On the Egalitarian Weights of Nations

Sascha Kurz
Dept. of Mathematics, University of Bayreuth

Nicola Maaser
Dept. of Economics, University of Bremen

Stefan Napel
Dept. of Economics, University of Bayreuth;
Public Choice Research Centre, Turku

January 31, 2013

Abstract

Voters from m disjoint constituencies (regions, federal states, etc.) are indirectly represented in an assembly which applies a weighted voting rule. All agents have single-peaked preferences over an interval and each delegate's preferences match the constituency's *median voter*. The collective decision corresponds to the assembly's *Condorcet winner*. Which voting weights $w_1, \ldots, w_m$ ought to be selected if constituency sizes differ and all voters are to have *a priori equal influence* on collective decisions? It is shown that representation is approximately egalitarian for weights proportional to the *square root* of constituency sizes if all ideal points are i.i.d. If, however, preferences are polarized along constituency lines then weights should induce a Shapley value *linear* in size.

Keywords: collective choice; institutional design; two-tier voting systems; equal representation; Shapley value; pivot probability; voting power; random order values

JEL codes: D02; D63; D70; H77

Napel gratefully acknowledges financial support from the Academy of Finland through the Public Choice Research Centre, Turku. We thank Matthew Braham, Ulrich Kamecke, Mario Larch, Ines Lindner, Vincent Merlin, Abraham Neyman and Andreas Nohn for very helpful comments on earlier drafts. We have also benefitted from discussions with Robert Hable and feedback on seminar presentations in Aachen, Amsterdam, Augsburg, Bamberg, Bayreuth, Bilbao, Bozen, Caen, Freiburg, Hamburg, Maastricht, Munich, Paderborn, Salerno, Tilburg, Turku, Zurich and the Economic Theory committee of the Verein für Socialpolitik. The usual caveat applies.

1 Introduction

The voting weights of delegations to electoral assemblies with a federal or divisional structure commonly vary in the size of the represented populations, but do so very differently. The US Electoral College, for instance, involves voter blocs that are broadly proportional to constituency size: each state has two votes (reflecting its two seats in the Senate) in addition to a number which is proportional to population (like House seats). California and Wyoming with around 11.9% and 0.2% of the US population thus end up holding around 10.2% and 0.6% of votes on the US President. In contrast, the most and least populous member states of the EU – Germany and Malta – currently have about 8.4% and 0.8% of votes in the Council of the European Union but comprise 16.3% and 0.1% of the EU population; the respective mapping from population size to voting weight is, very roughly, a square root function.[1] Delegates in other collective decision-making bodies, such as the Senate of Canada, the German Bundesrat, the Governing Council of the European Central Bank, and many a university senate or council of a multi-branch NGO, have voting weights that are yet more concave functions of the number of represented constituents, or even flat.

This paper analyzes the fairness of different voting weight arrangements from a theoretical perspective. Individuals vote on delegates or representatives in disjoint constituencies (bottom tier) and these representatives take collective decisions in a council, electoral college, or other assembly (top tier). We investigate the practically relevant, normative question: which simple function – possibly linear, possibly strictly concave or constant – *should* determine the top-tier voting weights of delegates from differently sized constituencies, such as US states or EU member countries? The considered objective is not one of efficiently aggregating private information (see, e.g., Bouton and Castanheira 2012) or of maximizing a utilitarian measure of welfare as investigated, for instance, by Barberà and Jackson (2006). We focus on the egalitarian criterion of *'one person, one vote'* and on providing all bottom-tier voters, at least a priori and under very stylized ideal conditions, with *equal influence* on the collective decision.

We assume a left–right policy spectrum and consider two-tier voting systems in which the respective median voter determines a constituency's top-tier policy position and the assembly's *Condorcet winner* defines its collective decision. The relation of *heterogeneity within each constituency* and *heterogeneity across constituencies* turns out to be the critical determinant of a fair voting weight allocation. Linear and square root weighting rules emerge in particularly prominent benchmark cases. The former is advisable for electorates that are polarized along constituency lines, i.e., exhibit significant heterogeneity across constituencies; while the latter is more egalitarian

[1] A least squares power-law regression of EU Council voting weights w_i on population sizes n_i results in $w_i = c \cdot n_i^{0.47}$ with $R^2 \approx 0.95$. The current Council voting rules involve two other but essentially negligible criteria, and will be changed in 2017 into a more proportional system.

when heterogeneity within each constituency is dominant. These are conclusions from two new asymptotic results (Theorems 1 and 2) which characterize the limit behavior of pivot probabilities for committee decisions on interval policy spaces.

The 'one person, one vote' principle is linked to the requirement of *anonymity* in social choice theory, that is, collective decisions shall depend only on the votes that the alternatives receive, not on whose votes these are. This general egalitarian norm is sometimes considered the minimum requirement for a decision-making procedure to be called 'democratic' (e.g., Dahl 1956, p. 37). It is straightforward to implement – at least in theory – in case of a direct, single-tier voting procedure or a two-tier one with symmetric constituencies. Complications arise when a two-tier system is asymmetric. A non-trivial integer apportionment problem already needs to be resolved for those assemblies, like parliaments, in which delegates from the same constituency can split their votes and thus reflect heterogeneity among constituents (see Balinski and Young 2001). Significant complications are added if all representatives of a constituency vote as a bloc (as in the US Electoral College, with two exceptions) or, equivalently, when the assembly contains a single delegate from each constituency who is endowed with a voting weight that increases in population size (as in the EU Council).

A way to adapt the principle to such situations has very vaguely been suggested by the US Supreme Court, requiring "that each citizen have an *equally effective* voice in the election" (cf. *Reynolds v. Sims*, 377 U.S. 533, 1964, p. 565; emphasis added by the authors). Here, we operationalize equal efficacy or influence by comparing the a priori probabilities of individual voters being decisive or pivotal for the collective decision. The corresponding joint event of (i) a given voter determining her delegate's vote and of (ii) this representative determining the assembly's collective decision is admittedly a rare one. Still, while all being close to zero, the resulting individual *pivot probabilities* can vary widely across constituencies when weights are chosen arbitrarily. They should not if an institutional designer wants to fix voting weights (or bloc sizes) which are fair at least from behind the constitutional 'veil of ignorance' – that is, when preference patterns of the day are ignored for practical or normative reasons.

The objective of equalizing the a priori influence of each citizen on collective decisions was first formally considered by Lionel S. Penrose in 1946, when the institutional design of a successor to the League of Nations – today's United Nations Organization (UNO) – was being discussed.[2] Penrose (1946) argued that the most intuitive solution to the weight allocation problem, i.e., weights proportional to constituency sizes, ignores "elementary statistics of majority voting". Namely, if there are only two policy alternatives ('yes' and 'no') and all individual decisions are statistically *independent and equiprobable* then the probability of an individual voter

[2]Informal investigations date back to anti-federalist writings by Luther Martin, a delegate from Maryland to the Constitutional Convention in Philadelphia in 1787. See Riker (1986).

being pivotal in her constituency with n_i voters, which for odd n_i corresponds to the probability of $n_i - 1$ voters being divided into 'yes' and 'no'-camps of same size, is approximately $\sqrt{2}/\sqrt{\pi n_i}$ (apply Stirling's formula when evaluating the binomial distribution function). So a voter from a constituency C_i which is four times larger than constituency C_j a priori faces a smaller probability of tipping the scales locally; but this probability is still half rather than only a quarter of the reference level. Consequently, top-tier voting weights should be such that the pivot probability of constituency C_i at the top tier is twice – not four times – that of C_j in order to equalize the indirect influence of all citizens.

The corresponding practical suggestion is also known as the *Penrose square root rule*. Despite the straightforward criticism that it treats voting decisions like coin tosses, the rule has provided a benchmark for numerous applied studies which consider the distribution of voting power in the US, EU, or IMF (including Felsenthal and Machover 2001, 2004; Grofman and Feld 2005; Fidrmuc et al. 2009; Leech and Leech 2009; Miller 2009, 2012; Kirsch and Langner 2011). And though practitioners may not care about Penrose's reasoning itself – for instance, when the EU's heads of state and government bargained on new, post-2017 voting rules for the Council – they have invoked Penrose's suggestion when it fitted their interests.[3]

The special role of square root weight allocation rules has been confirmed, qualified, and disputed in a number of studies on two-tier voting systems, both empirically (see Gelman et al. 2002; 2004) and theoretically. The respective constitutional objective functions and practical conclusions of these investigations vary. Besides the equalization of a priori influence (Chamberlain and Rothschild 1981; Felsenthal and Machover 1998; Laruelle and Valenciano 2008b; Kaniovski 2008), they consider utilitarian welfare maximization (e.g., Beisbart et al. 2005; Barberà and Jackson 2006; Beisbart and Bovens 2007; Laruelle and Valenciano 2008b; Koriyama et al. 2012) and the avoidance of majoritarian paradoxes like having a Bush majority in the 2000 Electoral College despite a Gore majority in the population at large (Felsenthal and Machover 1999; Kirsch 2007; Feix et al. 2008). Several departures from Penrose's independence and equiprobability assumptions have been considered in these studies. However, the literature has focused almost entirely on *binary* political decisions, with no scope for bargaining and strategic interaction.[4]

Existing results hence provide useful arguments in the thorny debate on the 'right'

[3]A particularly notorious case involved the then Polish president and prime minister in the negotiations of the Treaty of Lisbon. See, e.g., *The Economist* (2007, June 14th).

[4]We are aware of the following exceptions only: Laruelle and Valenciano (2008a) suggest a "neutral" top-tier voting rule when policy alternatives give rise to a Nash bargaining problem. Le Breton et al. (2012) investigate fair voting weights in case of the division of a transferable surplus, i.e., for a simplex of policy alternatives. Maaser and Napel (2007; 2012a; 2012b) report Monte Carlo simulation results for influence-based, majoritarian, and welfarist objective functions in the median voter environment which we will here investigate analytically.

weight allocation only to the extent that the assemblies in question indeed decide on dichotomous exogenous proposals. But many decisions involve several shades of grey. Members of the US Electoral College usually have binary options, but they face the survivors from a much larger field of initial contenders, with endogenous final political platforms. The EU Council more commonly decides on the level of subsidies, the scope of regulation, the scale of financial aid, etc. rather than on having a subsidy, regulation of an industry, or aid *per se*. It seems relevant, therefore, to analyze the fair choice of voting weights (and alternative objectives such as utilitarian welfare) for somewhat richer than binary policy spaces.

This paper considers the equalization of pivot probabilities for a one-dimensional convex policy space, i.e., for choices from a real *interval*. We assume single-peaked preferences with random ideal points for all voters, perfect congruence between preferences of a constituency's delegate and its *median voter*, and collective decisions which correspond to the *Condorcet winner* or the *core* of the cooperative game defined by preferences and weights of the delegates. The latter can be seen as a proxy for the equilibrium outcome of strategic bargaining (see, e.g., Banks and Duggan 2000).

In this model, the collective choice equals the *weighted median* among agents whose ideal points themselves are medians from disjoint samples. This is a very stylized representation of democratic decision making but yet richer than the binary model à la Penrose. The former nests the latter in case that ideal points have a discrete two-point distribution. In case of less trivial distributions, little can analytically be said about the order statistics of medians from *differently* sized samples; and next to nothing has so far been known about the combinatorial function therefrom which corresponds to the respective weighted median of non-identically distributed random variables.

We derive a general analytical result on the ratio of two delegates' pivot probabilities in an infinite increasing chain of collective decision bodies (Theorem 1). Each delegate i is characterized by his voting weight w_i and single-peaked preferences with a random ideal point λ_i that has the probability density function f_i. In line with the veil of ignorance perspective of constitutional design, this random variable is a priori assumed to have the same theoretical median M for all delegates. It is shown then that, under suitable regularity conditions, delegate i's probability of being pivotal in the sense that his ideal point constitutes the corresponding voting game's Condorcet winner is *asymptotically proportional to the probability density* f_i *at the expected median position* M *times his assigned voting weight* w_i.

This main analytical result has several practical corollaries for two-tier voting systems. In particular, if all individual voters are – behind a constitutional veil of ignorance – conceived of as having ideal points that are *independent and identically distributed (i.i.d.)*, then the sample median from a constituency C_i with n_i members is asymptotically normal with a standard deviation inversely proportional to the *square* *root* of n_i. The probability density of representative i's ideal point at the expected median position M is hence proportional to $\sqrt{n_i}$. It follows that voting weights w_i proportional to $\sqrt{n_i}$ render the top-tier pivot probabilities proportional to population sizes; this approximately equalizes the expected influence or efficacy of the vote across the population.[5]

The optimality of a square root allocation of voting weights is, however, restricted to the case of individual ideal points being i.i.d. and the use of a 50%-majority threshold. Assuming that voters' ideal points are subject to identical random shocks within constituencies – which introduces plausible positive correlation among members of the same constituency – implies greater variance of the respective sample medians. The latter's distributions become more and more similar across constituencies if, behind a veil of ignorance, the shock's distribution H is a priori identical for all constituencies and its variance σ_H^2 increases. If this measure σ_H^2 of *heterogeneity across constituencies* is sufficiently great relative to the *heterogeneity within each constituency*, which is captured by the variance σ_G^2 of the (conditional) ideal point distribution G under a zero shock, then an approximately *linear* weight allocation becomes optimal.

That a linear weighting rule is optimal in this case can be extended from simple majority to supermajority requirements. In particular, one can approximate the pivot probabilities of top-tier delegates by the *Shapley value* of the respective weighted voting game when the preference polarization across constituencies, as measured by σ_H^2/σ_G^2, is sufficiently large (Theorem 2). Even if the number of constituencies m is relatively small, one can hence achieve equal representation by finding voting weights such that the resulting Shapley value is proportional to population sizes, or as close to being proportional as is feasible.

The remainder of the paper is organized as follows. In Section 2, we spell out our model of two-tier decision making and the institutional design problem. Our main result for simple majority rule and $m \to \infty$, as well as its implications in case that individual ideal points are i.i.d. are presented in Section 3. We then explore the effect of adding heterogeneity across constituencies to that within, and study asymptotic behavior with respect to the 'across'-kind for fixed m in Section 4. We conclude in Section 5 and provide proofs of the two theorems in an appendix.

---

[5]The approximation can be improved if one bases the weight choice on the induced *Shapley value* or, with comparable effects, the *Penrose-Banzhaf power index*. See Dubey and Shapley (1979), Felsenthal and Machover (1998) or Laruelle and Valenciano (2008b) for good overviews on these and other power measures.

2 Model and Design Problem

We consider partitions $\mathfrak{C}^m = \{C_1, \ldots, C_m\}$ of a large number n of voters into $m < n$ disjoint *constituencies* with $n_i = |C_i| > 0$ members each. The preferences of any voter $l \in \{1, \ldots, n\} = \bigcup_i C_i$ are assumed to be single-peaked with *ideal point* ν^l in a convex one-dimensional *policy space* $X \subseteq \mathbb{R}$, i.e., in a finite or infinite real interval. These ideal points are conceived of as realizations of random variables with a priori identical, absolutely continuous distributions. A given profile $(\nu^1, \ldots, \nu^n)$ of ideal points could reflect voter preferences in an abstract left–right spectrum or on a specific one-dimensional policy issue (a transfer, an emission standard, a capital requirement, etc.).

A collective decision $x^* \in X$ on the issue at hand is taken by an assembly or council of representatives $\mathcal{R}^m$ which consists of one representative from each constituency. Without committing to any particular procedure for internal preference aggregation, political competition, lobbying or bargaining, it will be assumed that preferences of C_i's representative coincide with those of its respective median voter, i.e., representative i has the random ideal point

$$\lambda_i \equiv \text{median}\,\{\nu^l : l \in C_i\}. \tag{1}$$

For simplicity we take all n_i as odd numbers,[6] and leave aside agency problems or other reasons for why the preferences of a constituency's representative might not be congruent or at least sensitive to its median voter.[7]

In the top-tier assembly $\mathcal{R}^m$, constituency C_i has voting weight $w_i \geq 0$. Any coalition $S \subseteq \{1, \ldots, m\}$ of representatives which achieves a combined weight $\sum_{j \in S} w_j$ above

$$q^m \equiv 0.5 \sum_{j=1}^{m} w_j, \tag{2}$$

i.e., which has a *simple majority* of total weight, is winning and can pass proposals to implement some policy $x \in X$.

Let $\cdot\!:\!m$ be the random permutation of $\{1, \ldots, m\}$ that makes $\lambda_{k:m}$ the k-th leftmost ideal point among the representatives for any realization of $\lambda_1, \ldots, \lambda_m$ (that is, $\lambda_{k:m}$ is the k-th order statistic). We will disregard the zero probability events of two or more

[6] For an even number n_i, one could let each of the two middlemost ideal points in C_i define the representative i's preferences with equal probability. Or one works with the usual definition of the median, i.e., their arithmetic mean, and focuses on the probability of event $\{\partial x^*/\partial \nu^i > 0\}$ rather than the – no longer equivalent – event $\{x^* = \nu^i\}$ in what follows. Napel and Widgrén (2004) discuss in detail how *influence* in voting procedures can be quantified by outcome sensitivity measures like $\partial x^*/\partial \nu^i$.

[7] See, e.g., Gerber and Lewis (2004) for empirical evidence on how district median voters and partisan pressures jointly explain legislator preferences, and for a short discussion of the related theoretical literature. It is important to note that Theorems 1 and 2 will *not* require (1) to hold – they only assume λ_i's density to have certain properties.

constituencies having identical ideal points and define the random variable P by

$$P \equiv \min\Big\{j \in \{1,\dots,m\}\colon \sum_{k=1}^{j} w_{k:m} > q^m\Big\}. \tag{3}$$

Representative $P\colon m$'s ideal point, $\lambda_{P:m}$, cannot be beaten by any alternative $x \in X$ in a pairwise vote, i.e., it is in the *core* of the voting game defined by ideal points $\lambda_1, \dots, \lambda_m$, weights $w_1, \dots, w_m$ and quota q^m. We assume that the policy x^* agreed by $\mathcal{R}^m$ lies in the core. So x^* must equal $\lambda_{P:m}$ whenever the core is single-valued; then $\lambda_{P:m}$ actually beats every other alternative $x \in X$ and is the so-called *Condorcet winner* in $\mathcal{R}^m$. In order to avoid inessential case distinctions, we assume that $\mathcal{R}^m$ agrees on $\lambda_{P:m}$ also in the knife-edge case of the entire interval $[\lambda_{P-1:m}, \lambda_{P:m}]$ being majority-undominated, i.e.,[8]

$$x^* \equiv \lambda_{P:m}. \tag{4}$$

Representative $P\colon m$ will, therefore, generally be referred to as the *pivotal representative* or the *weighted median* of $\mathcal{R}^m$. Banks and Duggan (2000) and Cho and Duggan (2009) provide equilibrium analysis of non-cooperative legislative bargaining which supports policy outcomes inside or close to the core.

The event $\{x^* = \nu^l\}$ of voter l's ideal point coinciding with the collective decision almost surely entails that sufficiently small perturbations or idiosyncratic shifts of ν^l translate into identical shifts of x^*, so that $\partial x^*/\partial \nu^l > 0$. Voter l can then meaningfully be said to *influence*, be *decisive* or *pivotal* for, or even to *determine* the collective decision. This event has probability

$$p^l \equiv \Pr(x^* = \nu^l), \tag{5}$$

which depends on the joint distribution of $(\nu^1, \dots, \nu^n)$ and the voting weights $w_1, \dots, w_m$ that have been assigned in $\mathcal{R}^m$. Even though p^l will be very small given that the set of voters $\{1, \dots, n\}$ is assumed to be large, it would constitute a violation of the 'one person, one vote' principle if p^l/p^k differed substantially from unity for any $l, k \in \{1, \dots, n\}$.

We will assume throughout our analysis that voter ideal points are a priori *identically distributed*, in line with adopting a 'veil of ignorance'-perspective when one analyzes the efficacy of individual votes or a priori influence of voters. Moreover, it is assumed that ideal points are mutually *independent across constituencies*. We do, however, allow for a specific form of ideal points being *dependent within each*

[8] A sufficient condition for the core to be single-valued is that the vector of weights satisfies $\sum_{j\in S} w_j \neq q^m$ for each $S \subseteq \{1, \dots, m\}$. In the non-generic cases where this is violated, tie-breaking assumptions analogous to fn. 6 can be made. Note that no constituency's median voter will have an incentive to 'choose' a representative whose preferences differ from her own ones, that is, to misrepresent preferences, if x^* is determined by (4) (cf. Moulin 1980; Nehring and Puppe 2007).

constituency. Namely, we conceive of the ideal point ν^l of any voter $l \in C_i$ as the sum

$$\nu^l = \mu_i + \epsilon^l \tag{6}$$

of a constituency-specific random variable μ_i, which has distribution H, and a voter-specific random variable ϵ^l with absolutely continuous distribution G. Variables $\epsilon^1, \ldots, \epsilon^n$ and $\mu_1, \ldots, \mu_m$ are all taken to be mutually independent. Because G and H are the same for all voters $l \in \{1, \ldots, n\}$ and constituencies $C_i \in \mathfrak{C}^m$, ideal points $\nu^1, \ldots, \nu^n$ are indeed identically distributed.

The variance of G, σ_G^2, can be interpreted as a measure of *heterogeneity within each constituency*, reflecting the natural variation of political preferences. If distribution H of μ_i is non-degenerate, it reflects a common attitude component of preferences within each constituency. H's variance σ_H^2 is a measure of *heterogeneity across constituencies*: even though it is assumed that opinions in all constituencies vary between left–right, religious–secular, etc. in a similar manner, the locations of the respective ranges of opinion can differ between constituencies. The ideal points of two voters $l, k \in C_i$ from the same constituency hence are correlated, with a coefficient of $\sigma_H^2/(\sigma_H^2 + \sigma_G^2)$.

The case in which H is degenerate with $\sigma_H^2 = 0$ involves heterogeneity only within constituencies; the latter differ in size but voter ideal points ν^l are independent and identically distributed across the entire population. We regard this as a particularly important benchmark and will refer to it as the *i.i.d. case*.

With this notation, we can now state our objective of operationalizing the 'one person, one vote' principle more formally. Namely, given a partition $\mathfrak{C}^m = \{C_1, \ldots, C_m\}$ of n voters into constituencies and distributions G and H which describe heterogeneity of individual preferences within and across constituencies, we would like to find voting weights $w_1, \ldots, w_m$ such that each voter a priori has an equal chance of determining the collective decision $x^* \in X$ – that is, such that

$$\frac{p^l}{p^k} = 1 \textit{ for all } l, k \in \{1, \ldots, n\}. \tag{7}$$

For most combinations of $\mathfrak{C}^m$, G, and H, condition (7) cannot be satisfied by *any* weight vector $(w_1, \ldots, w_m)$. This is due to the discrete nature of weighted voting.[9] So the problem would need to be formulated more precisely as that of minimizing a specific notion of distance between $(1/n, \ldots, 1/n) \in \mathbb{R}^n$ and the probability vector $(p^1, \ldots, p^n)$ induced by $w_1, \ldots, w_m$.

Our actual concern, however, is *not* with finding the respective optimal solution to

[9] For instance, there are only 117 structurally different weighted voting games with $m = 5$ constituencies even if all majority thresholds between 0% and 100% are permitted. This number (related to *Dedekind's problem* in discrete mathematics) grows very fast, but the set of distinct feasible influence distributions remains finite.

such a (non-trivial) discrete minimization problem for a particular partition $\mathfrak{C}^m$ and specific distributions G and H. Rather, our objective is to find a *simple function* which maps $n_1, \ldots, n_m$ to weights $w_1, \ldots, w_m$ that induce $p^l/p^k \approx 1$ for all l and k, that is, which approximately satisfy the 'one person, one vote' criterion, for arbitrary partitions $\mathfrak{C}^m$.[10] Preferably, qualitative information on heterogeneity within and across constituencies should be sufficient to guide institutional design recommendations.

The stated assumptions imply that, when considering any given realization of μ_i, ideal points ν^l and ν^k are conditionally independent if $l, k \in C_i$ for some i. They are in any case identically distributed. In particular, $p^l = p^k$ holds for $l, k \in C_i$ irrespective of which G, H, and voting weights $w_1, \ldots, w_m$ are considered, and it must be the case that if $l \in C_i$ then

$$\Pr(\nu^l = \lambda_i) = \frac{1}{n_i}. \tag{8}$$

So, an individual voter's probability to be her constituency's median and to determine λ_i is *inversely proportional* to constituency C_i's population size.

The events $\{\nu^l = \lambda_i\}$ and $\{x^* = \lambda_i\}$ are independent given our statistical assumptions. (Note that the first event only entails information about the identity of C_i's median, not its location.) It follows that the probability p^l for an individual voter $l \in C_i$ influencing the collective decision x^* is $1/n_i$ times the probability of event $\{x^* = \lambda_i\}$ or, equivalently, of $\{P\colon m = i\}$. Letting

$$\pi_i(\mathcal{R}^m) \equiv \Pr(P\colon m = i) \tag{9}$$

denote the probability of constituency C_i's representative being pivotal in $\mathcal{R}^m$ (that is, of λ_i being the respective Condorcet winner in case of generic weights), our institutional design objective hence consists of solving the following

Problem of Equal Representation:
Find a simple mapping from constituency sizes $n_1, \ldots, n_m$ to voting weights $w_1, \ldots, w_m$ for the representatives in $\mathcal{R}^m$ such that

$$\frac{\pi_i(\mathcal{R}^m)}{\pi_j(\mathcal{R}^m)} \approx \frac{n_i}{n_j} \text{ for all } i, j \in \{1, \ldots, m\}. \tag{10}$$

One might conjecture that, if m is large enough and the weight distribution is not overly skewed, voting weight w_i should translate linearly into representative i's influence $\pi_i(\mathcal{R}^m)$.[11] But the distribution of the respective ideal points λ_i will certainly play a

[10] It is not necessary to specify exactly which functions are "simple" enough. *Power laws*, i.e., choosing $w_i = \beta n_i^\alpha$ for some $\alpha, \beta \in \mathbb{R}$, certainly qualify and turn out to constitute a sufficiently rich class of mappings.

[11] Asymptotic proportionality between weights and voting power has first been investigated by Penrose (1952) in the context of binary alternatives. Related formal results by Neyman (1982), Lindner and Machover (2004), Snyder et al. (2005), Jelnov and Tauman (2012) and Theorem 1 below suppose

role, too, and so the solution to this problem will depend on how heterogeneity of individual preferences within and across constituencies relate.

Note that, *if* the representatives' ideal points $\lambda_1, \ldots, \lambda_m$ were not only mutually independent but also had identical distributions $F_i = F_j$ for all $i, j \in \{1, \ldots, m\}$, then all orderings of $\lambda_1, \ldots, \lambda_m$ would a priori be equally likely. In this case, $\pi_i(\mathcal{R}^m)$ would coincide with i's *Shapley value* $\phi_i(v)$, where v is the characteristic function of the m-player cooperative game in which the worth $v(S)$ of a coalition $S \subseteq \{1, \ldots, m\}$ is 1 if $\sum_{j\in S} w_j > q^m$ and 0 otherwise, and[12]

$$\phi_i(v) \equiv \sum_{S\subseteq\{1,\ldots,m\}\setminus\{i\}} \frac{|S|!\cdot(m-|S|-1)!}{m!}[v(S\cup\{i\}) - v(S)]. \tag{11}$$

The way to solve the problem of equal representation would then simply be to search for a weighted voting game which induces a Shapley value proportional to $(n_1, \ldots, n_m)$. Unfortunately, if we assume that distribution G, which generates the private component ϵ^l in individual ideal points ν^l, is non-degenerate then (1) implies that $F_i = F_j$ if and only if $n_i = n_j$. The case in which this holds for all $i, j \in \{1, \ldots, m\}$ is the one in which equal representation is entirely trivial, i.e., achieved by giving all representatives identical weights because $n_1 = \ldots = n_m$.[13] In particular, F_j is a mean-preserving spread of F_i (i.e., F_i second-order stochastically dominates distribution F_j) if $n_i > n_j$ because the sample median of n_i independent draws from G has smaller variance than that of just $n_j < n_i$ draws; and the respective draw from H adds identical variance to both λ_i and λ_j.

In the i.i.d. benchmark case, in which $\sigma_H^2 = 0$ and the only acknowledged differences between two voters from distinct constituencies are the numbers of their fellow

that the *relative* weight of any given voter becomes negligible as more and more voters are added. The case when relative weights of a few large voters fail to vanish as $m \to \infty$ – giving rise to *oceanic games* and typically non-proportionality – has been treated by Shapiro and Shapley (1978) and Dubey and Shapley (1979). The limit behavior of pivot probabilities for *uniform* weights (as at the bottom tier) has been studied in more complex models than Penrose's by Chamberlain and Rothschild (1981), Myerson (2000), Gelman et al. (2002), and Kaniovski (2008).

[12]See Shapley (1953). For so-called *simple games* – in which $v(S) \in \{0,1\}$ for all $S \subseteq N$, $v(\emptyset) = 0$, $v(N) = 1$, and $v(S) = 1 \Rightarrow v(T) = 1$ if $S \subseteq T$ – the Shapley value is also referred to as the *Shapley-Shubik power index*, following the first suggestion of using ϕ in order to evaluate power in voting bodies by Shapley and Shubik (1954). We write $v = [q^m; w_1, \ldots, w_m]$ if v is defined by the weighted voting rule $[q^m; w_1, \ldots, w_m]$.

[13]We remark that re-partitioning the population into constituencies of equal size – i.e., appropriate redistricting – is, of course, a trivial possibility for altogether evading the considered problem. Our analysis is concerned with those cases where historical, geographical, cultural, and other reasons exogenously have defined a partition $\mathfrak{C}^m$ which cannot easily be changed. See Coate and Knight (2007) on socially optimal districting and Gul and Pesendorfer (2010) on strategic issues which arise for redistricting. We also disregard another relevant strategic feature of two-tier voting: incentives to allocate limited campaign resources to the constituencies. We refer the reader to Strömberg (2008).

constituents, one can be more specific than stochastic dominance. Namely, a standard result about the sample median of i.i.d. random variables is:

Lemma 1. *Let* $X_1, \ldots, X_s$ *be i.i.d. random variables with median* M *and with a density* f *that is continuous at* M *with* $f(\mathrm{M}) > 0$*. Then random variable* $Y = \mathrm{median}\{X_1, \ldots, X_s\}$ *is asymptotically* (M, σ^2)*-normally-distributed with*

$$\sigma^2 = \frac{1}{s\,[2f(\mathrm{M})]^2} \tag{12}$$

i.e., the re-scaled sample median $2f(\mathrm{M})\,\sqrt{s}[Y - \mathrm{M}]$ *of* $X_1, \ldots, X_s$ *converges in distribution to* $\mathbf{N}(0,1)$ *as* $s \to \infty$*.*

See, e.g., Arnold et al. (1992, Theorem 8.5.1) for a proof. It follows that in the i.i.d. case, ideal points $\lambda_1, \ldots, \lambda_m$ in assembly $\mathcal{R}^m$ are (approximately) normally distributed[14] with identical means but standard deviations that are *inversely proportional to the square root* of the respective constituency sizes. So, in the i.i.d. case, rather than all orderings of $\lambda_1, \ldots, \lambda_m$ being equally likely ($\phi(v)$'s implicit assertion), the representative of a constituency C_i which is four times larger than constituency C_j has twice the chances to find itself in the middle. (Recall that a normal distribution's density at the mean and median is inversely proportional to its standard deviation.) Then weights that are proportional to population sizes, or weights such that $\phi(v)$ is, would give representatives of large constituencies more a priori influence than is due.[15]

Before we make this reasoning precise in the following section, let us iterate that the considered median voter model of equal representation in two-tier decision making is an admittedly big simplification. Many collective decisions involve more than just a single dimension in which voter preferences differ. We ignore that voting might involve private information about some state variable (Feddersen and Pesendorfer 1996, 1997; Bouton and Castanheira 2012), and typical agency problems connected to imperfect monitoring and infrequent delegate selections. Empirical evidence highlights that a representative may take positions that differ significantly from his district's median when voter preferences within that district are sufficiently

[14]This approximation is very good already for rather moderate sample sizes. If, e.g., individual ideal points ν^l are standard uniformly distributed, i.e., $\epsilon^l \sim \mathbf{U}[0,1]$ and $\mu_i \equiv 0$, then λ_i is beta distributed with parameters $a = b = (n_i + 1)/2$. The corresponding beta and normal density functions can be regarded as identical for all practical purposes if $n_i > 100$. – Note that Lemma 1 is useful also in case of a non-degenerate distribution H of μ_i: it establishes that the precise distribution G of individual shocks does not matter for λ_i's distribution F_i; only $g(\mathrm{M})$, the sufficiently great n_i, and H do.

[15]We are unaware of any systematic empirical evidence for or against the claim that representatives from larger constituencies tend to be located more centrally in the relevant policy space. This theoretical prediction is testable in principle but rests on the two assumptions of aggregate preferences being determined by the median individual *and* individual ideal points being i.i.d. In view of the rapid transition towards indistinguishable distributions of representative ideal points when the i.i.d. assumption is given up (see Section 4), such claim is bound to be difficult to confirm in practice.

heterogeneous (see, e.g., Gerber and Lewis 2004). Still, we take it that the best intuitions about fairness are captured by simplifying thought experiments of a 'veil of ignorance' kind. The analysis of the described stylized world – no friction, particularly well-behaved preferences which are a priori identical for all – is useful in this way. It shows the limitations of and justifications for the simple intuition that weights should be proportional to the number of represented constituents, in a framework that goes beyond the binary world analyzed by Penrose (1946) and others.

3 Egalitarian Voting Weights for Many Constituencies

We will in this section consider situations in which the number m of constituencies is suitably large. Very few tangible results exist on the distribution of order statistics, like the median, from *differently* distributed random variables (the representative ideal points $\lambda_1, \ldots, \lambda_m$). And almost nothing seems to be known about the respective distribution of a *weighted median*, which is taken to define the collective decision $x^* \in X$ in our model. It turns out to be possible, nevertheless, to characterize the probability of some λ_i being the weighted median, i.e., the pivot probability $\pi_i(\mathcal{R}^m)$, as $m \to \infty$.

We conceive of $\mathcal{R}^1 \subset \mathcal{R}^2 \subset \mathcal{R}^3 \subset \ldots$ as an infinite chain of assemblies in which more and more constituencies $i \in \mathbb{N}$ have a representative with a voting weight $w_i \geq 0$ and a random ideal point λ_i with absolutely continuous distribution F_i. Some technical requirements will be imposed on the corresponding density f_i, but it is immaterial whether λ_i corresponds to the median of some set of other random ideal points, as formalized by definition (1); it could, alternatively, be an average of some ideal points (such as those of members of a coalition government or oligarchy) or reflect the interests of a constituency dictator. So while the problem of equal representation which we stated in Section 3 is the key motivation for investigating pivot probabilities $\pi_1(\mathcal{R}^m), \ldots, \pi_m(\mathcal{R}^m)$, the following characterization of their limiting behavior has more general applicability.[16] We will return to the specific issue of designing egalitarian two-tier voting systems after considering assemblies $\mathcal{R}^m$ with arbitrary ideal point distributions $F_1, \ldots, F_m$ and weighted voting rules $[q^m; w_1, \ldots, w_m]$.

For weight sequences $\{\boldsymbol{w}^m\}_{m \in \mathbb{N}}$ and associated weighted voting games $[q^m; w_1, \ldots, w_m]$ in which everyone's *relative* voting weight goes to zero as $m \to \infty$, the pivot probability $\pi_i(\mathcal{R}^m)$ of any given representative $i \in \mathbb{N}$ will converge to zero.[17] Still, $\pi_i(\mathcal{R}^m)/\pi_j(\mathcal{R}^m)$

[16] In particular, for given $F_1, \ldots, F_m$, $\pi(\mathcal{R}^m)$ amounts to a specific *quasivalue* or *random order value* for simple games. See, e.g., Monderer and Samet (2002).

[17] See Neyman (1982, Lemma 8.2) for an upper bound on the speed of convergence when $F_1, \ldots, F_m$ are identical.

need not converge. This is illustrated by the sequence $\{\boldsymbol{w}^m\}_{m\in\mathbb{N}}$ with

$$\boldsymbol{w}^m = (1, 2, \ldots, 2) \in \mathbb{R}^m. \tag{13}$$

Representative 1 is either a dummy player with $\pi_1(\mathcal{R}^m) = 0$ or, supposing that the ideal point distributions $F_1, \ldots, F_m$ are identical, $\pi_i(\mathcal{R}^m) = \frac{1}{m}$ for *all* $i = 1, \ldots, m$ depending on whether m is odd or even. So $\pi_1(\mathcal{R}^m)/\pi_2(\mathcal{R}^m)$ alternates between 0 and 1. More complicated examples of non-convergence can be constructed, e.g., by having $\{\boldsymbol{w}^m\}_{m\in\mathbb{N}}$ oscillate in a suitable fashion.

We rule out such possibilities by imposing a weak form of replica structure on the considered weights $w_1, w_2, w_3, \ldots$ and ideal point distributions $F_1, F_2, F_3, \ldots$ Specifically, we require that all representatives $i \in \mathbb{N}$ belong to one of an arbitrary but *finite number of representative types* $\theta \in \{1, \ldots, r\}$. All representatives of the same type θ have an identical weight and ideal point distribution; that is, there exists a mapping $\tau\colon \mathbb{N} \to \{1, \ldots, r\}$ such that $\tau(j) = \theta$ implies that λ_j has density f_θ and $w_j = w_\theta \geq 0$. We presume w.l.o.g. that $\tau(i) = i$ for $i \in \{1, \ldots, r\}$. And, avoiding somewhat contrived situations like in (13), we restrict attention to chains $\mathcal{R}^1 \subset \mathcal{R}^2 \subset \mathcal{R}^3 \subset \ldots$ in which each type θ maintains a *non-vanishing share*

$$\beta_\theta(m) \equiv |\{k \in \{1, \ldots, m\}\colon \tau(k) = \theta\}|/m \tag{14}$$

of representatives as $m \to \infty$. In particular, we presume the existence of $\beta > 0$ and $m^0 \in \mathbb{N}$ such that for all $m \geq m^0$ we have $\beta_\theta(m) \geq \beta > 0$.

The key requirements for the following result then are that (i) $F_1, F_2, F_3 \ldots$ have identical median M and that (ii) each distribution F_i has a density f_i which is locally continuous and positive at M. In order to allow the application of a powerful uniform convergence result by Neyman (1982) for the Shapley value, continuity will be strengthened to the requirement that each density f_i's variation at its median, $|f_i(x) - f_i(\mathrm{M})|$, can locally be bounded by a quadratic function cx^2. This bound follows readily if f_i is C^2 like the normal density functions singled out by Lemma 1, and could be relaxed to cx^a for any $a > 1$ if one used somewhat less round constants in the proof. Moreover, an unpublished extension by Abraham Neyman of his 1982 result could be employed in order to make do with just (ii). Details on this and the proof are presented in Appendix A.

Theorem 1. *Consider a finite number r of representative types and an infinite chain $\mathcal{R}^1 \subset \mathcal{R}^2 \subset \mathcal{R}^3 \subset \ldots$ of assemblies in which each type $\theta \in \{1, \ldots, r\}$ maintains a non-vanishing share. If for each type θ the ideal point distribution F_θ has median* M *and its density f_θ*

satisfies $f_\theta(\mathrm{M}) > 0$ *with* $|f_\theta(x) - f_\theta(\mathrm{M})| \le cx^2$ *for some* $c \ge 0$ *in a neighborhood of* M *then*

$$\lim_{m\to\infty} \frac{\pi_i(\mathcal{R}^m)}{\pi_j(\mathcal{R}^m)} = \frac{w_i f_i(\mathrm{M})}{w_j f_j(\mathrm{M})} \tag{15}$$

supposing that $w_j > 0$*.*

The key observation behind Theorem 1 is that, as m grows large, the pivotal member of $\mathcal{R}^m$ is most likely found very close to the common median M of ideal point distributions $F_1, \ldots, F_m$. Pivotality at location $x \in X$ requires that less than half the total weight of $\mathcal{R}^m$'s members is located in $(-\infty, x)$ and less than half the total weight is found in (x, ∞). In expectation, this occurs exactly at $x = \mathrm{M}$, and the probability for the realized weighted median in $\mathcal{R}^m$ to fall outside an ε-neighborhood of M turns out to approach zero exponentially fast as $m \to \infty$.

One can, therefore, restrict attention in the proof to an arbitrarily small interval $[-\varepsilon, \varepsilon]$ when m is sufficiently large and, w.l.o.g., $\mathrm{M} = 0$. The locally continuous densities $f_1(x), \ldots, f_m(x)$ can suitably be approximated by upper and lower bounds on this interval. Moreover, when we condition on the respective events $\{\lambda_j \in [-\varepsilon, \varepsilon]\}$, the bounds are almost identical for any $j = 1, \ldots, m$ when m is large. This makes all orderings of those representatives with ideal points in $[-\varepsilon, \varepsilon]$ conditionally equiprobable in very good approximation. Representative i's respective conditional pivot probability, therefore, corresponds to i's Shapley value in a 'subgame' which involves only the representatives j with realizations $\lambda_j \in [-\varepsilon, \varepsilon]$. It is possible to apply the uniform convergence result for the Shapley value proven by Neyman (1982) to each of these subgames. In the final step of the proof, it then remains to exploit that the probability of the condition $\{\lambda_i \in [-\varepsilon, \varepsilon]\}$ being true becomes proportional to λ_i's density at 0 when $\varepsilon \downarrow 0$.[18]

Theorem 1 provides a rather general answer to the posed problem of equal representation in the case that many constituencies are involved. In particular, comparison of equations (10) and (15) immediately yields

Corollary 1. *If m is sufficiently large then choosing*

$$(w_1, \ldots, w_m) \propto \left(\frac{n_1}{f_1(\mathrm{M})}, \ldots, \frac{n_m}{f_m(\mathrm{M})}\right) \tag{16}$$

achieves approximately equal representation (as formalized by condition (10)) if the technical

[18]At an intuitive level, one may even directly think of the limit case $\varepsilon = 0$: if one conditions on representative i's ideal point being located at $x = \mathrm{M}$, then each representative $j \neq i$ is equally likely found to i's left or right (because $F_j(\mathrm{M}) = \frac{1}{2}$). In this case, i's conditional pivot probability equals i's Penrose-Banzhaf power index, which, like the Shapley value, becomes proportional to $(w_1, \ldots, w_m)$ for the replica-like weight sequences that we consider. (See Lindner and Machover 2004 and Lindner and Owen 2007 on the corresponding limit result.)

constraints in Theorem 1 are satisfied by $f_1, \dots, f_m$.

If $m \ll \infty$, the approximation of the conditional pivot probabilities for ideal points in a neighborhood of the common median M, which is obtained by (a) considering the limit case of orderings being conditionally equiprobable and (b) by applying Neyman's limit result for the Shapley value, need not be very good. The latter source of imprecision can be avoided by computing the Shapley value $\phi(v)$ for the simple game $v = [q^m; w_1, \dots, w_m]$ which is defined by representatives' weights and simple majority rule as described in Section 2. The suggestion in Corollary 1 can hence be improved somewhat if (16) is replaced by

$$(w_1, \dots, w_m) \text{ such that } \phi(q^m; w_1, \dots, w_m) \propto \left(\frac{n_1}{f_1(\mathrm{M})}, \dots, \frac{n_m}{f_m(\mathrm{M})} \right). \tag{17}$$

We conclude this section by specifically considering the benchmark i.i.d. case, in which the ideal points $\nu^1, \dots, \nu^n$ correspond just to voter-specific random variables $\epsilon^1, \dots, \epsilon^n$ that are independent and identically distributed with a suitable probability density function g, and where λ_i corresponds to the median ideal point in C_i. In this case, the ideal points $\lambda_1, \dots, \lambda_m$ are asymptotically normally distributed by Lemma 1 with densities f_i that satisfy the quadratic bound condition of Theorem 1 and

$$f_i(\mathrm{M}) = \frac{1}{\sqrt{2\pi \cdot \frac{1}{n_i [2g(\mathrm{M})]^2}}} = \frac{2g(\mathrm{M})}{\sqrt{2\pi}} \sqrt{n_i} > 0. \tag{18}$$

Combining equation (18) and Corollary 1 we obtain:

Corollary 2 (Square root rule)**.** *If the ideal points of all voters are i.i.d., representative i's ideal point equals the median voter's ideal point in constituency C_i for all $i \in \{1, \dots, m\}$, and m is sufficiently large then*

$$(w_1, \dots, w_m) \propto \left(\sqrt{n_1}, \dots, \sqrt{n_m} \right) \tag{19}$$

or, better,

$$(w_1, \dots, w_m) \text{ such that } \phi(q^m; w_1, \dots, w_m) \propto \left(\sqrt{n_1}, \dots, \sqrt{n_m} \right) \tag{20}$$

achieves approximately equal representation.

4 Heterogeneity within vs. across constituencies

Corollary 2 derived a *square root* rule similar to that of Penrose (1946) for the i.i.d. case,[19] that is, for a degenerate distribution H of the constituency-specific μ_i-components of individual ideal points $\nu^l = \mu_i + \epsilon^l$. We now investigate the robustness of this rule regarding the degree of preference affiliation within each constituency. Non-degenerate shocks μ_i imply positive correlation within each constituency and give rise to *polarization of preferences* along constituency lines, which is measured by the ratio σ_H^2/σ_G^2.[20] It turns out that for sufficiently strong polarization, a *linear* weight allocation rule quickly performs better than strictly concave mappings

This is analytically seen most easily for the case in which all the involved distributions are normal. First, let all ϵ^l be distributed normally with zero mean and variance σ_G^2. Lemma 1 then implies that the median of $\{\epsilon^l\}_{l\in C_i}$ is approximately normally distributed with zero mean and variance $\pi\sigma_G^2/(2n_i)$. Second, let the constituency-specific preference component μ_i be normally distributed with zero mean and variance σ_H^2. Constituency C_i's aggregate ideal point λ_i then also has an approximately normal distribution with zero mean and variance $\pi\sigma_G^2/(2n_i) + \sigma_H^2$. Considering the corresponding densities at M = 0 for two representatives i and j yields

$$\frac{f_i(0)}{f_j(0)} = \left(\frac{\frac{\pi\sigma_G^2}{2n_i} + \sigma_H^2}{\frac{\pi\sigma_G^2}{2n_j} + \sigma_H^2} \right)^{-\frac{1}{2}}. \tag{21}$$

This ratio quickly approaches 1 as $\sigma_H^2 \to \infty$, or if $\sigma_H^2 > 0$ and $n_i, n_j \to \infty$. Corollary 1 then calls for $(w_1, \ldots, w_m) \propto (n_1, \ldots, n_m)$.

We pointed out in Section 3 that heterogeneity within each constituency will always give rise to different distributions of the sample medians when $n_i \neq n_j$. But the differences become small and no longer matter for pivotality in $\mathcal{R}^m$ when the heterogeneity across constituencies is sufficiently great. This is illustrated by Figure 1. It depicts the density functions of ideal points λ_i and λ_j when C_i is four times larger than constituency C_j, so that the standard deviation σ_i of the median of $\{\epsilon^l\}_{l\in C_i}$ is half the standard deviation $\sigma_j = \sigma$ of the median of $\{\epsilon^l\}_{l\in C_j}$. Panel (a) shows the densities when $\sigma_H^2 = 0$ (or when we condition on $\mu_i = \mu_j = 0$); panel (b) depicts the case when $\mu_i, \mu_j \sim \mathbf{U}[-6\sigma, 6\sigma]$. The densities in panel (b) are almost indistinguishable in a neighborhood of the median. This neighborhood's size increases in σ_H^2, and it coincides with the relevant policy range in which the Condorcet winner of $\mathcal{R}^m$ is most likely located under simple majority rule.

[19] Note that the *Penrose square root rule* does not refer to weights but top-tier pivot probabilities, which equal the Penrose-Banzhaf power index of the representatives in Penrose's binomial voting model (cf. fn. 5).

[20] The basic features of *polarization* are according to Esteban and Ray (1994, p. 824): (i) a high degree of homogeneity *within* groups, (ii) a high degree of heterogeneity *across* groups, and (iii) a small number of significantly sized groups. Ratio σ_H^2/σ_G^2 serves as a simple measure of polarization of ideal points $\nu^1, \ldots, \nu^n$ here, where groups are given exogenously. Esteban and Ray characterize polarization measures for the general case without an exogenous partition of the population.

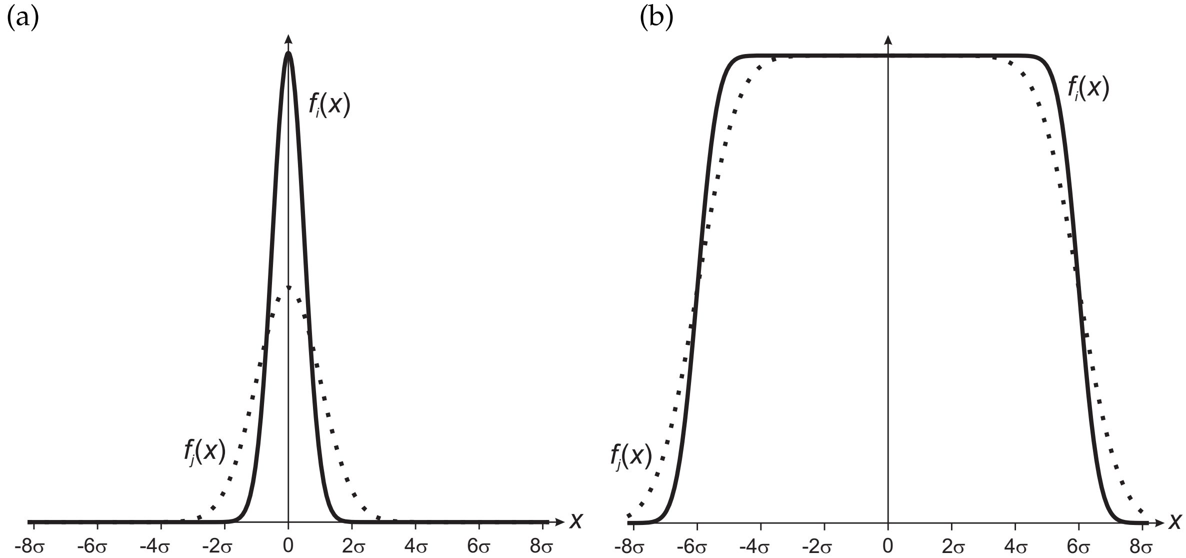

Figure 1: Densities of λ_i and λ_j when $n_i = 4n_j$ and (a) $\mu_i = \mu_j = 0$ or (b) $\mu_i, \mu_j \sim \mathbf{U}[-6\sigma, 6\sigma]$

Figure 1(b) reflects a preference dissimilarity or polarization ratio of $\sigma_H^2/\sigma_G^2 = 6\pi/n_j$.[21] This is tiny when one thinks of typical real-world population figures n_j, and suggests that the phase transition between optimality of a square root rule to optimality of a linear rule can be very fast. Figure 2 demonstrates that. The two panels consider (a) the US population's partition into 50 states and the District of Columbia, and (b) the current European Union with 27 member states (EU27). The dashed lines illustrate the (interpolated) optimal coefficients α^* as a function of σ_H^2/σ_G^2 when we search for the best rule in the class

$$(w_1, \ldots, w_m) \propto ({n_1}^{\alpha}, \ldots, {n_m}^{\alpha}) \tag{22}$$

for $\alpha \in \{0, 0.01, \ldots, 1.99, 2\}$;[22] the solid lines analogously depict α^* when one searches within the class of Shapley value-based rules

$$(w_1, \ldots, w_m) \text{ such that } \phi(q^m; w_1, \ldots, w_m) \propto ({n_1}^{\alpha}, \ldots, {n_m}^{\alpha})\,. \tag{23}$$

Optimality of the square root rule can be seen to break down quickly; already small degrees of preference dissimilarity across constituencies render a linear rule based on the Shapley value optimal.[23] This makes it possible to base institutional design

[21] Recall that the uniform distribution on $[a, b]$ has variance $(b-a)^2/12$; so here $\sigma_H^2 = 12\sigma^2$. If all ϵ^l are normal with variance σ_G^2 then $\sigma_j = \sigma$ corresponds to $\sigma_G^2 = (2n_j/\pi) \cdot \sigma^2$.

[22] Specifically, we consider $\epsilon^l \sim \mathbf{U}[-0.5, 0.5]$ and $\mu_i \sim \mathbf{N}(0, \sigma_H^2)$ with $0 \leq \sigma_H^2 \leq 10^{-6}$ and determine estimates of the pivot probabilities $\pi_i(\mathcal{R}^{27})$ which are induced by a given value of α via Monte Carlo simulation. The considered objective is to minimize $\|\cdot\|_1$-distance between individual pivot probabilities and the egalitarian ideal of $(1/n, \ldots, 1/n) \in \mathbb{R}^n$.

[23] That α^* fails to converge to 1 when the simpler weight-based rule in (22) is concerned attests to the

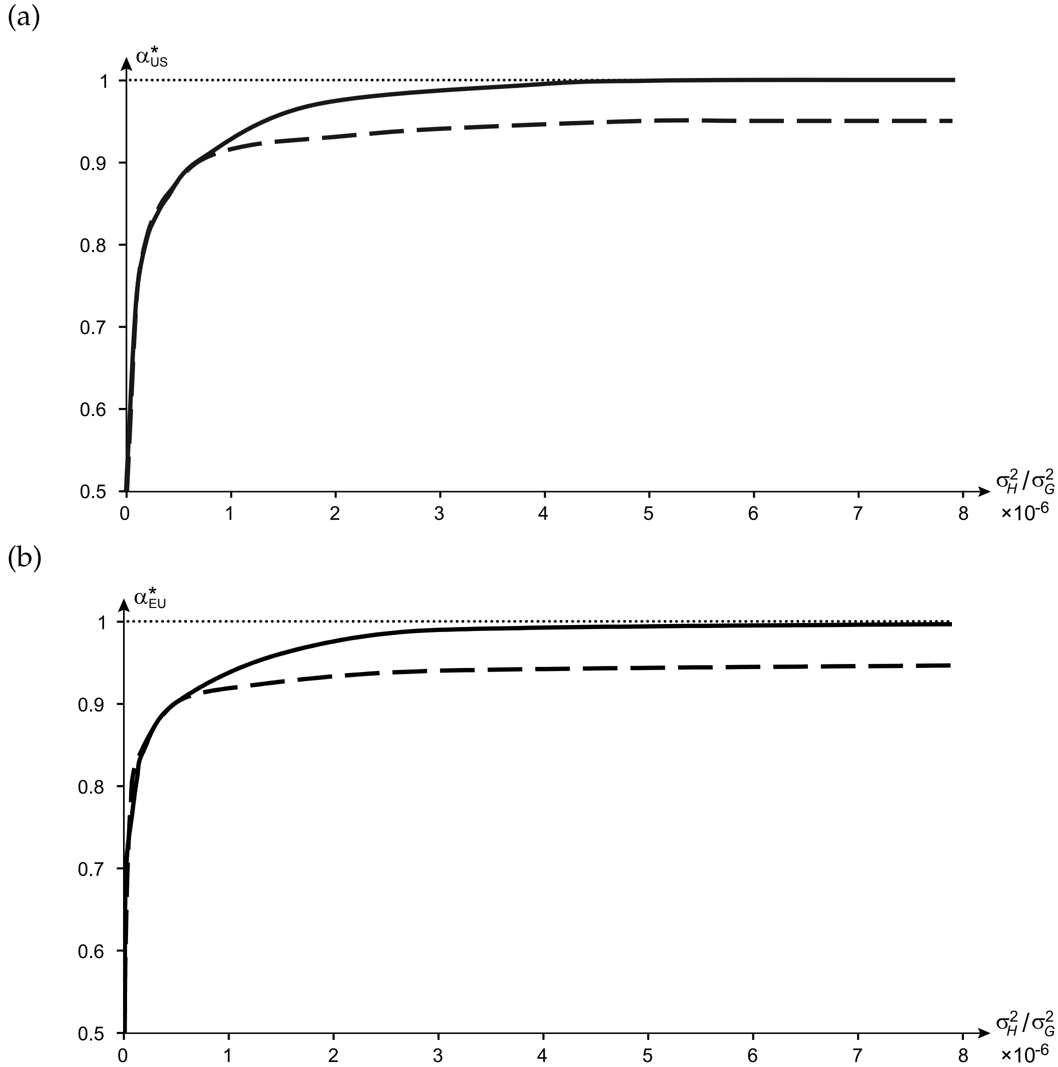

Figure 2: Best coefficient α for direct (dashed line) and Shapley value-based weight allocation rules (solid line) with (a) $n_1, \ldots, n_{51}$ defined by US population data and (b) $n_1, \ldots, n_{27}$ defined by EU27 population data

recommendations on rather qualitative assessments of polarization. It is not necessary to obtain precise estimates of σ_H^2/σ_G^2 in applications.

Note that Figure 2(b) considers real EU population data but counterfactually assumes Council decisions to be taken by a simple majority. However, Figure 1 suggests that a majority threshold of $q = 50\%$ may not be a critical condition for optimality of a linear Shapley rule, provided that $\sigma_H^2 > 0$. We will make this claim precise in the remainder of the section.

When we presume that assembly $\mathcal{R}^m$ uses the 50%-majority threshold defined in

combinatorial nature of weighted voting, which cannot totally be ignored even for $m = 27$ or $m = 51$. The respective α^*-induced $\|\cdot\|_1$-distances to the egalitarian ideal vary with the considered level of polarization. For both US and EU population partitions, they are minimal for $\sigma_H^2/\sigma_G^2 = 0$, then increase almost tenfold for $0 < \sigma_H^2/\sigma_G^2 < 1$, and eventually return to their original order of magnitude.

equation (2), the representative $P:m$ defined by (3) can be considered as the pivotal member of $\mathcal{R}^m$ without much qualification. We can generalize our model and consider arbitrary relative majority thresholds $q \in [0.5; 1)$ if we are willing to accept a weaker notion of pivotality. The complication is that the set of policy options that are q-majority undominated is no longer generically unique when $q > 0.5$; supermajority rules induce cores which typically consist of entire intervals. We can, nevertheless, generalize the quota definition in (2) to

$$q^m \equiv q \sum_{j=1}^{m} w_j, \tag{24}$$

for $q \in [0.5; 1)$ and consider the representative $P:m$ defined by (3) to be pivotal. This may be justified most easily by supposing that a legislative status quo $x^\circ \approx \infty$ exists and that formation of a winning coalition proceeds qualitatively in the same fashion as is sometimes assumed in order to motivate the Shapley value: coalition formation starts with the most enthusiastic supporters of change on the left, iteratively includes representatives further to the right, and gives all bargaining power to the first – and least enthusiastic – member who brings about the required supermajority to replace x°.[24]

Denote an m-member assembly $\mathcal{R}^m$ which uses the relative decision quota $q \in [0.5; 1)$ and chooses policy $x^* = \lambda_{P:m}$ as defined by (3)–(4) and (24) by $\mathcal{R}^{m,q}$. If $q > 0.5$, the corresponding pivot probabilities $\pi_i(\mathcal{R}^{m,q})$ and $\pi_j(\mathcal{R}^{m,q})$ of representatives i and j in general *fail* to exhibit the limit behavior with respect to m which is characterized in Theorem 1. So Corollaries 1 and 2 do not apply when $q > 0.5$.[25]

However, a second asymptotic relationship applies for $q = 0.5$ as well as $q \in (0.5; 1)$, for arbitrary *fixed* m, and without need for any kind of replica structure. Specifically, we can consider the situation in which given non-degenerate shock variables $\mu_1, \ldots, \mu_m$, whose common probability density h reflects preference heterogeneity across constituencies, are scaled by a non-negative factor t. Individual ideal points are then given by

$$\nu^l = t \cdot \mu_i + \epsilon^l \tag{25}$$

[24]Justifications for attributing most or all influence in $\mathcal{R}^m$ to representative $P:m$ in the supermajority case date back to Black (1948). Distance-dependent costs of policy reform, a strategic external agenda setter, or the need of assembly $\mathcal{R}^m$ to bargain with outsiders can motivate a focus on the core's extreme points. Status quo x° might also vary randomly on X such that it lies to the left or right of the core equiprobably (with $\pi_i(\mathcal{R}^m)$ then being i's pivot probability conditional on policy change).

[25]One can check numerically that when one considers rules $(w_1, \ldots, w_m) \propto (n_1{}^\alpha, \ldots, n_m{}^\alpha)$, the optimal coefficient $\alpha^*(q)$ for the i.i.d. case, where $\alpha^*(0.5) = 0.5$, increases *non-linearly* in q.

for $t \geq 0$. The corresponding ideal point of representative i from constituency C_i is

$$\lambda_i = t \cdot \mu_i + \tilde{\epsilon}_i \tag{26}$$

with

$$\tilde{\epsilon}_i = \text{median}\,\{\epsilon^l : l \in C_i\} \tag{27}$$

where we maintain the assumption that all μ_i and ϵ^l are mutually independent and respectively identically distributed for $i \in \{1, \ldots, m\}$ and $l \in \{1, \ldots, n\}$.

The i.i.d. case amounts to $t = 0$; and considering a large parameter t corresponds to investigating an electorate which is highly polarized along constituency lines. If we denote the pivot probability of representative i by $\pi_i(\mathcal{R}^{m,q,t})$ and the Shapley value of the weighted voting game $v = [q^m; w_1, \ldots, w_m]$ with q^m defined by (24) as $\phi(v)$, the following holds:

Theorem 2. *Consider an assembly $\mathcal{R}^{m,q}$ with an arbitrary number m of constituencies and the relative decision quota $q \in [0.5; 1)$. Let the ideal point of each representative $i \in \{1, \ldots, m\}$ be $\lambda_i = t \cdot \mu_i + \tilde{\epsilon}_i$, and suppose $\mu_1, \ldots, \mu_m$ and $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_m$ are all mutually independent random variables, $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_m$ have finite means and variances, and $\mu_1, \ldots, \mu_m$ have an identical bounded density. Then*

$$\lim_{t\to\infty} \frac{\pi_i(\mathcal{R}^{m,q,t})}{\pi_j(\mathcal{R}^{m,q,t})} = \frac{\phi_i(v)}{\phi_j(v)} \tag{28}$$

supposing that $\phi_j(v) > 0$.

The proof is provided in Appendix B and formalizes that the respective orderings of representatives which are induced by $\lambda_1, \ldots, \lambda_m$ and by $t \cdot \mu_1, \ldots, t \cdot \mu_m$ tend to coincide when t is large.[26] The theorem does not presume that $\tilde{\epsilon}_i$ satisfies (27); equality (28) applies also if λ_i is determined, e.g., by an oligarchy instead of the median voter of C_i. Moreover, it is worth noting that Theorem 2 does not impose any conditions like Theorem 1 on densities $g_1, \ldots, g_m$ or voting weights $w_1, \ldots, w_m$ in assembly $\mathcal{R}^{m,q}$. The Shapley value $\phi(v)$ automatically takes care of any combinatorial particularities associated with $w_1, \ldots, w_m$; and the convolution with $t \cdot \mu_i$'s bounded density, $\frac{1}{t}h\left(\frac{x}{t}\right)$, is sufficient to 'regularize' any (even non-continuous) distribution G_i of $\tilde{\epsilon}_i$. Applying Theorem 2 to the two-tier voting systems of Section 2, we conclude:

Corollary 3 (Linear Shapley rule)**.** *If the ideal points of voters can be conceived of as the sum of i.i.d. individual components ϵ^l and i.i.d. constituency components μ_i, and if heterogeneity across constituencies (variance of μ_i) is sufficiently great relative to heterogeneity within each*

[26] The density-driven intuition for Theorem 2 which is suggested by Figure 1(b) can also be made precise: under the additional assumption that the density h of the shock terms μ_i is Lipschitz continuous, the density functions of $\lambda_1, \ldots, \lambda_m$ converge uniformly to that of $t \cdot \mu_i$. A proof is available from the authors.

constituency (variance of ϵ^l) then

$$(w_1, \ldots, w_m) \text{ such that } \phi(q^m; w_1, \ldots, w_m) \propto (n_1, \ldots, n_m) \tag{29}$$

achieves approximately equal representation for any given relative decision quota $q \in [0.5; 1)$.

The indirect representation of bottom-tier voters which is achieved by this linear Shapley rule can fail to be close to egalitarian when m is small, the distribution of constituency sizes is extremely skewed or has small variance, or when q is close to 1. This is because the so-called *inverse problem* of finding weights which induce the desired Shapley value, or come close to it, often fails to have a good solution in these cases.[27] Still, provided that the considered heterogeneity across constituencies is sufficiently bigger than the heterogeneity within, the indirect representation achieved by weight allocation rule (29) is as egalitarian as possible for the given electorate's partition.

Whether Corollary 3 for the case of relevant preference affiliation within constituencies or Corollarly 2 for the i.i.d. case provides better guidance for designing a fair two-tier voting system in practice obviously depends. Some preference homogeneity within and dissimilarity across constituencies seems plausible – whether as the result of a sorting process ('voting with one's feet') à la Tiebout (1956), due to cultural uniformity fostered by geographical proximity and local policies (see Alesina and Spolaore 2003), or for other reasons. If constituencies correspond to entire nations, as in case of the EU Council or ECB Governing Council, citizens of a given constituency typically share more historical experience, traditions, language, communication etc. within constituencies than across. (This seems the key practical reason for why the issue of population size differences cannot trivially be resolved by redistricting.) However, the collective decisions that are taken by the top-tier assembly might be primarily about issues where opinions range over the same liberal–conservative, markets–government, dove–hawk, etc. spectrum in all constituencies. Moreover, there might be normative reasons outside the scope of our analysis for *pretending* that $\sigma_H^2 = 0$ even if it is not when one designs a presumably long-lasting, fair constitution. We, therefore, avoid any specific recommendations here for, say, new voting rules in the EU Council but warn that the i.i.d. presumption is more knife-edge and, therefore, seems to require special motivation.[28]

[27]For instance, the only feasible Shapley values in case of $m = 3$ constituencies are – up to isomorphisms – $(1/3, 1/3, 1/3)$, $(2/3, 1/6, 1/6)$ and $(1, 0, 0)$. A new approach to solving the inverse problem by using *integer linear programming* has been developed by Kurz (2012).

[28]A third alternative, inspired by the suggestion of "flexible" democratic mechanisms in other contexts (see Gersbach 2005, 2009), would be to specify different weighted voting rules for different policy domains. In some policy areas, such as competition policy, small or unstable between-constituency differences may call for square root weights; while fair decision making in other policy

5 Concluding remarks

This paper has developed two asymptotic results for the probability of being a decisive voter in order to address the practical issue of fair representation in two-tier voting systems. Our concern was the equalization of the *indirect influence* which bottom-tier voters can be expected to have on the top-tier decision in case of asymmetric constituencies and a one-dimensional convex policy space. Of course, many alternative criteria for choosing from all conceivable weighted voting rules exist (focusing, e.g., on consequentialist notions of welfare or information aggregation).

Penrose's square root rule has played a prominent role in the diplomatic debate concerning voting rules in the EU Council as well as in the general scientific discussion of two-tier voting over binary policy spaces. We have here provided it with a sound analytical foundation in a comparatively rich median voter environment where the 'one person, one vote' principle is operationalized as the goal to equalize pivot probabilities behind a constitutional veil of ignorance (Corollary 2).

Monte Carlo simulations by Maaser and Napel (2007) had first suggested that square root allocations perform well also outside Penrose's binomial world. More recent simulations even show that this extends from the considered influence-based criterion to utilitarian and majoritarian objective functions for specific example configurations (see Maaser and Napel 2012a; 2012b). However, there may be good reasons why most people consider a square root rule counterintuitive. The rule has been challenged on empirical grounds (especially by Gelman et al. 2004) and turns out to have limited robustness. It does not extend to supermajority rules; and it does badly if ideal points are correlated at the constituency level.

Our analysis has shown that a linear rule quickly performs better and becomes optimal for sufficiently strong preference affiliation within constituencies. Such a linear vs. square root dichotomy has arisen also in the binary voting model considered by Barberà and Jackson (2006), for a utilitarian institutional design objective. When Barberà and Jackson consider representatives who vote for candidates A or B in accordance with the majority of their constituents and each constituency's population is divided into the same number of groups within which individuals have identical preferences ("fixed-number-of-blocks model"), a linear weight allocation maximizes welfare. In contrast, if constituencies are made up of groups with fixed cardinality ("fixed-size-block model"), so that the number of homogeneous voter "blocks" is proportional to a constituency's population size, then welfare-maximizing weights vary with the square root of population sizes (provided blocks are sufficiently small). The fixed-number-of-blocks benchmark entails significant preference similarities

domains, such as agriculture or fisheries – with heterogenous shares of farmable land and some members landlocked, others islands – could involve linear weights.

within constituencies, while the latter model approaches situations with i.i.d. utilities as group size decreases. Other binary models, whose setups are all very different from ours, similarly allow the conclusion that linear weights are preferable in case of greater external than internal heterogeneity of the constituencies, while square root weights are called for in the i.i.d. case. Feix et al. (2008), for instance, show this when trying to minimize the probability of the top-tier assembly electing candidate *A* while the population majority prefers *B* as, e.g., for the 2000 US Electoral College.

The framework of this paper goes beyond the canonical case of binary collective decisions. It considers situations and two-tier voting systems that are of a non-binary nature, and incorporates voting under open agenda rule, bargaining, and other forms of strategic interaction in reduced form (by letting the respective constituency medians and Condorcet winner determine outcomes). Our two main results on weighted medians in interval spaces reinforce and generalize the pattern that has emerged in the literature on binary voting: as originally argued by Penrose (1946), ex ante independent and identical voters call for a weight allocation rule based on the square root of population sizes. However, sufficiently strong dissimilarity between constituencies renders most people's intuition correct – plain proportionality does the trick.

Appendices

A Proof of Theorem 1

A.1 Overview

Let us first give an overview of the five steps of the proof. In *Step 1*, we define a particular neighborhood I_m of the expected location of the weighted median of $\lambda_1, \ldots, \lambda_m$. This *essential interval* shrinks to {M} as $m \to \infty$. It is constructed such that the probabilities p_θ, $\overset{\sqsubset}{p}_\theta$, and $\overset{\sqsupset}{p}_\theta$ of a type-θ representative's ideal point falling inside I_m, inside I_m's left half, or inside I_m's right half, respectively, can suitably be bounded. Moreover, we decompose the deterministic total number $m_\theta = \beta_\theta(m) \cdot m$ of type-θ representatives in assembly $\mathcal{R}^m$ into the random numbers ${}^{\triangleleft}k_\theta$, k_θ, and $k^{\triangleright}_\theta$ of delegates with ideal points to I_m's left, inside I_m, and to I_m's right. Knowing the respective vector $\mathbf{k} = ({}^{\triangleleft}k_1, k_1, k^{\triangleright}_1, \ldots, {}^{\triangleleft}k_r, k_r, k^{\triangleright}_r)$ will be sufficient to determine whether the Condorcet winner is located inside I_m or not.

In *Step 2*, it is established that the weighted median of $\lambda_1, \ldots, \lambda_m$ is located inside the essential interval I_m with a probability that quickly approaches 1 as $m \to \infty$. As a corollary, the probability $\pi^\theta(\mathcal{R}^m)$ of the Condorcet winner having type θ converges to the corresponding *conditional* probability $\pi^\theta(\mathcal{R}^m|\mathcal{K})$ of a type-θ representative being pivotal where event $\mathcal{K}$ comprises all realizations of $\mathbf{k}$ such that $\mathcal{R}^m$'s weighted median lies inside I_m.

In *Step 3*, we show that the random orderings of the $k = \sum_{\theta \in \{1,\ldots,r\}} k_\theta$ representatives with ideal point realizations $\lambda_i \in I_m$ asymptotically become equiprobable as $m \to \infty$. It follows that, with a vanishing error, the respective conditional pivot probability $\pi^\theta(\mathcal{R}^m|\mathcal{K})$ equals the expected aggregate Shapley value of type-θ representatives in I_m.

In *Step 4*, the strong convergence result for the Shapley value by Neyman (1982) is applied to our setting. Neyman's result implies that the aggregate Shapley value of type-θ representatives with ideal points in I_m converges to their respective aggregate voting weight in each considered weighted voting 'subgame' among the representatives with ideal points $\lambda_i \in I_m$.

Having established that $\pi^\theta(\mathcal{R}^m)$ is asymptotically proportional to the aggregate voting weight of all type-θ representatives with ideal points inside I_m, aggregate probabilities are attributed to individual representatives in the final *Step 5*.

A.2 Proof

Step 1: Essential interval I_m and vector k

We begin by identifying a neighborhood of M and a sufficiently great number of representatives such that both the densities f_θ and the numbers of type-θ representatives in $\mathcal{R}^m$ can suitably be bounded. This leads to the definition of intervals I_m around M which later steps will focus on. Bounds for the probabilities of a type-θ representative's ideal point falling inside I_m, and more specifically into I_m's left or right halves, are provided in Lemma 2. The final part of *Step 1* introduces the vector **k** as a type-specific summary of how many ideal points are located to the left of I_m, inside I_m, and to its right.

First note that

$$0 < \underline{u} \equiv \min_{\theta' \in \{1,\ldots,r\}} f_{\theta'}(\mathrm{M}) \leq f_\theta(\mathrm{M}) \leq \overline{u} \equiv \max_{\theta' \in \{1,\ldots,r\}} f_{\theta'}(\mathrm{M}) \tag{30}$$

for every $\theta \in \{1, \ldots, r\}$. Using the continuity of f_θ in a neighborhood $(\mathrm{M} - \varepsilon_1, \mathrm{M} + \varepsilon_1)$ of M, which is implied by $|f_\theta(x) - f_\theta(\mathrm{M})| \leq cx^2$, we can choose $0 < \varepsilon_2 \leq \varepsilon_1$ such that

$$\frac{5}{6} f_\theta(\mathrm{M}) \leq f_\theta(x) \leq \frac{7}{6} f_\theta(\mathrm{M}) \tag{31}$$

for all $x \in [\mathrm{M} - \varepsilon_2, \mathrm{M} + \varepsilon_2]$ and any specific $\theta \in \{1, \ldots, r\}$. Inequality (30) can be used in order to obtain bounds

$$\frac{1}{2} \underline{u} \leq f_\theta(x) \leq 2\overline{u} \tag{32}$$

for all $x \in [\mathrm{M} - \varepsilon_2, \mathrm{M} + \varepsilon_2]$ and all $\theta \in \{1, \ldots, r\}$ which do not depend on θ. Due to the existence of some $m^0 \in \mathbb{N}$ such that $\beta_\theta(m) \geq \beta > 0$ for all $m \geq m^0$, we can also choose $0 < \varepsilon_3 \leq \varepsilon_2$ such that

$$\beta_\theta(m) \geq \beta > 0 \tag{33}$$

for all $m \geq \frac{1}{{\varepsilon_3}^{8/3}}$ and all $\theta \in \{1, \ldots, r\}$. And we can determine $0 < \varepsilon_4 \leq \varepsilon_3$ such that

$$24 < \underline{u}\beta \cdot (m\beta)^{\frac{1}{40}} \leq \underline{u}\beta m_\theta^{\frac{1}{40}} \tag{34}$$

for all $m \geq \frac{1}{{\varepsilon_4}^{8/3}}$, where $m_\theta \equiv \beta_\theta(m) \cdot m$.

Then define

$$\varepsilon(m) \equiv m^{-\frac{3}{8}} \tag{35}$$

and note that $\varepsilon(m) \leq \varepsilon_4$ iff $m \geq m^1 \equiv \frac{1}{{\varepsilon_4}^{8/3}} \geq m^0$. So, whenever we consider a sufficiently large number of representatives (specifically, $m \geq m^1$), inequalities (31)–

(34) are satisfied. We refer to

$$I_m \equiv [\mathrm{M} - \varepsilon(m), \mathrm{M} + \varepsilon(m)] \tag{36}$$

as the *essential interval*. The probability of an ideal point of type θ to fall inside I_m is

$$p_\theta \equiv \int_{\mathrm{M}-\varepsilon(m)}^{\mathrm{M}+\varepsilon(m)} f_\theta(x)dx. \tag{37}$$

For realizations in the left and right halves of I_m we respectively obtain the probabilities

$$\overset{\sqsubset}{p}_\theta \equiv \int_{\mathrm{M}-\varepsilon(m)}^{\mathrm{M}} f_\theta(x)dx \quad \text{and} \quad \overset{\sqsupset}{p}_\theta \equiv \int_{\mathrm{M}}^{\mathrm{M}+\varepsilon(m)} f_\theta(x)dx, \tag{38}$$

with $\overset{\sqsubset}{p}_\theta + \overset{\sqsupset}{p}_\theta = p_\theta$.

Lemma 2. *For $m \geq m^1$ we have*

$$\frac{5}{3} f_\theta(\mathrm{M})\varepsilon(m) \leq \quad p_\theta \quad \leq \frac{7}{3} f_\theta(\mathrm{M})\varepsilon(m), \tag{39}$$

$$\frac{5}{6} f_\theta(\mathrm{M})\varepsilon(m) \leq \overset{\sqsubset}{p}_\theta, \overset{\sqsupset}{p}_\theta \leq \frac{7}{6} f_\theta(\mathrm{M})\varepsilon(m), \tag{40}$$

$$\underline{u}\beta m_\theta^{-\frac{3}{8}} \leq \quad p_\theta \quad \leq 4\overline{u} m_\theta^{-\frac{3}{8}}, \text{ and} \tag{41}$$

$$\frac{1}{2}\underline{u}\beta m_\theta^{-\frac{3}{8}} \leq \overset{\sqsubset}{p}_\theta, \overset{\sqsupset}{p}_\theta \leq 2\overline{u} m_\theta^{-\frac{3}{8}}. \tag{42}$$

Proof. The inequalities can be concluded from (31)–(33), $m_\theta = \beta_\theta m$, and $\beta < 1$. □

Now for any realization $\boldsymbol{\lambda}$ of the ideal points in assembly $\mathcal{R}^m$, let

$$k_\theta \equiv \#\{j\colon \tau(j) = \theta \text{ and } \lambda_j \in [\mathrm{M} - \varepsilon(m), \mathrm{M} + \varepsilon(m)]\} \tag{43}$$

denote the number of type-θ representatives with a policy position in the essential interval, i.e., no more than $\varepsilon(m)$ away from the expected sample median M. Analogously, let

$$^{\triangleleft}k_\theta \equiv \#\{j\colon \tau(j) = \theta \text{ and } \lambda_j \in (-\infty, \mathrm{M} - \varepsilon(m))\} \tag{44}$$

and

$$k^{\triangleright}_\theta \equiv \#\{j\colon \tau(j) = \theta \text{ and } \lambda_j \in (\mathrm{M} + \varepsilon(m), \infty)\} \tag{45}$$

denote the random number of type-θ representatives to the left and to the right of I_m.

One can conceive of $\boldsymbol{\lambda}$-realizations as the results of a two-part random experiment: in the first part, it is determined for each λ_j whether it is located to the right of I_m, to its left, or inside I_m, e.g., by drawing a vector $\boldsymbol{l} = (l_1, \ldots, l_m)$ of independent random variables where $l_i = 1$ (-1) indicates a realization of λ_i to the right (left) of I_m and $l_i = 0$ indicates $\lambda_i \in I_m$ (with probabilities $\frac{1}{2} - \overset{\sqsubset}{p}_\theta$, $\frac{1}{2} - \overset{\sqsupset}{p}_\theta$, and p_θ, respectively). This already fixes $\overleftarrow{k}_\theta$, k_θ, and $\overrightarrow{k}_\theta$ for each $\theta \in \{1, \ldots, r\}$ and is summarized by the vector

$$\mathbf{k} = (\overleftarrow{k}_1, k_1, \overrightarrow{k}_1, \ldots, \overleftarrow{k}_r, k_r, \overrightarrow{k}_r). \tag{46}$$

In the second part, the exact ideal point locations are drawn. It will turn out that those outside I_m can be ignored with vanishing error; and the k_θ type-θ ideal points inside have conditional densities $\hat{f}_\theta$ with

$$\hat{f}_\theta(x) \equiv \frac{f_\theta(x)}{p_\theta} \quad \text{for } x \in I_m. \tag{47}$$

Step 2: Type θ's aggregate pivot probability $\pi^\theta(\mathcal{R}^m)$ converges to the conditional probability $\pi^\theta(\mathcal{R}^m|\mathcal{K})$ of type θ being pivotal in I_m

We next appeal to *Hoeffding's inequality*[29] in order to obtain bounds on the probability that the shares of representatives $\frac{\overleftarrow{k}_\theta}{m_\theta}$, $\frac{k_\theta}{m_\theta}$, and $\frac{\overrightarrow{k}_\theta}{m_\theta}$ with ideal points to the left, inside, or right of I_m deviate by more than a specified distance from their expectations. These bounds will imply that one can condition on the pivotal ideal point lying inside I_m in later steps of the proof with an exponentially decreasing error.

Hoeffding's inequality concerns the average $\overline{X} \equiv \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ of n independent bounded random variables $X_i \in [a_i, b_i]$ and guarantees

$$\Pr\left\{\left|\overline{X} - \mathbf{E}[\overline{X}]\right| > t\right\} \leq 2 \exp\left(\frac{-2t^2 n^2}{\sum_{i=1}^{n} (b_i - a_i)^2}\right). \tag{48}$$

Our specific construction will involve only random variables $X_i \in [0, 1]$, so that

$$\Pr\left\{\left|\overline{X} - \mathbf{E}[\overline{X}]\right| > t\right\} \leq 2 \exp\left(-2t^2 n\right). \tag{49}$$

We will put $n = m_\theta$ for a fixed $\theta \in \{1, \ldots, r\}$, so that $n \to \infty$ as $m \to \infty$, and choose $t = n^{-\frac{2}{5}}$, which implies $t(n) \ll \varepsilon(m)$. For this choice

$$\Pr\left\{\left|\overline{X} - \mathbf{E}[\overline{X}]\right| > n^{-\frac{2}{5}}\right\} \leq 2 \exp\left(-2n^{\frac{1}{5}}\right), \tag{50}$$

[29] See Hoeffding (1963, Theorem 2).

i.e., the probability of "extreme realizations" exponentially goes to zero as $m \to \infty$ (and hence $n = m_\theta \to \infty$).

Lemma 3. *For each $\theta \in \{1, \dots, r\}$ we have:*

$$
\begin{array}{llcl}
\text{(I)} & \Pr\left\{\frac{{}^{\triangleleft}k_\theta}{m_\theta} \in \left[\frac{1}{2} - \overset{\sqsubset}{p}_\theta - m_\theta^{-\frac{2}{5}}, \frac{1}{2} - \overset{\sqsubset}{p}_\theta + m_\theta^{-\frac{2}{5}}\right]\right\} & \geq & 1 - 2\exp\left(-2m_\theta^{\frac{1}{5}}\right) \\
\text{(II)} & \Pr\left\{\frac{k_\theta}{m_\theta} \in \left[p_\theta - m_\theta^{-\frac{2}{5}}, p_\theta + m_\theta^{-\frac{2}{5}}\right]\right\} & \geq & 1 - 2\exp\left(-2m_\theta^{\frac{1}{5}}\right) \\
\text{(III)} & \Pr\left\{\frac{k^{\triangleright}_\theta}{m_\theta} \in \left[\frac{1}{2} - \overset{\sqsupset}{p}_\theta - m_\theta^{-\frac{2}{5}}, \frac{1}{2} - \overset{\sqsupset}{p}_\theta + m_\theta^{-\frac{2}{5}}\right]\right\} & \geq & 1 - 2\exp\left(-2m_\theta^{\frac{1}{5}}\right).
\end{array}
$$

Proof. Let $\theta \in \{1, \dots, r\}$ be arbitrary but fixed. For statement (I) we consider the $n = m_\theta$ indices $j_1 \dots, j_{m_\theta} \in \{1, \dots, m\}$ of type θ and denote by X_i the random variable which is 1 if the realization λ_{j_i} lies inside the interval $(-\infty, \mathrm{M} - \varepsilon(m))$ and zero otherwise. In the notation of Hoeffding's inequality we have $\overline{\mathrm{X}} = \frac{{}^{\triangleleft}k_\theta}{m_\theta}$. Since the probability that λ_{j_i} lies in the left half of I_m is given by $\overset{\sqsubset}{p}_\theta$ and $\int_{-\infty}^{\mathrm{M}} f_\theta(x)dx = \int_{\mathrm{M}}^{\infty} f_\theta(x)dx = \frac{1}{2}$, the probability that λ_{j_i} lies in the interval $(-\infty, \mathrm{M} - \varepsilon(m))$ is given by $\frac{1}{2} - \overset{\sqsubset}{p}_\theta$. Thus we have $\mathbf{E}[\overline{\mathrm{X}}] = \frac{1}{2} - \overset{\sqsubset}{p}_\theta$ and (50) implies (I). The statements (II) and (III) follow along the same lines (namely, by letting X_i be the characteristic function of intervals $[\mathrm{M} - \varepsilon(m), \mathrm{M} + \varepsilon(m)]$ and $(\mathrm{M} + \varepsilon(m), \infty)$, respectively). Note that ${m_\theta}^{-2/5} \ll \varepsilon(m) = m^{-3/8}$ for large m. $\square$

We can use the bounds on p_θ in (41) and that $\beta m \leq m_\theta \leq m$ for $m \geq m^1 \geq m^0$ in order to conclude from (II) that for any given $\theta \in \{1, \dots, r\}$

$$
\underline{u}\beta^2 \varepsilon(m) \cdot m - m^{\frac{3}{5}} \leq k_\theta \leq 4\overline{u}\varepsilon(m) \cdot m + m^{\frac{3}{5}} \tag{51}
$$

with a probability of at least $1 - 2 \cdot \exp\left(-2{m_\theta}^{\frac{1}{5}}\right)$. A further implication of observations (I)–(III) is:

Lemma 4. *For $m \geq m^1$ the inequalities*

$$
\begin{aligned}
{}^{\triangleleft}k_\theta &< \frac{1}{2}m_\theta && (52) \\
k^{\triangleright}_\theta &< \frac{1}{2}m_\theta && (53) \\
{}^{\triangleleft}k_\theta + \frac{2}{3}k_\theta &> \frac{1}{2}m_\theta && (54) \\
k^{\triangleright}_\theta + \frac{2}{3}k_\theta &> \frac{1}{2}m_\theta && (55)
\end{aligned}
$$

are simultaneously satisfied for all $\theta \in \{1, \dots, r\}$ *with a probability of at least* $1 - 6r \cdot \exp\left(-2(\beta m)^{\frac{1}{5}}\right)$.

Proof. The events considered in statements (I), (II), and (III) of Lemma 3 are realized for *all* $\theta \in \{1, \ldots, r\}$ with a joint probability of at least

$$\left(1 - 2\exp\left(-2(\beta m)^{\frac{1}{5}}\right)\right)^{3r} \geq 1 - 6r\exp\left(-2(\beta m)^{\frac{1}{5}}\right), \tag{56}$$

since $m_\theta \geq \beta m$ for $m \geq m^0$ and $(1-x)^k \geq (1-kx)$ is valid for all $x \in [0,1]$ and $k \in \mathbb{N}$. If $m \geq m^1$, we then have

$$\overset{\triangleleft}{k}_\theta \leq \left(\frac{1}{2} - \overset{\sqsubset}{p}_\theta\right) \cdot m_\theta + {m_\theta}^{\frac{3}{5}} \leq \frac{m_\theta}{2} - \frac{u\beta {m_\theta}^{\frac{5}{8}}}{2} + {m_\theta}^{\frac{3}{5}} = \frac{m_\theta}{2} - {m_\theta}^{\frac{3}{5}} \underbrace{\left(\frac{u\beta m_\theta^{\frac{1}{40}}}{2} - 1\right)}_{>0} < \frac{1}{2} m_\theta \tag{57}$$

for any $\theta \in \{1, \ldots, r\}$. The first inequality follows directly from (I), the second inequality uses (42), and the final inequality follows from (34). Analogous inequalities pertain to $\overset{\triangleright}{k}_\theta$.

Moreover, we can conclude

$$\begin{aligned}
\overset{\triangleleft}{k}_\theta + \frac{2}{3}k_\theta \quad &\geq \quad \left(\frac{1}{2} - \overset{\sqsubset}{p}_\theta\right) \cdot m_\theta - {m_\theta}^{\frac{3}{5}} + \frac{2p_\theta}{3} m_\theta - \frac{2}{3}{m_\theta}^{\frac{3}{5}} && (58)\\
&= \quad \frac{m_\theta}{2} - \frac{5}{3}{m_\theta}^{\frac{3}{5}} + \left(\frac{2p_\theta}{3} - \overset{\sqsubset}{p}_\theta\right) m_\theta && (59)\\
&= \quad \frac{m_\theta}{2} + \frac{5}{3}{m_\theta}^{\frac{3}{5}} \left(\frac{2\overset{\sqsupset}{p}_\theta}{5}{m_\theta}^{\frac{2}{5}} - \frac{\overset{\sqsubset}{p}_\theta}{5}{m_\theta}^{\frac{2}{5}} - 1\right) && (60)\\
&\geq \quad \frac{m_\theta}{2} + \frac{5}{3}{m_\theta}^{\frac{3}{5}} \left(\frac{1}{10} \cdot \frac{3}{7} p_\theta \cdot {m_\theta}^{\frac{2}{5}} - 1\right) && (61)\\
&\geq \quad \frac{m_\theta}{2} + \frac{5}{3}{m_\theta}^{\frac{3}{5}} \underbrace{\left(\frac{u\beta {m_\theta}^{\frac{1}{40}}}{24} - 1\right)}_{>0} > \frac{1}{2} m_\theta. && (62)
\end{aligned}$$

The first inequality uses (I) and (II); the second one employs (39) and (40); the third applies (41); and the final one invokes (34). Analogous inequalities pertain to $\overset{\triangleright}{k}_\theta + \frac{2}{3}k_\theta$. □

Lemma 4 implies that the respective unweighted sample median among representatives of type θ is located within I_m for all $\theta \in \{1, \ldots, r\}$ with a probability that quickly approaches 1. The same must *a fortiori* be true for the pivotal assembly member, i.e., the weighted median among all representatives.

We collect in the set $\mathcal{K}$ all $\mathbf{k} = (\overset{\triangleleft}{k}_1, k_1, \overset{\triangleright}{k}_1, \ldots, \overset{\triangleleft}{k}_r, k_r, \overset{\triangleright}{k}_r)$ such that the events considered by Lemma 3, (I)–(III), are realized for *all* $\theta \in \{1, \ldots, r\}$. The inequalities in Lemma 4 then hold for any $\mathbf{k} \in \mathcal{K}$. We can decompose the probability $\pi^\theta(\mathcal{R}^m)$ of some type-θ representative being pivotal into conditional probabilities $\pi^\theta(\mathcal{R}^m|\mathcal{K})$ and $\pi^\theta(\mathcal{R}^m|\neg\mathcal{K})$

which respectively concern only λ-realizations where $\mathbf{k} \in \mathcal{K}$ and $\mathbf{k} \notin \mathcal{K}$. Then Lemma 4 implies

$$\begin{aligned} \pi^\theta(\mathcal{R}^m) &= \Pr\{\mathcal{K}\} \cdot \pi^\theta(\mathcal{R}^m|\mathcal{K}) + \Pr\{\neg\mathcal{K}\} \cdot \pi^\theta(\mathcal{R}^m|\neg\mathcal{K}) \\ &= \pi^\theta(\mathcal{R}^m|\mathcal{K}) + O(exp(-2m^{\frac{1}{5}})). \end{aligned} \tag{63}$$

Step 3: $\pi^\theta(\mathcal{R}^m|\mathcal{K})$ converges to the expectation of type θ's Shapley value inside I_m

Now condition on some $\mathbf{k} \in \mathcal{K}$ such that exactly $\sum_\theta k_\theta = k$ ideal points fall inside the essential interval, where k is asymptotically proportional to $\varepsilon(m) \cdot m = m^{\frac{5}{8}}$ by (51). Label them $1, \ldots, k$ for ease of notation and let $\varrho \in \mathcal{S}_k$ denote an arbitrary element of the space $\mathcal{S}_k$ of permutations which bijectively map $(1, \ldots, k)$ to some $(j_1, \ldots, j_k)$. The conditional probability for the event that the k ideal points located in I_m are ordered exactly as they are in ϱ by the second step of the experiment is

$$p(\varrho|\mathbf{k}) \equiv \int_{-\varepsilon(m)}^{\varepsilon(m)} \int_{x_{j_1}}^{\varepsilon(m)} \cdots \int_{x_{j_{k-1}}}^{\varepsilon(m)} \hat{f}_{j_1}(x_{j_1}) \ldots \hat{f}_{j_k}(x_{j_k})\, dx_{j_k} \ldots dx_{j_2} dx_{j_1}. \tag{64}$$

Lemma 5. *For all $m \geq m^1$, any $\mathbf{k} \in \mathcal{K}$ with $\sum_\theta k_\theta = k$ and permutation $\varrho \in \mathcal{S}_k$ we have*

$$p(\varrho|\mathbf{k}) = \frac{1}{k!} + \frac{1}{k!} \cdot O(m^{-\frac{1}{8}}). \tag{65}$$

Proof. The premise $|f_\theta(x) - f_\theta(\mathrm{M})| \leq cx^2$ for $x \in I_m$ permits us to choose $\delta \in O(\varepsilon(m)^2)$ with $\delta \leq \frac{1}{2}$ such that

$$(1-\delta) \cdot f_\theta(\mathrm{M}) \leq f_\theta(x) \leq (1+\delta) \cdot f_\theta(\mathrm{M}) \tag{66}$$

and, equivalently,

$$(1-\delta) \cdot \hat{f}_\theta(\mathrm{M}) \leq \hat{f}_\theta(x) \leq (1+\delta) \cdot \hat{f}_\theta(\mathrm{M}) \tag{67}$$

for all types $1 \leq \theta \leq r$ and all $x \in I_m$. Integrating (66) on I_m yields

$$2\varepsilon(m)(1-\delta) \cdot f_\theta(\mathrm{M}) \leq p_\theta \leq 2\varepsilon(m)(1+\delta) \cdot f_\theta(\mathrm{M}). \tag{68}$$

With these bounds we can conclude from $\hat{f}_\theta(\mathrm{M}) = \frac{f_\theta(\mathrm{M})}{p_\theta}$ that

$$\frac{1-\delta}{2\varepsilon(m)} \leq \frac{1}{2\varepsilon(m)(1+\delta)} \leq \hat{f}_\theta(\mathrm{M}) \leq \frac{1}{2\varepsilon(m)(1-\delta)} \leq \frac{1+2\delta}{2\varepsilon(m)} \tag{69}$$

because $1/(1-\delta) \leq 1 + 2\delta$.

Using $(1-\delta)^k \geq 1 - k\delta$ and $(1+\delta)^k \leq 1 + 2k\delta$ for $k\delta \leq 1$,[30] and noting that the hypercube $[0,1]^k$ can be partitioned into $k!$ polytopes $\{x \in [0,1]^k : x_{j_1} \leq x_{j_2} \leq \ldots \leq x_{j_k}\}$ with equal volume, inequality (67) yields

$$\begin{aligned}
p(\varrho|\mathbf{k}) &\geq (1-\delta)^k \int_{-\varepsilon(m)}^{\varepsilon(m)} \int_{x_{j_1}}^{\varepsilon(m)} \cdots \int_{x_{j_{k-1}}}^{\varepsilon(m)} \hat{f}_{j_1}(\mathrm{M}) \ldots \hat{f}_{j_k}(\mathrm{M})\, dx_{j_k} \ldots dx_{j_2} dx_{j_1} && (70)\\
&= \frac{(1-\delta)^k}{k!} \cdot \hat{f}_{j_1}(\mathrm{M}) \ldots \hat{f}_{j_k}(\mathrm{M}) \int_{-\varepsilon(m)}^{\varepsilon(m)} \int_{-\varepsilon(m)}^{\varepsilon(m)} \cdots \int_{-\varepsilon(m)}^{\varepsilon(m)} 1\, dx_{j_k} \ldots dx_{j_2} dx_{j_1} && (71)\\
&= \frac{(1-\delta)^k}{k!} \cdot \hat{f}_{j_1}(\mathrm{M}) \ldots \hat{f}_{j_k}(\mathrm{M}) \cdot (2\varepsilon(m))^k && (72)\\
&\overset{(69)}{\geq} \frac{(1-\delta)^{2k}}{k!} \geq \frac{1-2k\delta}{k!} && (73)
\end{aligned}$$

and, analogously,

$$\begin{aligned}
p(\varrho|\mathbf{k}) &\leq (1+\delta)^k \int_{-\varepsilon(m)}^{\varepsilon(m)} \int_{x_{j_1}}^{\varepsilon(m)} \cdots \int_{x_{j_{k-1}}}^{\varepsilon(m)} \hat{f}_{j_1}(\mathrm{M}) \ldots \hat{f}_{j_k}(\mathrm{M})\, dx_{j_k} \ldots dx_{j_2} dx_{j_1} && (74)\\
&= \frac{(1+\delta)^k}{k!} \cdot \hat{f}_{j_1}(\mathrm{M}) \ldots \hat{f}_{j_k}(\mathrm{M}) \cdot (2\varepsilon(m))^k && (75)\\
&\overset{(69)}{\leq} \frac{(1+\delta)^k(1+2\delta)^k}{k!} \leq \frac{(1+2\delta)^{2k}}{k!} \leq \frac{1+8k\delta}{k!}. && (76)
\end{aligned}$$

This implies

$$\left| p(\varrho|\mathbf{k}) - \frac{1}{k!} \right| \leq \frac{8k\delta}{k!}. \tag{77}$$

Because $k \in O(m^{\frac{5}{8}})$ and $\delta \in O(m^{-\frac{6}{8}})$, the relative error $|p(\varrho|\mathbf{k}) - (k!)^{-1}| \big/ (k!)^{-1}$ tends to zero at least as fast as $O(m^{-\frac{1}{8}})$. □

So even though the probabilities of the orderings $\varrho \in \mathcal{S}_k$ of the k agents inside I_m differ depending on which specific ϱ is considered and what are the involved representative types (i.e., which $\mathbf{k}$ is considered), these differences vanish and all orderings become equiprobable as m gets large.

Type θ's conditional pivot probability can be written as

$$\pi^\theta(\mathcal{R}^m|\mathcal{K}) = \sum_{\mathbf{k}\in\mathcal{K}} P(\mathbf{k}) \cdot \Big\{ \sum_{\varrho\in\mathcal{S}_k \,:\, \psi(\mathbf{k},\varrho)=\theta} p(\varrho|\mathbf{k}) \Big\}, \tag{78}$$

---

[30]The first statement is easily seen by induction on k. The second follows from

$$(1+\delta)^k = \sum_{j=0}^{k} \binom{k}{j} \delta^j \leq 1 + \sum_{j=1}^{k} \frac{1}{j!} \underbrace{(k\delta)^j}_{\leq k\delta} \leq 1 + k\delta \underbrace{\sum_{j=1}^{k} \tfrac{1}{j!}}_{\leq e-1} \leq 1 + 2k\delta.$$

Since k is asymptotically proportional to $m^{\frac{5}{8}}$ and $\varepsilon(m)^2 = m^{-\frac{6}{8}}$ we can choose $\delta \in O(m^{-\frac{6}{8}})$ with $(k\delta)^j \leq k\delta$ for $j \geq 1$ whenever m is large enough.

where $P(\mathbf{k})$ denotes the probability of $\mathbf{k}$ conditional on event $\{\mathbf{k} \in \mathcal{K}\}$ and function $\psi \colon \mathcal{K} \times \mathcal{S}_k \to \{1, \ldots, r\}$ identifies the type θ' of the pivotal member in $\mathcal{R}^m$ when $\mathbf{k}$ describes how the representative types are divided between I_m and its left or right, and ϱ captures the ordering inside I_m. Lemma 5 approximates the probability of ordering ϱ conditional on $\mathbf{k}$ as $1/k!$, and one thus obtains

$$\pi^{\theta}(\mathcal{R}^m|\mathcal{K}) = \sum_{\mathbf{k} \in \mathcal{K}} P(\mathbf{k}) \cdot \phi_{\theta}(\mathbf{k}) + O(m^{-\frac{1}{8}}) \tag{79}$$

with

$$\phi_{\theta}(\mathbf{k}) = \sum_{\varrho \in \mathcal{S}_k \,:\, \psi(\mathbf{k},\varrho)=\theta} \frac{1}{k!}. \tag{80}$$

Because a constant factor $\frac{1}{k!}$ pertains to each ordering $\varrho \in \mathcal{S}_k$, $\phi_{\theta}(\mathbf{k})$ equals the probability that, as the weights $w_1, w_2, \ldots, w_k$ of the k representatives inside I_m are accumulated in *uniform* random order, the threshold $q(\mathbf{k}) \equiv q^m - \sum_{\theta \in \{1,\ldots,r\}} {}^{\triangleleft}k_{\theta} w_{\theta}$ is first reached by the weight of a type-θ representative. The term $\phi_{\theta}(\mathbf{k})$ is, therefore, simply the aggregated *Shapley value* of the type-θ representatives in the weighted voting game defined by quota $q(\mathbf{k})$ and weight vector $(w_1, w_2, \ldots, w_k)$. Equation (79) states that $\pi^{\theta}(\mathcal{R}^m|\mathcal{K})$ converges to the expectation of this Shapley value $\phi_{\theta}(\mathbf{k})$.

Step 4: Type θ's Shapley value $\phi_{\theta}(\mathbf{k})$ converges to θ's relative weight in I_m

Condition $\mathbf{k} \in \mathcal{K}$ implies $\frac{1}{3} \cdot \sum_{\theta \in \{1,\ldots,r\}} k_{\theta} w_{\theta} \leq q(\mathbf{k}) \leq \frac{2}{3} \cdot \sum_{\theta \in \{1,\ldots,r\}} k_{\theta} w_{\theta}$ (see Lemma 4). And our premises guarantee that the relative weight of each individual representative in I_m shrinks to zero. The "Main Theorem*" in Neyman (1982), therefore, has the following corollary:

Lemma 6 (Neyman 1982)**.** *Given that* $\mathbf{k} \in \mathcal{K}$*,*

$$\phi_{\theta}(\mathbf{k}) = \frac{k_{\theta} w_{\theta}}{\sum_{\theta'=1}^{r} k_{\theta'} w_{\theta'}} \cdot (1 + \mu(m)) \quad \textit{with} \quad \lim_{m \to \infty} |\mu(m)| = 0. \tag{81}$$

Proof. Neyman's theorem considers an infinite sequence of weighted voting games $[q^n; \boldsymbol{w}^n]$ with n voters whose individual relative weights w_i^n approach 0, and in which the relative quota q^n is bounded away from 0 and 100% (or at least $\lim_{n\to\infty} q^n / (\max_i w_i^n) = \infty$). Neyman establishes that[31]

$$\lim_{n \to \infty} |\phi_{T_n}(q^n; \boldsymbol{w}^n) - \sum_{i \in T_n} w_i^n| = 0 \tag{82}$$

holds for any sequence of voter subsets $T_n \subseteq \{1, \ldots, n\}$, where $\phi_{T_n}(q^n; \boldsymbol{w}^n)$ denotes their

[31] We somewhat specialize his finding and adapt the notation.

aggregate Shapley value. (We here consider $q^n = q(\mathbf{k})/w_\Sigma$, $\boldsymbol{w}^n = (w_1, w_2, \ldots, w_k)/w_\Sigma$ and $T_n = \{i \in N\colon \tau(i) = \theta\}$ for $N = \{1, \ldots, k\}$ and $w_\Sigma = \sum_{i \in N} w_i$.[32])

It is trivial that (81) holds if $w_\theta = 0 = \phi_\theta(\mathbf{k})$. So we can assume $w_\theta > 0$, and because there is at least the proportion $\beta > 0$ of representatives from each type in I_m for large m, the aggregate relative weight of θ-type representatives in I_m is bounded away from 0, i.e.,[33]

$$\liminf_{m\to\infty} \frac{k_\theta w_\theta}{\sum_{\theta'=1}^r k_{\theta'} w_{\theta'}} > 0. \tag{83}$$

Therefore, not only the absolute error $\tilde{\mu}(m)$ made in approximating $\phi_\theta(\mathbf{k}) = \phi_{T_n}(q^n; \boldsymbol{w}^n)$ by $\frac{k_\theta w_\theta}{\sum_{\theta'=1}^r k_{\theta'} w_{\theta'}}$ but also the relative error $\mu(m) \equiv \tilde{\mu}(m) / \frac{k_\theta w_\theta}{\sum_{\theta'=1}^r k_{\theta'} w_{\theta'}}$ must vanish as $m \to \infty$. □

Step 5: Attributing aggregate pivot probabilities to individual representatives

It then remains to disaggregate the pivot probabilities $\pi^\theta(\mathcal{R}^m)$ and $\pi^{\theta'}(\mathcal{R}^m)$ of types θ and θ' to individual representatives i and j. The aggregate relative weight of type-θ representatives in the essential interval satisfies

$$\frac{k_\theta w_\theta}{\sum_{\theta'=1}^r k_{\theta'} w_{\theta'}} = \frac{\beta_\theta(m) m p_\theta w_\theta (1 + O(m^{-\frac{2}{5}}))}{\sum_{\theta'=1}^r \beta_{\theta'}(m) m p_{\theta'} w_{\theta'} (1 - O(m^{-\frac{2}{5}}))} = \frac{\beta_\theta(m) p_\theta w_\theta}{\sum_{\theta'=1}^r \beta_{\theta'}(m) p_{\theta'} w_{\theta'}} \left(1 + O(m^{-\frac{2}{5}})\right) \tag{84}$$

for any $\mathbf{k} \in \mathcal{K}$ (see (II) in Lemma 3).[34] Combining this with equations (63), (79) and (81) yields

$$\lim_{m\to\infty} \frac{\pi^\theta(\mathcal{R}^m)}{\pi^{\theta'}(\mathcal{R}^m)} = \lim_{m\to\infty} \frac{\beta_\theta(m) p_\theta w_\theta}{\beta_{\theta'}(m) p_{\theta'} w_{\theta'}} = \lim_{m\to\infty} \frac{\beta_\theta(m) f_\theta(\mathrm{M}) w_\theta}{\beta_{\theta'}(m) f_{\theta'}(\mathrm{M}) w_{\theta'}} \tag{85}$$

for arbitrary $\theta, \theta' \in \{1, \ldots, r\}$. Here, the final equality uses

$$\lim_{m\to\infty} \frac{p_\theta}{p_{\theta'}} = \lim_{m\to\infty} \frac{\int_{-\varepsilon(m)}^{\varepsilon(m)} f_\theta(x)dx}{\int_{-\varepsilon(m)}^{\varepsilon(m)} f_{\theta'}(x)dx} = \frac{f_\theta(\mathrm{M})}{f_{\theta'}(\mathrm{M})}, \tag{86}$$

which can be deduced from (68).

Our main result then follows from noting that the $m_\theta = \beta_\theta(m) \cdot m$ representatives of type θ in assembly $\mathcal{R}^m$ are symmetric to each other and, therefore, must have identical pivot probabilities in $\mathcal{R}^m$. Hence

$$\lim_{m\to\infty} \frac{\pi_i(\mathcal{R}^m)}{\pi_j(\mathcal{R}^m)} = \lim_{m\to\infty} \frac{\pi^{\tau(i)}(\mathcal{R}^m)/\beta_{\tau(i)}(m)}{\pi^{\tau(j)}(\mathcal{R}^m)/\beta_{\tau(j)}(m)} = \frac{f_i(\mathrm{M}) w_i}{f_j(\mathrm{M}) w_j}. \tag{87}$$

[32] Our notation leaves some inessential technicalities implicit: $\mathcal{K}$ really refers to a family of such sets, parameterized by m; we implicitly consider a sequence of $\mathbf{k}$-vectors such that $n = k \to \infty$ as $m \to \infty$.

[33] The limit itself need not exist because our premises do not rule out that, e.g., m_θ is periodic in m.

[34] To see the second equality note that for $y \in (0, \frac{1}{2})$ we have $\frac{1}{1-y} = 1 + y + y^2 + \ldots \leq 1 + 2y = 1 + O(y)$. Similarly, $\frac{1}{1-y} \geq 1 + y = 1 + O(y)$ and so $\frac{1}{1-y} = 1 + O(y)$.

■

A.3 Remarks

Let us end this appendix with remarks on possible further generalizations. First, the quadratic bound on f_θ's variation in a neighborhood of M could be relaxed by choosing different constants in equations (35) and (49): $t(m_\theta) = m_\theta^{-b_1}$ with $b_1 < \frac{1}{2}$ is all that is needed in order to ensure a vanishing error probability in (49); and $\varepsilon(m) = m^{-b_2}$ with $b_2 < b_1$ in (35) is sufficient for $\varepsilon(m) \gg t(m_\theta)$. Then a local bound $|f_\theta(x) - f_\theta(\mathrm{M})| \leq cx^a$ for $a > \frac{1-b_2}{b_2}$ is sufficient to establish Lemma 5. Requirement $b_2 < b_1 < \frac{1}{2}$ leaves generous room for $a < 2$, but implies $a > 1$.

Second, it is actually sufficient to assume *local continuity* of all f_θ at M, rather than any strengthening of this,[35] if one appeals to an unpublished result by Abraham Neyman. When, as in our setting, all voting weights have the same order of magnitude, the uniform convergence theorem of Neyman (1982) for the Shapley value can be generalized to hold for all *random order values* that are 'sufficiently close' to the Shapley value. More specifically, consider the expected marginal contribution of a voter $i \in \{1, \ldots, k\}$

$$\Phi_i(v) \equiv \sum_{\varrho \in \mathcal{S}_k} p(\varrho) \cdot [v(T_i(\varrho) \cup \{i\}) - v(T_i(\varrho))] \tag{88}$$

in a weighted voting game $v = [q; w_1, \ldots, w_k]$, where any given permutation $\varrho \in \mathcal{S}_k$ on $N = \{1, \ldots, k\}$ has probability $p(\varrho)$, and $T_i(\varrho) \subset N$ denotes the set of i's predecessors in ϱ, i.e., $T_i(\varrho) = \{j\colon \varrho(j) < \varrho(i)\}$. The random order value $\Phi(v)$ equals the Shapley value $\phi(v)$ if $p(\varrho) = \frac{1}{k!}$. This equiprobability can, for instance, be obtained by letting ϱ be defined by the order statistics of a vector of random variables $\boldsymbol{X} = (X_1, \ldots, X_k)$ with mutually independent and $[0,1]$-uniformly distributed $X_1, \ldots, X_k$. The latter assumption can be relaxed somewhat without destroying the asymptotic proportionality of i's weight w_i and $\Phi_i(v)$ which Neyman (1982) has established when $\Phi(v) = \phi(v)$:

Theorem 3 (Neyman, personal communication)**.** *Fix $L > 1$. For every $\varepsilon > 0$ there exist $\delta > 0$ and $K > 0$ such that if v is the weighted voting game $v = [q; w_1, \ldots, w_k]$ with $w_1, \ldots, w_k > 0$, $\sum_{i=1}^k w_i = 1$, $K \cdot \max_i w_i < q < 1 - K \cdot \max_i w_i$, $\max_{i,j} w_i/w_j < L$, and $\{p(\varrho)\}_{\varrho \in \mathcal{S}_k}$ in (88) is defined by the order statistics of independent $[0,1]$-valued random variables $X_1, \ldots, X_k$ with densities f_i such that $1 - \delta < f_i(x) < 1 + \delta$ for every $x \in [0,1]$ and*

[35] Local continuity of f_θ is obviously necessary: a modification of $f_\theta(\mathrm{M})$ – with $f_\theta(x)$ unchanged for $x \neq \mathrm{M}$ – would affect $w_i f_\theta(\mathrm{M})$ but not $\pi_i(\mathcal{R}^m)$. Also the requirement of *positive density* at the common median cannot be relaxed. This is seen, e.g., by considering densities f_i, f_j where $f_i(x) = 0$ on a neighborhood $N_\varepsilon(\mathrm{M})$ while $f_j(\mathrm{M}) = 0$ with $f_j(x) > 0$ for $x \in N_\varepsilon(\mathrm{M}) \setminus \{\mathrm{M}\}$; then $\pi_i(\mathcal{R}^m)/\pi_j(\mathcal{R}^m)$ converges to 0 rather than w_i/w_j.

$i \in \{1, \ldots, k\}$ *then*

$$\sum_{i=1}^{k} |w_i - \Phi_i(v)| < \varepsilon. \tag{89}$$

Of course, one can equivalently let $\{p(\varrho)\}_{\varrho \in \mathcal{S}_k}$ be defined by the order statistics of independent I_m-valued random variables with densities $\hat{f}_1, \ldots, \hat{f}_k$, instead of $[0,1]$-valued ones, if the theorem's condition $1 - \delta < f_i(x) < 1 + \delta$ is replaced by the requirement that $\frac{1-\delta}{2\varepsilon(m)} < \hat{f}_i(x) < \frac{1+\delta}{2\varepsilon(m)}$ for all $x \in I_m$.

The values of δ and L which one obtains for a given ε in Theorem 3 apply to *any* value of k. We consider the weighted voting subgames played by the $k = \sum_{\theta \in \{1,\ldots,r\}} k_\theta$ representatives with realizations $\lambda_i \in I_m$ for given $\mathbf{k} \in \mathcal{K}$. The relative weight of any such representative i, $\hat{w}_i = w_i / \sum_{\theta \in \{1,\ldots,r\}} k_\theta w_\theta$, approaches zero as $m \to \infty$; and so does the maximum relative weight. Recalling that the corresponding subgame's relative quota $\hat{q} = q(\boldsymbol{k}) / \sum_{\theta \in \{1,\ldots,r\}} k_\theta w_\theta$ is bounded by $\frac{1}{3} \leq \hat{q} \leq \frac{2}{3}$, the condition $K \cdot \max_i \hat{w}_i < \hat{q} < 1 - K \cdot \max_i \hat{w}_i$ is satisfied when m is sufficiently large. Any null players with $w_i = 0$ can w.l.o.g. be removed from consideration. Then all weights have the same order of magnitude, i.e., the choice of L such that $\max_{i,j} \hat{w}_i / \hat{w}_j < L$ holds for all $\mathbf{k} \in \mathcal{K}$ is trivial.

Moreover, the conditional densities $\hat{f}_\theta$ in our setup satisfy $\frac{1-\delta}{2\varepsilon(m)} < \hat{f}_i(x) < \frac{1+\delta}{2\varepsilon(m)}$ for every $\theta \in \{1, \ldots, r\}$ and $x \in I_m$ when m is large enough. Specifically, continuity of f_θ in a neighborhood of M implies that for any given $\varepsilon > 0$ there exists $\Delta(\varepsilon) > 0$ with $\lim_{\varepsilon \downarrow 0} \Delta(\varepsilon) = 0$ such that

$$(1 - \Delta(\varepsilon)) \cdot f_\theta(\mathrm{M}) \leq f_\theta(x) \leq (1 + \Delta(\varepsilon)) \cdot f_\theta(\mathrm{M}) \tag{90}$$

for all $x \in [\mathrm{M} - \varepsilon, \mathrm{M} + \varepsilon]$ and all $\theta \in \{1, \ldots, r\}$ (cf. inequality (31)). Similarly to inequality (39) we then conclude

$$(1 - \Delta(\varepsilon)) f_\theta(\mathrm{M}) \cdot 2\varepsilon \leq p_\theta \leq (1 + \Delta(\varepsilon)) f_\theta(\mathrm{M}) \cdot 2\varepsilon. \tag{91}$$

Combining the last two inequalities with inequality (47) yields

$$\frac{(1 - \Delta(\varepsilon))}{(1 + \Delta(\varepsilon)) \cdot 2\varepsilon} \leq \hat{f}_\theta(x) \leq \frac{(1 + \Delta(\varepsilon))}{(1 - \Delta(\varepsilon)) \cdot 2\varepsilon}. \tag{92}$$

So considering $\varepsilon = \varepsilon(m)$ and any fixed δ, the conditional densities $\hat{f}_\theta$ satisfy $\frac{1-\delta}{2\varepsilon(m)} < \hat{f}_i(x) < \frac{1+\delta}{2\varepsilon(m)}$ for every $\theta \in \{1, \ldots, r\}$ and $x \in I_m$ when m is sufficiently large.

Hence, all premises in Neyman's unpublished Theorem 3 are satisfied by the corresponding weighted voting subgames of agents with ideal points in I_m. Theorem 3, therefore, ensures the approximate weight proportionality of the aggregate random order value Φ of the type-θ representatives. Now if one recalls (78) and notices that the bracketed sum equals $\Phi(v)$ with $v = [\hat{q}; \hat{w}_{j_1}, \ldots, \hat{w}_{j_k}]$ when $j_1, \ldots, j_k$ denote the representatives with ideal points in I_m, we can replace Lemmata 5–6 by the following:

Lemma 7.

$$\pi^{\theta}(\mathcal{R}^m|\mathcal{K}) = \frac{k_\theta w_\theta}{\sum_{\theta'=1}^{r} k_{\theta'} w_{\theta'}} \cdot (1 + \mu(m)) \quad \textit{with} \quad \lim_{m\to\infty} |\mu(m)| = 0. \tag{93}$$

The proof of Theorem 1 can then be concluded by appealing to (63), hence

$$\lim_{m\to\infty} \frac{\pi^{\theta}(\mathcal{R}^m|\mathcal{K})}{\pi^{\theta'}(\mathcal{R}^m|\mathcal{K})} = \lim_{m\to\infty} \frac{\pi^{\theta}(\mathcal{R}^m)}{\pi^{\theta'}(\mathcal{R}^m)}, \tag{94}$$

and equations (85)–(87). Importantly, the presumption $|f_\theta(x) - f_\theta(\mathrm{M})| \leq cx^2$ for $x \in [\mathrm{M} - \varepsilon_1, \mathrm{M} + \varepsilon_1]$, which Lemma 5 required, is *not* needed by Lemma 7. It can hence be replaced in Theorem 1 by the simpler requirement that each f_θ is continuous in a neighborhood of M.

Finally, the assumption that only a finite number of different densities and weights are involved in the chain $\mathcal{R}^1 \subset \mathcal{R}^2 \subset \mathcal{R}^3 \subset \ldots$ could be loosened. However, it is critical that each representative's relative weight vanishes as $m \to \infty$ in order to apply Neyman's results; the asymptotic relation (15) fails to hold, for instance, for a chain with $w_1 = \sum_{j>1} w_j$. And because our result depends on a vanishing *relative* error, which is considered neither by Neyman (1982) nor Theorem 3,[36] it is similarly important that the aggregate relative weight of each type of representatives is bounded away from zero. For instance, with just one representative having weight $w_1 = 1$ and $\beta_2(m) = m-1$ ones with $w_2 = 2$ (see equation (13)), $\lim_{m\to\infty} \pi_1(\mathcal{R}^m) = \lim_{m\to\infty} \pi_j(\mathcal{R}^m) = 0$ for any $j \neq 1$ but the limit of $\pi_1(\mathcal{R}^m)/\pi_j(\mathcal{R}^m)$ may fail to exist.

B Proof of Theorem 2

The result easily follows from the definition of the Shapley value and the fact that the orderings which are induced by the realizations of the vectors $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ will coincide with a probability which tends to 1 as t approaches infinity. To see the latter, ignore any null events in which several ideal points or constituency shocks coincide and let $\hat{\varrho}(\mathbf{x})$ denote the permutation of $\{1, \ldots, m\}$ such that $x_i < x_j$ whenever $\hat{\varrho}(i) < \hat{\varrho}(j)$ for the real-valued vector $\mathbf{x} = (x_i)_{i\in\{1,\ldots,m\}}$. We then have:

Lemma 8. *For $i \in \{1, \ldots, m\}$ and $t > 0$ let $\lambda_i^t \equiv t \cdot \mu_i + \tilde{\varepsilon}_i$, where $\mu_1, \ldots, \mu_m$ and $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_m$ are all mutually independent random variables, $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_m$ have finite means and variances, and*

[36] See, however, Lindner and Machover (2004), where conditions very similar to ours are considered for the Shapley and Banzhaf values, and the related discussion by Lindner and Owen (2007).

$\mu_1, \ldots, \mu_m$ *have an identical bounded density. Then*

$$\lim_{t\to\infty} \Pr(\hat{\varrho}(\lambda^t) = \varrho) = \lim_{t\to\infty} \Pr(\hat{\varrho}(\mu) = \varrho) = \frac{1}{m!} \tag{95}$$

for each permutation ϱ *of* $\{1, \ldots, m\}$.

Proof. Let us denote the finite variance of $\tilde{\epsilon}_i$ by σ_i^2 and let $U \equiv (\max_i |\mathbf{E}[\tilde{\epsilon}_i]|)^3$. We can choose a real number k such that the bounded density function h of μ_i, with $i \in \{1, \ldots, m\}$, satisfies $h(x) \le k$ for all $x \in \mathbb{R}$. For any given realization $\mu_j = x$, the probability of the independent random variable μ_i assuming a value inside interval $(x - 4t^{-\frac{2}{3}}, x + 4t^{-\frac{2}{3}})$ is bounded above by $k \cdot 8t^{-\frac{2}{3}}$. We can infer that the event $\left\{|\mu_i - \mu_j| < 4t^{-\frac{2}{3}}\right\}$, which is identical to the event $\left\{|t\mu_i - t\mu_j| < 4t^{\frac{1}{3}}\right\}$, has a probability of at most $k \cdot 8t^{-\frac{2}{3}}$ for any $i \neq j \in \{1, \ldots, m\}$. And we can conclude from Chebyshev's inequality that $\Pr(|\tilde{\epsilon}_i - \mathbf{E}[\tilde{\epsilon}_i]| < t^{\frac{1}{3}})$ is at least $1 - \sigma_i^2 \cdot t^{-\frac{2}{3}}$. For $t \ge U$, we have $|\mathbf{E}[\tilde{\epsilon}_i]| \le t^{\frac{1}{3}}$; and if $|\tilde{\epsilon}_i - \mathbf{E}[\tilde{\epsilon}_i]| < t^{\frac{1}{3}}$ holds then also

$$2t^{\frac{1}{3}} > |\mathbf{E}[\tilde{\epsilon}_i]| + |\tilde{\epsilon}_i - \mathbf{E}[\tilde{\epsilon}_i]| \ge |\tilde{\epsilon}_i| \tag{96}$$

by the triangle inequality. Hence, the probability for (96) to hold when $t \ge U$ is $\Pr(|\tilde{\epsilon}_i| < 2t^{\frac{1}{3}}) \ge 1 - \sigma_i^2 \cdot t^{-\frac{2}{3}}$ for each $i \in \{1, \ldots, m\}$.

Now consider the joint event that (i) $|t\mu_i - t\mu_j| \ge 4t^{\frac{1}{3}}$ for *all* pairs $i \neq j \in \{1, \ldots, m\}$ and (ii) that $|\tilde{\epsilon}_i| < 2t^{\frac{1}{3}}$ for *all* $i \in \{1, \ldots, m\}$. In this event, the ordering of $\lambda_1^t, \ldots, \lambda_m^t$ is determined entirely by the realization of $t\mu_1, \ldots, t\mu_m$; in particular, $\hat{\varrho}(\boldsymbol{\lambda}^t) = \hat{\varrho}(\boldsymbol{\mu})$. Using the mutual independence of the considered random variables this joint event must have a probability of at least

$$\prod_{s=1}^{\binom{m}{2}} \left(1 - k \cdot 8t^{-\frac{2}{3}}\right) \cdot \prod_{i=1}^{m} \left(1 - \sigma_i^2 \cdot t^{-\frac{2}{3}}\right) \ge 1 - \left(8k\binom{m}{2} + \sum_{i=1}^{m} \sigma_i^2\right) \cdot t^{-\frac{2}{3}} \tag{97}$$

for $t \ge U$. The right hand side clearly tends to 1 as t approaches infinity. It hence remains to acknowledge that any ordering $\hat{\varrho}(\boldsymbol{\mu})$ has an equal probability of $1/m!$ because $\mu_1, \ldots, \mu_m$ are i.i.d. □

References

Alesina, A. and E. Spolaore (2003). *The Size of Nations*. Cambridge, MA: MIT Press.

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja (1992). *A First Course in Order Statistics*. New York, NY: John Wiley & Sons.

Balinski, M. L. and H. P. Young (2001). *Fair Representation – Meeting the Ideal of One Man, One Vote* (Second ed.). Washington, D.C.: Brookings Institution Press.

Banks, J. S. and J. Duggan (2000). A bargaining model of collective choice. *American Political Science Review 94*(1), 73–88.

Barberà, S. and M. O. Jackson (2006). On the weights of nations: Assigning voting weights in a heterogeneous union. *Journal of Political Economy 114*(2), 317–339.

Beisbart, C. and L. Bovens (2007). Welfarist evaluations of decision rules for boards of representatives. *Social Choice and Welfare 29*(4), 581–608.

Beisbart, C., L. Bovens, and S. Hartmann (2005). A utilitarian assessment of alternative decision rules in the Council of Ministers. *European Union Politics 6*(4), 395–419.

Black, D. (1948). The decisions of a committee using a special majority. *Econometrica 916*(3), 245–261.

Bouton, L. and M. Castanheira (2012). One person, many votes: Divided majority and information aggregation. *Econometrica 80*(1), 43–87.

Chamberlain, G. and M. Rothschild (1981). A note on the probability of casting a decisive vote. *Journal of Economic Theory 25*(1), 152–162.

Cho, S. and J. Duggan (2009). Bargaining foundations of the median voter theorem. *Journal of Economic Theory 144*(2), 851–868.

Coate, S. and B. Knight (2007). Socially optimal districting: A theoretical and empirical exploration. *Quarterly Journal of Economics 122*(4), 1409–1471.

Dahl, R. A. (1956). *A Preface to Democratic Theory*. Chicago, IL: University of Chicago Press.

Dubey, P. and L. Shapley (1979). Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research 4*(2), 99–131.

The Economist (2007, June 14th). The square root or death: The Germans fret that Poland may block a summit accord. Available at `http://www.economist.com/node/9341010/print`.

Esteban, J. and D. Ray (1994). On the measurement of polarization. *Econometrica 62*(4), 819–851.

Feddersen, T. and W. Pesendorfer (1997). Voting behavior and information aggregation in elections with private information. *Econometrica 65*(5), 1029–1058.

Feddersen, T. J. and W. Pesendorfer (1996). The swing voter's curse. *American Economic Review 86*(3), 408–424.

Feix, M. R., D. Lepelley, V. Merlin, J.-L. Rouet, and L. Vidu (2008). Majority efficient representation of the citizens in a federal union. Mimeo, Université de la Réunion, Université de Caen, and Université d'Orléans.

Felsenthal, D. and M. Machover (1998). *The Measurement of Voting Power – Theory and Practice, Problems and Paradoxes*. Cheltenham: Edward Elgar.

Felsenthal, D. and M. Machover (1999). Minimizing the mean majority deficit: The second square-root rule. *Mathematical Social Sciences 37*(1), 25–37.

Felsenthal, D. and M. Machover (2001). The Treaty of Nice and qualified majority voting. *Social Choice and Welfare 18*(3), 431–464.

Felsenthal, D. and M. Machover (2004). Analysis of QM rules in the draft Constitution for Europe proposed by the European Convention, 2003. *Social Choice and Welfare 23*(1), 1–20.

Fidrmuc, J., V. Ginsburgh, and S. Weber (2009). Voting on the choice of core languages in the European Union. *European Journal of Political Economy 25*(1), 56–62.

Gelman, A., J. N. Katz, and J. Bafumi (2004). Standard voting power indexes don't work: An empirical analysis. *British Journal of Political Science 34*(1133), 657–674.

Gelman, A., J. N. Katz, and F. Tuerlinckx (2002). The mathematics and statistics of voting power. *Statistical Science 17*(4), 420–435.

Gerber, E. R. and J. B. Lewis (2004). Beyond the median: Voter preferences, district heterogeneity, and political representation. *Journal of Political Economy 112*(6), 1364–1383.

Gersbach, H. (2005). *Designing Democracy: Ideas for better rules*. Berlin: Springer.

Gersbach, H. (2009). Democratic mechanisms. *Journal of the European Economic Association 7*(6), 1436–1469.

Grofman, B. and S. L. Feld (2005). Thinking about the political impacts of the Electoral College. *Public Choice 123*(1-2), 1–18.

Gul, F. and W. Pesendorfer (2010). Strategic redistricting. *American Economic Review 100*(4), 1616–1641.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association 58*(301), 13–30.

Jelnov, A. and Y. Tauman (2012). Voting power and proportional representation of voters. Mimeo, Tel Aviv University and SUNY, Stony Brook.

Kaniovski, S. (2008). The exact bias of the Banzhaf measure of power when votes are neither equiprobable nor independent. *Social Choice and Welfare 31*(2), 281–300.

Kirsch, W. (2007). On Penrose's square-root law and beyond. *Homo Oeconomicus 24*(3-4), 357–380.

Kirsch, W. and J. Langner (2011). Invariably suboptimal: An attempt to improve the voting rules of the Treaties of Nice and Lisbon. *Journal of Common Market Studies 49*(6), 1317–1338.

Koriyama, Y., J.-F. Laslier, A. Macé, and R. Treibich (2012). Optimal apportionment. Cahier no. 2012-1, École Polytechnique ParisTech.

Kurz, S. (2012). On the inverse power index problem. *Optimization 61*(8), 989–1011.

Laruelle, A. and F. Valenciano (2008a). Bargaining in committees of representatives: The 'neutral' voting rule. *Journal of Theoretical Politics 20*(1), 93–106.

Laruelle, A. and F. Valenciano (2008b). *Voting and Collective Decision-Making*. Cambridge: Cambridge University Press.

Le Breton, M., M. Montero, and V. Zaporozhets (2012). Voting power in the EU Council of Ministers and fair decision making in distributive politics. *Mathematical Social Sciences 63*(2), 159–173.

Leech, D. and R. Leech (2009). Reforming IMF and World Bank governance: In search of simplicity, transparency and democratic legitimacy in the voting rules. Warwick Economic Research Papers 914, University of Warwick.

Lindner, I. and M. Machover (2004). L. S. Penrose's limit theorem: Proof of some special cases. *Mathematical Social Sciences 47*(1), 37–49.

Lindner, I. and G. Owen (2007). Cases where the Penrose limit theorem does not hold. *Mathematical Social Sciences 53*(3), 232–238.

Maaser, N. and S. Napel (2007). Equal represenation in two-tier voting systems. *Social Choice and Welfare 28*(3), 401–420.

Maaser, N. and S. Napel (2012a). The mean voter, the median voter, and welfare-maximizing voting weights. Discussion Paper 59, Public Choice Research Centre, University of Turku.

Maaser, N. and S. Napel (2012b). A note on the direct democracy deficit in two-tier voting. *Mathematical Social Sciences 63*(2), 174–180.

Miller, N. R. (2009). A priori voting power and the U.S. Electoral College. *Homo Oeconomicus 26*(3-4), 341–380.

Miller, N. R. (2012). Why the Electoral College is good for political science (and public choice). *Public Choice 150*(1-2), 1–25.

Monderer, D. and D. Samet (2002). Variations on the Shapley value. In R. J. Aumann and S. Hart (Eds.), *Handbook of Game Theory*, Volume 3, Chapter 54. Amsterdam: North-Holland.

Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice 35*(4), 437–455.

Myerson, R. B. (2000). Large Poisson games. *Journal of Economic Theory 94*(1), 7–45.

Napel, S. and M. Widgrén (2004). Power measurement as sensitivity analysis – A unified approach. *Journal of Theoretical Politics 16*(4), 517–538.

Nehring, K. and C. Puppe (2007). The structure of strategy-proof social choice – Part I: General characterization and possibility results on median spaces. *Journal of Economic Theory 135*(1), 269–305.

Neyman, A. (1982). Renewal theory for sampling without replacement. *Annals of Probability 10*(2), 464–481.

Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society 109*(1), 53–57.

Penrose, L. S. (1952). *On the Objective Study of Crowd Behaviour*. London: H. K. Lewis & Co.

Riker, W. H. (1986). The first power index. *Social Choice and Welfare 3*(4), 293–295.

Shapiro, N. Z. and L. S. Shapley (1978). Values of large games, I: A limit theorem. *Mathematics of Operations Research 3*(1), 1–9.

Shapley, L. S. (1953). A value for *n*-person games. In H. W. Kuhn and A. W. Tucker (Eds.), *Contributions to the Theory of Games*, Volume II, pp. 307–317. Princeton, NJ: Princeton University Press.

Shapley, L. S. and M. Shubik (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review 48*(3), 787–792.

Snyder, James M., Jr., M. T. Ting, and S. Ansolabehere (2005). Legislative bargaining under weighted voting. *American Economic Review 95*(4), 981–1004.

Strömberg, D. (2008). How the Electoral College influences campaigns and policy: The probability of being Florida. *American Economic Review 98*(3), 769–807.

Tiebout, C. (1956). A pure theory of local expenditure. *Journal of Political Economy 64*(5), 416–424.