i - \sum_i \theta(-\lambda_i)x_i > 2t - \sum_i x_i = 2t - \sum_i \theta(\lambda_i)x_i - \sum_i \theta(-\lambda_i)x_i.$$

The last terms on the l.h.s. and r.h.s. are the same, so we end up with

$$\sum_i \theta(\lambda_i)x_i > 2t - \sum_i \theta(\lambda_i)x_i$$

or

$$\sum_i \theta(\lambda_i)x_i > t.$$

We can now prove the following proposition. Its meaning will be explained after the proof.

**Proposition 1** *Let $R$ be a weighted rule with weights $x_i$ and threshold $t$. Let $S$ be the weighted rule with the same weights, but with threshold $(1-t)$. Suppose that there is no voting profile $(\lambda_1, \ldots, \lambda_n)$, for which the threshold is exactly met, i.e. for which $\sum_i \theta(\lambda_i)x_i = t$ or, equivalently, for which $\sum_i \lambda_i x_j = 2t - 1$. Then the expected utilities $E[U_i]$s are the same under $R$ and $S$ for each $i = 1, \ldots, n$.*

*Proof* We start with Eq. (4) from Sect. 3

$$E[U_i] = \sum_{\lambda 1} \cdots \sum_{\lambda n} p^1_{\lambda 1} \cdots p^n_{\lambda n} w^i_{\lambda i} D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)).$$

According to Lemma 1, under rule $R$, $D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)) = \theta(\sum_j \lambda_j x_j - (2t-1))$. Likewise, under rule $S$, $D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)) = \theta(\sum_j \lambda_j x_j - (2(1-t)-1)) = \theta(\sum_j \lambda_j x_j + (2t-1))$. Thus, under rule $S$:

$$E[U_i] = \sum_{\lambda 1} \cdots \sum_{\lambda n} p^1_{\lambda 1} \cdots p^n_{\lambda n} w^i_{\lambda i} \theta\left(\sum_j \lambda_j x_j + (2t-1)\right).$$

For each $k = 1, \ldots, n$, the sum extends over $\lambda_k = -1$ and $\lambda_k = 1$. We switch to new summation indices $\kappa_k = -\lambda_k$ and obtain

$$E[U_i] = \sum_{\kappa 1} \cdots \sum_{\kappa n} p^1_{-\kappa 1} \cdots p^n_{-\kappa n} w^i_{-\kappa i} \theta\left(-\sum_j \kappa_j x_j + (2t-1)\right).$$

For any argument $x \neq 0$, we have $\theta(-x) = 1 - \theta(x)$. If there is no voting profile, under which the threshold $t$ is exactly met, then, according to Lemma 1, $0 \neq$

$-\sum_j \kappa_j xj + (2t-1)$. Thus,

$$E[U_i] = \sum_{\kappa 1} \cdots \sum_{\kappa n} p^1_{-\kappa 1} \cdots p^n_{-\kappa n} w^i_{-\kappa i} \left(1 - \theta\left(\sum_j \kappa_j x_j - (2t-1)\right)\right).$$

Under the default model, $p^k_{-\kappa k} = p^k_{\kappa k} = 0.5$ and $w^k_{-\kappa k} = -w^k_{\kappa k}$. Thus,

$$\begin{aligned} \mathrm{E}[U_i] &= -\sum_{\kappa 1} \cdots \sum_{\kappa n} p^1_{\kappa 1} \cdots p^n_{\kappa n} w^i_{\kappa i} \left(1 - \theta\left(\sum_j \kappa_j x_j - (2t-1)\right)\right) \\ &= \sum_{\kappa 1} \cdots \sum_{\kappa n} p^1_{\kappa 1} \cdots p^n_{\kappa n} w^i_{\kappa i} \theta\left(\sum_j \kappa_j x_j - (2t-1)\right) \\ &\quad - \sum_{\kappa 1} \cdots \sum_{\kappa n} p^1_{\kappa 1} \cdots p^n_{\kappa n} w^i_{\kappa i}. \end{aligned}$$

In the last equation, the second term of the r.h.s. disappears for symmetry reasons—the summation extends over $\kappa_k = -1$ and $\kappa_k = 1$, and $w^k_{-\kappa k} = -w^k_{\kappa k}$, whereas all $p^k_{\kappa k}$ equal 0.5. The remaining part is just $\mathrm{E}[U_i]$ under rule $R$, with summation indices $\kappa_k$ instead of $\lambda_k$.

Proposition 1 entails that, in investigating the utility distribution under different weighted voting rules, we need not check the whole interval of thresholds [0, 1], if we restrict ourselves to the default model. It suffices to study the interval [0.5, 1]. This is why we showed only curves for thresholds from 0.5 on up in Sect. 3.

In applying Proposition 1, one always has to check that the threshold is not exactly met for a voting profile. However, for a decision board with a finite number of representatives, there are only finitely many thresholds $t$ that are exactly met for some voting profile. Thus, if we start with fixed weights and plot the expected utility for some constituency for different thresholds, then the curve will be symmetric with respect to 0.5 except for finitely many points.

In Sect. 4 we prove that standard voting power is proportional to expected utility under the default model. It follows immediately that, under the conditions of Proposition 1, voting power is not affected, if we switch the threshold from $t$ to $(1-t)$.

## Appendix 2. Maximizing a weighted sum of voting powers

In this Appendix we prove Theorem 1, which concerns the quantity $B = \sum_i y_i \beta'_i$. We assume that $\sum_i y_i = 1$. We need the following result from the voting power literature (Lemma 3.3.12 and Corollary 3.3.13 in FM 1998, pp. 57).

**Lemma 2** *Consider a decision rule R for a board where n votes are cast.*

a. Suppose that the voting profile $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ yields acceptance under $R$, i.e. $D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)) = 1$. Suppose, furthermore, that $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is minimal in the following sense: If $(\lambda_1', \lambda_2', \ldots, \lambda_n')$ is a different voting profile, for which, $\lambda_i' \leq \lambda_i$, for all $i$, then $R$ yields rejection. Consider now a decision rule $S$ that yields rejection under $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and is identical with $R$ otherwise. The voting powers under $S$, call them $\beta_i'(S)$, can then be expressed in terms of the voting powers under $R$, call them $\beta_i'(R)$ in the following way: $\beta_i'(S) = \beta_i'(R) - 2^{1-n}$, if $\lambda_i = 1$; and $\beta_I'(S) = \beta_i'(R) + 2^{1-n}$ otherwise.
b. Likewise, suppose that the voting profile $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ yields rejection under $R$, i.e. $D^R((\lambda_1, \lambda_2, \ldots, \lambda_n)) = 0$. Suppose, furthermore, that $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is maximal in the following sense: If $(\lambda_1', \lambda_2', \ldots, \lambda_n')$ is a different voting profile, for which $\lambda_i' \geq \lambda_i$ for all $i$, then $R$ yields acceptance. Consider now a decision rule $S$ that yields acceptance under $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and is identical with $R$ otherwise. The voting powers under $S$ can then be expressed in terms of the voting powers under $R$ in the following way: $\beta_i'(S) = \beta_i'(R) + 2^{1-n}$, if $\lambda_i = 1$; and $\beta_i'(S) = \beta_i'(R) - 2^{1-n}$ otherwise.

The proof of Theorem 1 proceeds as follows (cf. FM 1998, proof of Theorem 3.3.14, p. 59, their proof goes back to Dubey and Shapley 1979, p. 107). Assume that decision rule $R$ yields maximum $B = \sum_i y_i \beta_i'$. Suppose that, contrary to the second claim in Theorem 1, there is a voting profile $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ that yields acceptance under $R$ and under which $\sum_i y_i \theta(\lambda_i) < 0.5$. Without loss of generality we can assume that $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is minimal in the sense of Lemma 2a—if $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is not yet minimal, there must be a different minimal voting profile $(\lambda_1', \lambda_2', \ldots, \lambda_n')$ with $\sum_i y_i \theta(\lambda_i') < 0.5$ because of monotonicity. Let us now construct a rule $S$ as it is done in Lemma 2a. According to Lemma 2a, $\beta_i'(S) = \beta_i'(R) - 2^{1-n}$, if $\lambda_i = 1$, and $\beta_i'(S) = \beta_i'(R) + 2^{1-n}$ otherwise. It follows that, under $S$, $B(S) = \sum_i y_i \beta_i'(S) = \sum_i y_i \beta_i'(S)(\theta(\lambda_i) + \theta(-\lambda_i)) = \sum_i y_i(\beta_i'(R) - 2^{1-n})\theta(\lambda_i) + \sum_i y_i(\beta_i'(R) + 2^{1-n})\theta(-\lambda_i) = B(R) - 2^{1-n} \sum_i y_i \lambda_i$. Now we assumed that $\sum_i y_i \theta(\lambda_i) < 0.5$. It follows from Lemma 1, that $\sum_i y_i \lambda_i < 0$. Therefore, $B(S)$ is larger than $B(R)$, which is a contradiction with our assumption that $B(R)$ is maximal. The first claim of Theorem 1, viz. that any voting profile with $\sum_i y_i \theta(\lambda_i) > 0.5$ is accepted under $R$, is proved in the same way by using part b of Lemma 2.

**Acknowledgements** We are grateful to Stephan Hartmann, Maurice Salles, Robert Sugden and the audience of the conference "The Philosophical Aspects of Social Choice and Welfare" in Caen, June 2005, for their contributions and/or comments. Our research was supported by the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government through a Sofja Kovalevskaja Award.

## References

Banzhaf JF (1965) Weighted voting doesn't work: a mathematical analysis. Rutgers Law Rev 19:317–334
Barberà S, Jackson M (2006) On the weights of nations: assigning voting power to heterogeneous voters. J Polit Econ 114(2):317–339
Bentham J (1980) Correspondence. In: Dinwiddy JR (ed) vol IV. Athlone, London

Beisbart C, Bovens L, Hartmann S (2005) A utilitarian assessment of alternative decision rules in the council of ministers. European Union of Politics 6:395–418 with appendix at http://www.uni-konstanz.de/eup/Data/6-4/Beisbart2_app.doc

Bovens L, Hartmann S (2007) Welfare, voting and the constitution of a federal assembly. In: Galavotti MC, Scazzieri R, Suppes P (eds) Reasoning, rationality and probability, (forthcoming) CSLI Publications, Stanford

Bovens L, Beisbart C (2007) Factions in Rousseau's 'Du Contrat Social' and Federal Representation. Analysis 67:12–20

Brams SJ, Davis MD (1974) The 3/2's Rule in presidential campaigning. Am Polit Sci Rev 68:113–134

Coelho D (2005) Maximin choice of voting rules for committees. Econ Gov 6:159–175

Dubey P, Shapley LS (1979) Mathematical properties of the Banzhaf power index. Math Oper Res 4:99–131

Feller W (1968) An introduction to probability theory and its applications, vol 1, 3rd edn. Wiley, New York

Felsenthal DS, Machover M (1998) The measurement of voting power: theory and practice, problems and paradoxes. Edward Elgar, Cheltenham

Felsenthal DS, Machover M (2002) Enlargement of the EU and weighted voting in its councils, revised version 2002. online at http://eprints.lse.ac.uk/archive/00000407/01/euenbook.pdf (checked April 2007)

Grofman B, Feld SL (2005) Thinking about the political impacts of the electoral college. Public Choice 123:1–18

Harrison R (1993) Democracy. Routledge, London

Lindner I, Machover M (2004) L. S. Penrose's limit theorem: proof of some special cases. Math Social Sci 47:37–49

Penrose LS (1952) On the objective study of crowd behaviour. H. K. Lewis & Co, London

Rae DW (1969) Decision rules and individual values in constitutional choice. Am Polit Sci Rev 63:40–53

Rawls J (1971) A theory of justice. Harvard University Press, Cambridge

Rousseau JJ (1993 [1762]) The social contract and discourses. Translated by GDH Cole. Everyman, London

Schweizer U (1990) Calculus of consent: A game-theoretic perspective. J Inst Theor Econ 146:28–54

Shapley LS (1953), A value for $n$-Person games. In: Kuhn HW, Tucker AW (eds) Contributions to the theory of games, vol II. Princeton University Press, Princeton, pp 307–317

Shapley LS, Shubik M (1954) A method for evaluating the distribution of power in a committee system. Am Polit Sci Rev 48:787–792

Straffin Jr PD, David MD, Brams SJ (1982) Power and satisfaction in an ideologically divided voting body. In: Holler M (ed) Voting, power, and voting power. Physica-Verlag, Würzburg, pp 239–255

Widgrén M (1994) Voting power in the EC decision making and the consequences of two different enlargements. Eur Econ Rev 38:1153–1170

Życzkowski K, Słomczyński W (2004) Voting in the European Union: The square root system of Penrose and a critical point, Mimeo. http://arxiv.org/ftp/cond-mat/papers/0405/0405396.pdf (checked April 2007)