Impartial Resolution of Random Conflicts*

Niels Boissonnet†
Department of Economics, IMW, and JICE, Bielefeld University

Yves Breitmoser‡
Department of Economics, IMW, and JICE, Bielefeld University

A planner repeatedly chooses between two randomly drafted allocations, and whenever people disagree she must side with someone. We test welfare measures by a minimal requirement of impartiality: no one should win almost every conflict. Standard indices fail it. Since they ignore which proposals are drafted, a drafting process tailored to them can make one person a near-dictator. More strikingly, under a given process that treats people identically, the egalitarian-equivalent index lets the person who needs goods together more win almost every conflict under leximin once goods are many, and other rules only change who wins; under the same process, scoring each person by how her bundle ranks among the drafts splits the conflicts evenly. The freedom behind this is general: for any anonymous Paretian rule, some rescaling of unchanged preferences makes it side with a designated person almost always. We then derive the rankings that impartiality requirements single out. If a randomly drafted bundle, given to one person alone, must beat any benchmark equally often whoever receives it, a Paretian ranking that uses only frequencies is an anonymous rule applied to percentiles, each person's share of drafts that she ranks no higher. Hammond equity selects their leximin, and consistency across drafting processes selects their sum, a continuous Borda rule that reward models trained on pairwise comparisons already compute. These rankings treat people alike in expectation under every rule; in actual decisions they do so only under a condition on preferences and proposals that no scale can create.

1. INTRODUCTION

A planner must repeatedly choose between two allocations drafted at random. Some choices are easy: one allocation is better for everyone, and the Pareto principle settles the matter. Many are not. When one person prefers the first allocation and another the second, the two are in *conflict*, and whatever the planner does, she sides with one of them against the other. When can such a planner be called impartial? A minimal requirement is that no one should win almost every conflict. Call a person's *favor rate* the probability that the planner sides with her, given a conflict that the planner resolves; a person whose favor rate is close to one is a *near-dictator*. The planner still respects unanimity, but follows her whenever people disagree. We read the exclusion of near-dictators as a normative test and apply it to the tools of welfare economics.

These tools split the planner's choice in two steps. First, a welfare index puts people's situations on a common scale. Second, an aggregation rule, such as leximin, the utilitarian sum or the Nash product, then trades one person's welfare against another's. The index carries the interpersonal comparisons, which ordinal preferences leave open. We show that the standard indices fail the test. We then derive the rankings that impartiality requirements single out, and what they guarantee.

The standard indices measure welfare in resources. The egalitarian-equivalent index scores a bundle by the fraction of the whole endowment that its recipient finds equally good (Pazner and Schmeidler, 1978; Fleurbaey and Maniquet, 2011). Money-metric utility scores it by the cost of its cheapest equivalent at reference prices (Samuelson, 1974), and opportunity-equivalent indices by the size of an equivalent reference opportunity set (Piacquadio, 2017). The motivating idea is that

* We are grateful for all comments. Financial support by Verbundinitiative InChange funded by MKW NRW is greatly appreciated.
† Corresponding author. Department of Economics, IMW, and JICE, Bielefeld University
‡ Department of Economics, IMW, and JICE, Bielefeld University

differences in tastes are not, by themselves, grounds for priority: nobody should gain or lose standing merely because of what she likes. When preferences are homothetic, all these indices coincide.

Part of their failure is expected. These indices ignore which proposals are drafted, so a drafting process tailored to the index can make one person a near-dictator under leximin, even a process that treats everyone symmetrically and gives everyone nearly the same bundle (Proposition 1). But a planner does not choose the drafting process, and ignoring it does not by itself imply failure under a given one. The striking failure concerns a process that is given. Take two people who differ only in how much they need goods together, and a process that splits each good independently and treats them identically. As goods multiply, the egalitarian-equivalent index lets the one who needs goods together more win almost every conflict that leximin resolves, and leximax hands almost every conflict to the other (Theorem 2). Their resources have the same distribution. A difference in tastes is all that separates them, and it decides who prevails. More people do not dilute the effect: among any number, the one who needs goods together most wins almost every conflict that leximin resolves.

Changing the rule changes the winner, not the outcome. When one person needs goods in fixed proportions and the other cares only about her average share, the utilitarian rule sides almost always with the second, and the Nash rule with the first; both converge at a speed close to the inverse square root of the number of goods, leximin exponentially. Other indices fail through the reference they fix. Equivalent income reverses the egalitarian-equivalent results, because its reference bundle stretches one person's range far below the other's (Proposition 2). With non-homothetic preferences, the reference prices of the money metric alone can move a person's favor rate from 0.97 to 0.03.

Behind these failures lies a general freedom. To rank allocations, an anonymous rule needs welfare scores, and ordinal preferences do not fix them; we call the choice of scores a *calibration*. The calibration matters only in conflicts: if one proposal helps someone and harms no one, the Pareto principle settles the comparison under every calibration. When people disagree, the calibration can decide. For any anonymous rule that respects Pareto improvements, some calibration of unchanged preferences makes the rule side with one designated person in almost every conflict (Theorem 1). The rule need not be a weighted sum, continuous or egalitarian. Relative utilitarianism, which normalizes each person's von Neumann–Morgenstern utility between her best and worst alternatives, is a case in point: every calibration is such a utility for some attitudes toward risk. An anonymous rule cannot favor a person by name, but it can favor her through the numbers that represent her welfare. This is the sense in which anonymity is not impartiality.

The natural repair is to let the index account for how often bundles are drafted. Done in one way, this works in expectation: score each person by her *percentile*, the share of drafted proposals that give her a bundle she ranks no higher. Like an opportunity-equivalent index, a percentile measures a bundle against a reference; here the reference is what the drafting process might give its recipient. Done in any other way, the repair fails for some rule (Theorem 6). In the two-person economies of Theorem 2, percentiles also pass the test in actual decisions. Under the same drafting processes, as goods multiply, they give each person half of the conflicts under leximin, leximax, the utilitarian rule and the Nash rule, while the egalitarian-equivalent index hands almost all of them to one person (Theorem 7).

Which rankings do impartiality requirements single out? We answer without choosing a scale first, by placing requirements on the social ranking of allocations itself. The central one, *equal standing*, concerns how the ranking compares people. Draw a proposal at random and give one person, alone, the bundle it assigns her. The chance that society prefers this to a fixed benchmark must not depend on who that person is. The egalitarian-equivalent leximin ranking fails it in the economies of Theorem 2. A second requirement, *frequency independence*, restricts the information the ranking may use: about each position, only how often the drafts would place each person below it. With strong Pareto and anonymity, the two requirements make the ranking a fixed anonymous rule applied to percentiles (Theorem 3). A percentile of 0.7 means that she likes her bundle at least as

much as what she would receive in 70% of proposals. Percentiles are not assumed. Equal standing forces them when society compares one person's bundle with another's, and frequency independence carries that comparison over to whole allocations. This second step is where the restriction bites. Without it, a ranking that adds a small resource term to the sum of percentiles meets all the other requirements that, below, single out that sum (Section 3.2).

One further requirement then selects a criterion. Hammond equity selects *percentile leximin*, which gives priority to the person worst placed relative to what she might have received (Corollary 1). Consistency across drafting processes selects *percentile utilitarianism*, the sum of percentiles (Theorem 4). The requirement is simple: if society prefers one allocation to another whichever of two offices drafts the proposals, it keeps this preference when both offices draft in given proportions. The argument is Harsanyi's, with drafting processes in the role of lotteries. The sum of percentiles is the expected number of people for whom an allocation is at least as good as a random draft, a continuous Borda count.

The same sum is already in use. Language models are fine-tuned with reward models learned from pairwise comparisons: two responses to a prompt are sampled, and an annotator states which one she prefers. When annotators disagree, the learned reward ranks a finite set of responses by their Borda count over annotators (Siththaranjan et al., 2024). Draw the responses from a law instead, and the count becomes the sum of the annotators' percentiles (Section 4.9). What we establish about percentile rankings, including what they cannot guarantee, therefore bears on a technology in wide use.

What do these rankings guarantee? The test excludes near-dictators; its natural strengthening asks that favor rates be equal, and we measure the rankings against it. Percentile rankings equalize *expected* favor rates, computed as if each person's welfare were drawn from her own prospects, independently of the others'; percentile scales are the only scales that do so under every anonymous rule (Theorem 6). Actual favor rates are another matter. A rule compares the scores that occur together in a proposal. When the two scores are 0.2 and 0.8, one person may be more likely to hold the lower value, although both have the same overall distribution. With independently drawn proposals, a scale equalizes favor rates under every anonymous rule if and only if it is a common transformation of percentiles and, given the percentile values that occur together, every person is equally likely to hold each value (Theorem 5). The second condition concerns preferences and proposals, and no scale can create it. It can fail merely because preferences differ, even when proposals treat people identically, and it is not a pathology of unusual rules (Theorem O1). When no scale works for every rule, a planner can still choose one for the rule she uses (Theorem O2).

The two kinds of measurement express two neutralities toward tastes. Resource indices give no one more goods because of what she likes; percentiles give no one a better standing. Theorem 2 shows that they can conflict, and ordinal information cannot decide between them. If a person's need for goods in fixed proportions is a need rather than a taste, placing her low may be justified, and equal standing is the wrong benchmark. Our results do not reject resource indices. They show what impartiality between tastes requires, and what is given up when welfare is measured in resources.

The results separate two tasks that welfare analysis often merges: measuring each person's welfare, and deciding how society trades one person's welfare against another's. Information about each person separately can put scores on a common footing; only information about which welfare positions occur together can show whether a rule built on those scores treats people impartially. Public appraisal draws the same line. The UK *Green Book* separates appraisal advice from policy objectives and asks how costs and benefits fall on different groups and places (HM Treasury, 2026), and distributional choices enter cost-benefit analysis (Sunstein, 2024). Our model does not evaluate that guidance or replace monetary appraisal; it isolates what ordinal rankings and a specified proposal process can establish when comparable preference strengths are unavailable.

Related literature. Our informational restriction goes back to the foundations of welfare eco-

nomics: ordinal preferences alone leave interpersonal comparisons undetermined (Sen, 1970, 1977; d'Aspremont and Gevers, 1977; Roberts, 1980). The indices we test respond by building comparisons from reference bundles, prices or opportunities and from fairness requirements (Pazner and Schmeidler, 1978; Fleurbaey and Maniquet, 2011; Piacquadio, 2017); equivalent income fixes a reference for all goods but one (Fleurbaey and Blanchet, 2013; Decancq et al., 2015). Money-metric utility depends on the reference prices once preferences are not homothetic (Blackorby and Donaldson, 1988); our test makes this dependence measurable. These approaches share our premise that differences in tastes are not by themselves grounds for priority, but they draw a different consequence from it. Our results do not reject them or other compensatory welfare criteria (Fleurbaey and Zuber, 2021); they show that neutrality in resources and neutrality in standing are distinct requirements.

Another response enriches the information. Harsanyi (1955) uses preferences over lotteries. Dhillon and Mertens (1999) characterize relative utilitarianism, the sum of utilities normalized by each person's best and worst alternatives, and Karni and Weymark (2024) derive it from the impartial treatment of conflicting lottery preferences of equal strength. Sprumont (2013) characterizes relative egalitarianism, the leximin of the same normalized utilities. Percentile utilitarianism and percentile leximin are their counterparts with the proposal law in place of the two extreme alternatives, and equal standing counts how often people are placed above a benchmark rather than weighing intensities.

Statistical normalization puts utility scales on a common footing (MacAskill, 2016); MacAskill et al. (2020) argue that variance normalization gives everyone "equal say". Unlike resource indices, it depends on the proposal law, and in our examples it comes close to equal favor rates. Equal standing is a different requirement, and it selects percentiles instead. Quantile fairness compares a bundle with what its recipient might receive under other allocations (Babichenko et al., 2024), and the sum of percentiles is a Borda count relative to the proposal law (Young, 1974); Section 3.3 discusses what this relation implies.

A recent literature treats the training of reward models as a problem of preference aggregation (Conitzer et al., 2024; Ge et al., 2024). Ge et al. (2024) show that the standard Bradley–Terry–Luce estimator violates Pareto optimality and pairwise majority consistency, and Chakraborty et al. (2024) maximize the smallest reward across annotator groups. We add two observations. The standard estimator computes the continuous analogue of a criterion that impartiality requirements single out in a distributive setting. And any rule that aggregates group rewards must first put them on a common scale, which is the calibration problem of this paper; percentiles are the scale that gives every group the same prospects whatever the rule.

Favor rates are closest to voting *success*, not pivotal power (Rae, 1969; Laruelle and Valenciano, 2005). Unlike quantitative versions of Arrow's theorem (Kalai, 2002; Mossel, 2012), our near-dictatorship results keep anonymity and strong Pareto and vary only measurement.

Throughout, the two proposals are drawn independently and their law stays fixed when scores or rules vary. Our characterizations make no claim about other pair laws, such as a reform paired with its status quo. The paper thus identifies what impartial measurement and impartial rankings can achieve in a specified decision problem, not an institution-independent welfare index.

2. WELFARE RANKINGS AND IMPARTIALITY

Society ranks the proposals submitted to it; it does not search for the best possible allocation. The model keeps three objects apart: which bundles can be allocated, how proposals arise, and how society ranks a submitted pair. A change of welfare scale then leaves people's preferences and the proposals untouched.

2.1. Allocations and proposals

A bundle lists an agent's shares of the available goods, and an allocation lists one bundle for each agent. Formally, let Ω be a nonempty, possibly infinite, set of goods. The bundle space is $X = [0, 1]^{\Omega}$

with the product topology, and $N = \{1, \dots, n\}$ is the set of agents, with $n \geq 2$. An allocation $\mathbf{x} = (x_1, \dots, x_n)$ assigns each agent a bundle. The vectors 0_Ω and 1_Ω contain none and all of each good. We write $\mathbf{e}_i(x)$ for the allocation that gives bundle x to agent i and nothing to everyone else.

The set of feasible allocations is[1]

$$\mathcal{X} = \left\{ \mathbf{x} \in X^n : \textstyle\sum_i x_i \leq 1_\Omega \right\}.$$

Each agent i has a continuous, monotone weak order $\succsim_i$ over her own bundles, with strict part $\succ_i$ and indifference part[2] $\sim_i$.

Our approach relies on a new object, the *pair law*: a probability law π on $\mathcal{X} \times \mathcal{X}$ that describes how often each pair of feasible allocations comes up for a decision. Offices, committees or outside experts draft proposals; two of them reach the floor, and the planner must adopt one without writing a third. The law π records which pairs of proposals reach the floor together. The planner ranks the two proposals before her; she need not look for a better allocation outside them, and if someone submitted one, the Pareto requirement below would apply to it. The pair law counts rejected proposals too. We hold it fixed when the welfare scale or the way of ranking proposals changes.

Assumption 1. *There is a probability law μ on $\mathcal{X}$ such that*

$$\pi = \mu \otimes \mu.^{3}$$

The pair law π has full support on $\mathcal{X} \times \mathcal{X}$, and $\pi\{(\mathbf{x}, \mathbf{y}) : x_i \sim_i z\} = 0$ for every agent i and every bundle $z \in X$.

The first condition requires that the two proposals be independent draws from one law μ, the *proposal law*.[4] One proposal then carries no information about the other, and the proposal law alone fixes the pair law. Nor does the order of the two proposals matter: $(\mathbf{x}, \mathbf{y})$ and $(\mathbf{y}, \mathbf{x})$ have the same law. We maintain Assumption 1 throughout the paper. Independence excludes a reform paired with its status quo, strategic submissions and drafts that respond to a common shock; Section 5 returns to these cases.

The second condition requires that every pair of feasible allocations, or one close to it, can come to a vote, and that no welfare level of an agent occurs with positive probability in a proposal (*atomless welfare levels*). Both conditions pass to the proposal law: μ has full support on $\mathcal{X}$, and $\mu\{\mathbf{x} : x_i \sim_i z\} = 0$.[5]

2.2. Scores, anonymity and Pareto agreement

A *calibration* consists of continuous, onto representations $\mathbf{m} = (m_1, \dots, m_n)$, where $m_i : X \to [0, 1]$ represents $\succsim_i$. Its score vector is $\mathbf{m}(\mathbf{x}) = (m_1(x_1), \dots, m_n(x_n))$. The common endpoints satisfy $m_i(0_\Omega) = 0$ and $m_i(1_\Omega) = 1$. They fix ranges, not interpersonal comparisons: an increasing

[1] Only the results on the Dirichlet family and the verification of condition (T) use this particular form of $\mathcal{X}$. The other results hold for any feasible set contained in $\mathcal{X}$ that has four properties: it is compact; it contains every $\mathbf{e}_i(1_\Omega)$, so each agent can receive everything; giving someone less never makes an allocation infeasible (comprehensiveness); and feasibility does not depend on who receives a given bundle (permutation invariance). Their proofs use nothing else. Such sets describe, for instance, indivisible goods, each of which must go entirely to one agent.

[2] Monotonicity means that $x \geq y$ implies $x \succsim_i y$, and that an increase in every coordinate is strictly preferred. Preferences stay fixed unless we say otherwise.

[3] When Ω is uncountable, $\mathcal{X}$ is compact but not metrizable, and the Borel sets of $\mathcal{X} \times \mathcal{X}$ need not be generated by products of Borel sets. We then take μ to be a Radon measure and $\mu \otimes \mu$ its Radon product, which is defined on every Borel set of $\mathcal{X} \times \mathcal{X}$ and gives each product $G \times E$ the probability $\mu(G)\mu(E)$. For countable Ω no such care is needed.

[4] This law is unique: $\mu(E) = \pi(E \times \mathcal{X}) = \pi(\mathcal{X} \times E)$ for every Borel set $E \subseteq \mathcal{X}$.

[5] Every nonempty open subset of $\mathcal{X} \times \mathcal{X}$ contains a product $G \times H$ of nonempty open sets, and $\pi(G \times H) = \mu(G)\mu(H)$; conversely, $\mu(G) = \pi(G \times \mathcal{X})$. Likewise, $\pi\{(\mathbf{x}, \mathbf{y}) : x_i \sim_i z\} = \mu\{\mathbf{x} : x_i \sim_i z\}$.

transformation of one agent's scores preserves her ranking but can change how her scores compare with others'.

An *aggregator* is a weak order R on $[0,1]^n$—with strict part P and indifference part I—that satisfies strong Pareto, that is, $\mathbf{a} \geq \mathbf{b}$ implies $\mathbf{a}\,\mathrm{R}\,\mathbf{b}$, strictly if $\mathbf{a} \neq \mathbf{b}$; and anonymity, that is, $\mathbf{a}\,\mathrm{I}\,\sigma\mathbf{a}$ for every permutation σ. A calibration and an aggregator together rank the proposals: the calibration turns each proposal into a score vector, and the aggregator ranks the score vectors. Formally, $(\mathbf{m}, \mathrm{R})$ ranks $\mathbf{x}$ strictly above $\mathbf{y}$, written $\mathbf{x} \succ^{\mathbf{m},\mathrm{R}} \mathbf{y}$, when $\mathbf{m}(\mathbf{x})\,\mathrm{P}\,\mathbf{m}(\mathbf{y})$, and declares a *social tie* when $\mathbf{m}(\mathbf{x})\,\mathrm{I}\,\mathbf{m}(\mathbf{y})$. We write $\succ$ for $\succ^{\mathbf{m},\mathrm{R}}$ when the calibration and the rule are clear. Scores are continuous, so the set of pairs with $\mathbf{x} \succ^{\mathbf{m},\mathrm{R}} \mathbf{y}$ is Borel whenever the strict part of R is; we assume this of every aggregator, and every explicit rule below satisfies it.

Leximin compares the lowest scores first, then the next lowest if necessary; leximax reverses that priority; utilitarianism sums the scores. Each rule adds a distributive judgment to the calibration. For instance, applying one increasing transformation to every score preserves leximin and leximax, but it can change a utilitarian ranking, because it changes differences between levels.

Anonymity treats scores symmetrically, not bundles. Swapping two numbers in a score vector leaves an anonymous rule indifferent. Swapping two physical bundles between people with different preferences need not: the swap changes the scores as well as their owners. Anonymity therefore does not declare identical material bundles equally good for everyone. Conversely, an unequal social choice need not violate anonymity. Leximin may favor the lower-scoring person precisely because it applies the same priority to whoever occupies that position. What remains open is how ordinal preferences determine who counts as lower-scoring in the first place.

Single-recipient allocations show what an interpersonal scale does. In $\mathbf{e}_i(x)$, only agent i receives a nonzero bundle. Anonymity and strong Pareto therefore rank $\mathbf{e}_i(x)$ against $\mathbf{e}_j(y)$ by whether $m_i(x)$ exceeds $m_j(y)$, with indifference at equality. A rescaling that preserves all these comparisons must give bundles with the same old score the same new score, whoever receives them. Since every score function covers $[0,1]$, only a common increasing transformation can do so. Section 3 derives this split from requirements on the ranking of allocations.

2.3. Conflicts and favor rates

A conflict occurs when one person prefers proposal $\mathbf{x}$ and another prefers $\mathbf{y}$. When $(\mathbf{m}, \mathrm{R})$ puts $\mathbf{x}$ strictly above $\mathbf{y}$, whoever prefers $\mathbf{x}$ is favored and whoever prefers $\mathbf{y}$ is overruled; when it declares a tie, neither counts. A comparison on which everyone agrees is not a conflict, and strong Pareto settles it under every calibration. For any number of agents, define

$$F_i(\mathbf{m}, \mathrm{R}) = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \succ^{\mathbf{m},\mathrm{R}} \mathbf{y},\ x_i \succ_i y_i,\ y_j \succ_j x_j \text{ for some } j \neq i\},$$
$$O_i(\mathbf{m}, \mathrm{R}) = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \succ^{\mathbf{m},\mathrm{R}} \mathbf{y},\ y_i \succ_i x_i,\ x_j \succ_j y_j \text{ for some } j \neq i\}.$$

Agent i is favored on F_i and overruled on O_i: in both sets someone disagrees with her, and the rule picks her preferred proposal in the first and the other one in the second. Her *favor rate* is

$$f_i(\mathbf{m}, \mathrm{R}) = \frac{\pi(F_i(\mathbf{m}, \mathrm{R}))}{\pi(F_i(\mathbf{m}, \mathrm{R})) + \pi(O_i(\mathbf{m}, \mathrm{R}))},$$

whenever the denominator is positive. We drop the arguments of f_i, F_i and O_i when the calibration and the rule are clear from the context.

The calibration enters these sets in one place only. The conditions $x_i \succ_i y_i$ and $y_j \succ_j x_j$ involve preferences alone: they decide which pairs are conflicts, and no calibration changes them. The condition $\mathbf{x} \succ^{\mathbf{m},\mathrm{R}} \mathbf{y}$, that is, $\mathbf{m}(\mathbf{x})\,\mathrm{P}\,\mathbf{m}(\mathbf{y})$, depends on where the calibration places each agent and on how the rule compares the two score vectors: it decides who wins. A change of scale therefore changes who wins each conflict, not which pairs are conflicts. Under Assumption 1 the proposal law fixes the pair law, so favor rates depend on μ, the calibration and the rule only.

A favor rate records who wins the pairs that actually reach the floor. These pairs link the agents' fortunes: a proposal generous to one person is often stingy with another, and the rule compares the positions that occur together. What does the calibration do on its own? To see it, compute the same rate in a fictitious society where each person's score is drawn from her own distribution, independently of everyone else's. Formally, agent i's *score distribution* ν_i is the law of $m_i(x_i)$ when $\mathbf{x}$ is drawn from μ, and $\bar{\nu} = \nu_1 \otimes \cdots \otimes \nu_n$ is the law of a score vector whose coordinates are independent draws from these distributions. Draw two score vectors $\mathbf{a}, \mathbf{b}$ independently from $\bar{\nu}$; their joint law is $\bar{\pi} = \bar{\nu} \otimes \bar{\nu}$. Agent i prefers $\mathbf{a}$ when $a_i > b_i$, and

$$\bar{F}_i(\mathbf{m}, \mathrm{R}) = \{(\mathbf{a}, \mathbf{b}) : \mathbf{a}\,\mathrm{P}\,\mathbf{b},\ a_i > b_i,\ b_j > a_j \text{ for some } j \neq i\},$$
$$\bar{O}_i(\mathbf{m}, \mathrm{R}) = \{(\mathbf{a}, \mathbf{b}) : \mathbf{a}\,\mathrm{P}\,\mathbf{b},\ a_i < b_i,\ a_j > b_j \text{ for some } j \neq i\}.$$

Her *expected favor rate* is

$$\bar{f}_i(\mathbf{m}, \mathrm{R}) = \frac{\bar{\pi}(\bar{F}_i(\mathbf{m}, \mathrm{R}))}{\bar{\pi}(\bar{F}_i(\mathbf{m}, \mathrm{R})) + \bar{\pi}(\bar{O}_i(\mathbf{m}, \mathrm{R}))},$$

whenever the denominator is positive. It is the rate one expects from each person's prospects alone, before knowing how the proposals pair their fortunes. The calibration and the primitives enter it through the score distributions only. How one person's position varies with another's within a proposal leaves it unchanged, and the actual rate differs from it by exactly what this link adds.

How far can the calibration move these rates? As far as one likes. For any aggregator, some calibration makes the planner side with a designated person in almost every conflict, actual or expected.

Theorem 1. *Let* R *be any aggregator and* $i \in N$. *There are calibrations* $\mathbf{m}^{(K)}$, $K \geq 1$, *with* $f_i(\mathbf{m}^{(K)}, \mathrm{R}) \to 1$ *and* $\bar{f}_i(\mathbf{m}^{(K)}, \mathrm{R}) \to 1$ *as* $K \to \infty$. *Moreover, on a set of pairs whose* π*-probability tends to* 1, $\mathbf{x} \succ^{\mathbf{m}^{(K)}, \mathrm{R}} \mathbf{y}$ *if and only if* $x_i \succ_i y_i$*; and on a set of score pairs whose* $\bar{\pi}$*-probability tends to* 1, $\mathbf{a}\,\mathrm{P}\,\mathbf{b}$ *if and only if* $a_i > b_i$.

A suitable choice of scales makes the rule follow agent i's own ranking in almost every comparison.[6] The aggregator, the preferences and the proposal law stay fixed; only the numerical representations change, so no one receives a larger coefficient in a welfare sum. The theorem covers leximin and abstract, possibly discontinuous aggregators. In every case, each person's new scores are an increasing transformation of her old ones, so her preference ranking is preserved. The second conclusion locates the favoritism. The same scales work in the fictitious society where no one's fortune depends on anyone else's, so the scale alone produces it; the proposals need not help. Anonymity does not prevent this. Do the calibrations that welfare economics uses avoid it?

2.4. Standard calibrations put to the test

Welfare economics offers calibrations that use preferences and resources only. Two are standard. The *egalitarian-equivalent* calibration scores a bundle by the fraction of the whole endowment that its recipient finds equivalent to it: $m_i(x) = t$ when $x \sim_i t\, 1_\Omega$ (Pazner and Schmeidler, 1978; Fleurbaey and Maniquet, 2011). The *money-metric* calibration fixes reference prices $p \gg 0$ for the goods and scores a bundle by the cost $e_i(p, x)$ of the cheapest bundle its recipient finds at least as good, divided

[6] The proof first finds finitely many score levels at which an increase in the designated agent's score is decisive, whatever the permitted changes in the other scores. It then places most of her score probability near these levels, and most of the other agents' scores in suitable narrow intervals. As the number of levels grows, comparisons that go against her become rare. The construction keeps full support and needs no continuity of the aggregator. Unlike quantitative Arrow results (Kalai, 2002; Mossel, 2012), the argument keeps anonymity and strong Pareto and varies only measurement. Nor does it overturn the Pareto benchmark: unanimous improvements already agree with the designated person. Its force lies elsewhere. The exceptions can be made arbitrarily rare *among conflicts*.

by $e_i(p, 1_\Omega)$ (Samuelson, 1974). Both measure welfare in resources: a person's score is the quantity of goods, along the equal ray or at the reference prices, that would leave her as well off. Neither looks at what others receive, nor at which proposals are drafted.

The two calibrations often coincide. Suppose that agent i's preferences are homothetic, and represent them by the function u_i that is homogeneous of degree one with $u_i(1_\Omega) = 1$. Then $x \sim_i u_i(x)\, 1_\Omega$, so the egalitarian-equivalent score is $u_i(x)$. The cheapest bundle at least as good as x costs $u_i(x)$ times the cheapest bundle at least as good as 1_Ω, so the money-metric score is $u_i(x)$ as well, whatever the reference prices.[7] The two differ only when preferences are not homothetic, and the money metric then depends on the reference prices (Blackorby and Donaldson, 1988).

The same argument covers every index that rescales a reference. Fix a compact set B of bundles that contains a strictly positive bundle, and score x by the factor τ such that x is as good as the best bundle in τB, divided by the factor for 1_Ω. A reference ray (Fleurbaey and Maniquet, 2011) is the case of a single bundle, the money metric that of a budget set, and the opportunity-equivalent index of Piacquadio (2017) that of a general opportunity set. With homothetic preferences, the best bundle in τB is worth τ times the best bundle in B, so every such index equals u_i, whatever B. What we establish below for the egalitarian-equivalent calibration and homothetic preferences therefore holds for the whole family.

How often does a rule built on these calibrations side with each person when interests conflict? Under leximin, the answer can be: almost always with the same person. Part of this is not surprising. None of these indices looks at the proposals, so none can adapt to them, and a drafting process tailored to the index exploits this.

Proposition 1. *Let* $\mathbf{m}$ *be any calibration, let* R *be leximin, and let* A *be an agent. Suppose that some bundle* z *with* $z(\omega) < \frac{1}{n}$ *for every good satisfies* $m_A(z) < m_j(z)$ *for every* $j \neq A$. *There are proposal laws* μ_k, *each invariant to every permutation of the agents and with* $\mu_k \otimes \mu_k$ *satisfying Assumption 1, such that* $f_A(\mathbf{m}, \mathrm{R}) \to 1$ *and* $\bar{f}_A(\mathbf{m}, \mathrm{R}) \to 1$ *as* $k \to \infty$. *With* $n = 2$, *the same laws give* $f_A \to 0$ *and* $\bar{f}_A \to 0$ *under leximax. In particular, with* $n = 2$, *a calibration chosen without regard to the proposal law makes one agent a near-dictator under some symmetric law unless the two agents rank alike all bundles below* $\frac{1}{2}\, 1_\Omega$.

The laws μ_k concentrate on allocations that give every agent a bundle close to z: the drafts treat the agents symmetrically and give them nearly the same bundle. The index still places A lowest in almost every proposal, because her taste converts z into a smaller score, and leximin protects the lowest. The proof uses only the continuity of the scores, so the proposition applies to every index of this section. For the egalitarian-equivalent calibration and a Leontief agent A, any z with unequal amounts of the goods will do: $m_A(z)$ is the smallest amount, and anyone who values more of every good gives z a higher score. Percentile scales escape this construction, because they move with the law: concentrating the drafts near z changes where z stands in each person's distribution.

This part of the favoritism is the expected price of ignoring the drafts. But the laws μ_k are built for the purpose, and a planner does not choose the drafting process. Nor does independence from the law imply failure under a given law. Score each person by her percentile under the uniform law, computed once and for all: the scale ignores the actual law, yet it equalizes expected favor rates under every rule whenever the actual law is uniform (Theorem 6). The question that matters is what an index does under a law that is given and treats people alike.

We answer it for many goods, $\Omega = \{1, \ldots, K\}$, and any number of agents. A proposal law μ_K is a *product law* if it splits each good independently according to a common law Q on $T = \{s \in [0,1]^n : \sum_i s_i \leq 1\}$, the share s_i going to agent i. With two agents, A and B, T is the triangle of pairs (a, b) with $a + b \leq 1$. We assume that Q has a continuous density bounded away

[7] The cost is computed over all nonnegative bundles, with preferences extended homothetically. Restricting the cheapest bundle to $X = [0,1]^\Omega$ would let the bound $z(\omega) \leq 1$ separate the two calibrations, for reasons unrelated to tastes.

from zero on T; then $\mu_K \otimes \mu_K$ satisfies Assumption 1. The law treats the agents alike when Q is symmetric, that is, unchanged when agents swap their shares. Agent i has CES preferences with equal weights and parameter $\rho_i \in [-\infty, 1]$. The lower ρ_i, the more she needs the goods together: at $\rho_i = -\infty$ she ranks bundles by her smallest share (Leontief), at $\rho_i = 1$ by her average share. Her egalitarian-equivalent score is the power mean of her shares, $m_i(x) = \left(\frac{1}{K}\sum_\omega x(\omega)^{\rho_i}\right)^{1/\rho_i}$, with the minimum at $\rho_i = -\infty$ and the geometric mean at $\rho_i = 0$. As goods multiply, it concentrates around a constant c_i, her *typical score*: if s denotes her share of a good under Q, $c_i = (\mathbb{E}[s^{\rho_i}])^{1/\rho_i}$ for $\rho_i \in (-1, 1]$, $\rho_i \neq 0$, $c_i = \exp \mathbb{E}[\log s]$ for $\rho_i = 0$, and $c_i = 0$ for $\rho_i \leq -1$. When Q is symmetric, all agents' shares have the same distribution and power means increase with their order, so $c_i < c_j$ whenever $\rho_i < \rho_j$ and $\rho_j > -1$. We also use the utilitarian rule, which sums the scores, and the Nash rule, which multiplies them.

Theorem 2. *Let* **m** *be the egalitarian-equivalent calibration and let μ_K be product laws.*

(a) Suppose that at most one agent has $\rho_i \leq -\frac{1}{3}$. If $c_A < c_j$ for every $j \neq A$, then $f_A \to 1$ and $\bar{f}_A \to 1$ under leximin as $K \to \infty$. If $c_Z > c_j$ for every $j \neq Z$, then $f_Z \to 1$ and $\bar{f}_Z \to 1$ under leximax.

(b) Let $n = 2$. If A ranks bundles by her smallest share and B by her average share, there are constants $C, r > 0$ such that, for every $K \geq 2$,

$$1 - f_A^{\text{leximin}} \leq Ce^{-rK}, \qquad f_A^{\text{util}} \leq C\,\frac{\log K}{\sqrt{K}}, \qquad 1 - f_A^{\text{Nash}} \leq C\sqrt{\frac{\log K}{K}},$$

and the same bounds hold for expected favor rates.

Part (a) needs no tailored law and no extreme taste. When Q is symmetric, all agents receive shares with the same distribution, and the only difference between them is how much they need the goods together. That difference decides almost every conflict: the index scores the most complementary agent lowest, and leximin protects her; leximax protects the agent with the highest typical score, the one who needs the goods together least. With two agents, $f_A + f_B = 1$, so A loses almost every conflict under leximax. Agents with $\rho_A = 0$ and $\rho_B = \frac{1}{2}$ suffice; under the uniform law, A wins 63%, 78% and 95% of the conflicts leximin resolves with 4, 16 and 64 goods. A third agent does not dilute the effect. Add one with $\rho = 1$: with 64 goods, A still wins 97% of her conflicts under leximin, and the new agent wins 93% of hers under leximax. With a Leontief agent facing one who cares about her average share, the convergence is fast: 78%, 96% and 99.9% with two, four and eight goods.[8] The law need not treat the agents alike. When it does not, the leximin near-dictator is the agent with the lowest typical score, whether her tastes or the drafts place her there. Resource indices are motivated by the view that differences in tastes are not grounds for priority. Under a symmetric product law, a difference in tastes is all that separates the agents, and it decides who prevails.

Part (b) shows that changing the rule changes the winner, not the outcome. Leximin and leximax compare levels. The utilitarian and Nash rules compare gains: in a conflict, with w_i and ℓ_i agent i's scores in the proposal she prefers and in the other one, they compare $w_A - \ell_A$ with $w_B - \ell_B$, and $\log(w_A/\ell_A)$ with $\log(w_B/\ell_B)$. The Leontief agent's gain is of order $1/K$ in absolute terms but of order one relative to her level, which is itself of order $1/K$; the other agent's gain is of order $1/\sqrt{K}$ in both senses. The utilitarian rule therefore sides with B and the Nash rule with A, both at a speed

[8] Monte Carlo estimates from 10^5 to 4×10^5 pairs of uniform drafts for each number of goods. With a fixed set of goods and a fixed law, the favor rate stays below a bound smaller than one, whatever the preferences. An egalitarian-equivalent index lies between the smallest and the largest amount in the bundle, so proposals close to the proportional divisions $(0.4 \cdot 1_\Omega, 0.1 \cdot 1_\Omega)$ and $(0.3 \cdot 1_\Omega, 0.2 \cdot 1_\Omega)$ receive nearly the same indices under every preference, and leximin resolves their conflicts against A. The bound need not be informative: under the uniform law with four goods, a Leontief agent facing one who cares only about her best good wins 99% of resolved conflicts.

close to $1/\sqrt{K}$, against an exponential speed under leximin.[9] No rule among the four treats the agents alike. Part (b) is stated for two agents because the utilitarian rule changes character with more. The Nash rule still sides with the Leontief agent, whose relative gain outweighs the others' combined relative gains. The utilitarian rule instead ignores her: her absolute gain is too small to matter, and the others' total decides. Facing two agents who care about their average share, she wins 96% of her conflicts under the Nash rule with 64 goods, and 42% under the utilitarian rule.[10]

Both results hold for expected favor rates. The favoritism therefore comes from the scale, not from how the drafts pair the agents' fortunes. Under a common percentile scale, by contrast, no proposal law creates expected favoritism (Section 4.6), and with two agents, under the laws of Theorem 2, percentiles also give each agent half of the actual conflicts as goods multiply (Section 4.8).

2.5. Indices with a free reference

Some welfare measures fix a reference that is not a rescaled set. They fail the test for a further reason: the reference can decide on its own who wins.

Equivalent income. Choose one good as income, here good 1, and fix the other goods at reference levels, here one unit each. A person's equivalent income for x is the income y such that $x \sim_i (y, 1, \dots, 1)$ (Fleurbaey and Blanchet, 2013; Decancq et al., 2015), where preferences are extended to negative incomes when needed. It is not a calibration in our sense, because the agents' ranges differ; leximin, leximax and the utilitarian rule apply to real-valued scores unchanged.

Proposition 2. *Under the product laws of Theorem 2, with two agents, let A rank bundles by her smallest share and B by her average share. The equivalent incomes are* $y_A(x) = \min_\omega x(\omega)$ *and* $y_B(x) = \sum_\omega x(\omega) - (K-1)$. *As* $K \to \infty$, $f_A \to 0$ *under leximin and under the utilitarian rule, and* $f_A \to 1$ *under leximax; the same holds for expected favor rates, and the convergence under leximin and leximax is exponentially fast.*

The reversal of Theorem 2 comes from the ranges, not from the tastes: y_A lies in $[0, 1]$, while y_B ranges over $[-(K-1), 1]$, so the reference places B's ordinary bundles far below A's. Rescaling each range affinely to $[0, 1]$ leaves y_A unchanged and turns y_B into the average share, which gives back the egalitarian-equivalent index. Under the uniform law, B wins 98% of the conflicts leximin resolves with two goods and all of them, to three decimals, from four goods on.

Money metric with non-homothetic preferences. The reference prices then decide. In a two-good economy where the two agents' tastes are mirror images of each other and the uniform law treats them alike, equal reference prices give each agent exactly half of the conflicts leximin resolves; moving the price of the first good from 0.001 to 0.999 moves A's favor rate from 0.97 to 0.03.[11] This makes measurable the objection of Blackorby and Donaldson (1988): the reference prices, not the tastes, decide who wins.

Relative utilitarianism. Dhillon and Mertens (1999) normalize each person's von Neumann–Morgenstern utility between her best and worst alternatives. Every calibration is such a normalized utility for some attitudes toward risk, so Theorem 1, applied to the utilitarian rule, shows that some

[9] A local limit argument gives $\sqrt{K}\, f_A^{\text{util}} \to 1/(\sqrt{\pi}\, f_a(0)\, \sigma_b)$ and $\sqrt{K}\,(1 - f_A^{\text{Nash}}) \to \sigma_b/(\sqrt{\pi}\, \mu_b)$, where $f_a(0)$ is the density of A's share at zero and μ_b, σ_b are the mean and standard deviation of B's share; under the uniform law, 1.197 and 0.399. Monte Carlo estimates of f_A^{util} are 0.41, 0.26, 0.14 and 0.07 with 4, 16, 64 and 256 goods, and of f_A^{Nash} are 0.80, 0.91, 0.95 and 0.98. With two CES agents with $\rho_A, \rho_B > -\frac{1}{2}$, both gains are of order $1/\sqrt{K}$, and the utilitarian and Nash rates converge to intermediate values given by a bivariate central limit argument: 0.52 and 0.61 for $\rho_A = 0$ and $\rho_B = \frac{1}{2}$ under the uniform law, while the leximin rate tends to one. Only leximin and leximax turn a difference in curvature alone into near-dictatorship.

[10] Monte Carlo estimates from 2×10^5 pairs of uniform drafts. Under the utilitarian rule, her rate falls from 0.52 with 4 goods to 0.46 with 16.

[11] The utilities are $u_A = \frac{1}{2}x_1^{0.3} + \frac{1}{2}x_2^{0.9}$ and $u_B = \frac{1}{2}x_1^{0.9} + \frac{1}{2}x_2^{0.3}$. Monte Carlo estimates from 4×10^5 pairs; the Walrasian prices with equal endowments are equal by symmetry and also give one half.

risk attitudes make relative utilitarianism side with a designated person in almost every conflict. With risk neutrality in egalitarian-equivalent units, it is the utilitarian rule of Theorem 2(b).

Variance normalization (MacAskill et al., 2020) rescales each person's utility to mean zero and unit variance under the proposal law. It depends on the law and so escapes Proposition 1. It equalizes two moments of the score distributions, not the distributions, so it does not equalize expected favor rates under every rule (Theorem 6). In practice it comes close: applied to egalitarian equivalents for a Leontief agent and an average-share agent under the uniform law, it gives rates between 0.43 and 0.49 under leximin, leximax and the utilitarian rule, for 2 to 64 goods. Depending on the law is therefore not enough; Section 4.6 identifies the dependence that is.

3. IMPARTIAL SOCIAL ORDERINGS

Section 2 shows that standard indices fail the test, and that any anonymous rule fails it under some calibration. Which rankings of allocations, then, are impartial when preference intensities cannot be compared? This section answers without choosing a welfare scale first. It places requirements on the ranking of allocations directly and derives the rankings that meet them. One impartiality requirement on how the ranking compares people forces percentile comparisons. A requirement on the information the ranking uses then makes it a function of percentiles alone. One further requirement selects a criterion: an equity condition gives the leximin of percentiles, and consistency across drafting processes gives their sum.

3.1. Social ordering functions

A social ranking may depend on preferences and on the drafting process. An *environment* $\xi = (\succsim, \mu)$ consists of a profile of continuous, monotone preferences and a proposal law μ for which $\mu \otimes \mu$ satisfies Assumption 1; Ξ is the set of environments.

Percentiles. Agent i's percentile of a bundle is the share of proposals that give her a bundle she ranks no higher:

$$U_i^\xi(x) = \mu\{\mathbf{x}' : x_i' \precsim_i x\}, \qquad U^\xi(\mathbf{x}) = (U_1^\xi(x_1), \dots, U_n^\xi(x_n)).$$

It depends only on her ordinal ranking and the proposal law: $U_i^\xi(x) = 0.7$ means that agent i likes bundle x at least as much as what she receives in 70% of proposals. Percentiles measure a bundle against what the same person might receive. By Lemma O2, U_i^ξ is a continuous representation of $\succsim_i$ onto $[0,1]$, and $U_i^\xi(X_i)$ is uniform on $[0,1]$ when $\mathbf{X} = (X_1, \dots, X_n)$ is a random proposal with law μ: everyone's percentiles are spread evenly.

Social ordering functions. A *social ordering function* assigns to every environment ξ a weak order $\succeq_\xi$ on X^n, with strict part $\succ_\xi$ and indifference part $\sim_\xi$. Following Fleurbaey and Maniquet (2011), it ranks every profile of bundles, feasible or not; the planner applies it to the proposals that reach the floor. Each choice of a calibration in every environment, followed by a fixed aggregator, defines a social ordering function. Not every social ordering function arises this way, and the requirements below decide which ones qualify.

Single-recipient allocations reveal how a ranking compares people (Section 2.2). A *position* (i, x) is agent i holding bundle x, and the social ordering function compares positions through the allocations $\mathbf{e}_i(x)$:

$$(i,x) \sqsupseteq_\xi (j,y) \quad \text{if and only if} \quad \mathbf{e}_i(x) \succeq_\xi \mathbf{e}_j(y),$$

with strict part $\sqsupset_\xi$. In words, i with x is at least as well placed as j with y when society would rather give x to i alone than y to j alone. All allocations $\mathbf{e}_i(0_\Omega)$ coincide, so the empty bundle places everyone alike.

Definition 1. *A social ordering function satisfies* strong Pareto (P) *if, in every environment, $x_i \succsim_i y_i$ for all i implies $\mathbf{x} \succeq_\xi \mathbf{y}$, strictly if some of these preferences is strict;* anonymity (A) *if renaming the*

agents, together with their preferences and their shares in every proposal, renames the ranking and changes nothing else;[12] *and* continuity (C) *if, in every environment, for every position* (j, y) *and agent* i*, the set of bundles* x *with* $(i, x) \sqsupset_\xi (j, y)$ *is open. It satisfies* full continuity (C*) *if every* $\succeq_\xi$ *is continuous on* X^n.

These conditions are standard. Continuity (C) concerns single-recipient allocations only: it excludes breaking ties between people of equal standing by name.

Equal standing (ES). *In every environment* $\xi = (\gtrsim, \mu)$*, for every position* (j, y) *and all agents* i, i'*,*

$$\mu\{\mathbf{x} : (i, x_i) \sqsupset_\xi (j, y)\} = \mu\{\mathbf{x} : (i', x_{i'}) \sqsupset_\xi (j, y)\}.$$

Fix a benchmark in which one person alone holds a bundle. Draw a proposal, and give person i, alone, the bundle it assigns her. Equal standing requires that the chance that society prefers this to the benchmark not depend on who i is. The condition uses no welfare scale: it involves the ranking's own comparisons and frequencies under the proposal law.

Proposition 3. *For every environment* ξ *let* $\mathbf{m}^\xi$ *be a calibration, and let* R *be an aggregator. The social ordering function that ranks* $\mathbf{x}$ *weakly above* $\mathbf{y}$ *when* $\mathbf{m}^\xi(\mathbf{x})\,\mathrm{R}\,\mathbf{m}^\xi(\mathbf{y})$ *satisfies* (ES) *if and only if, in every environment, all agents have the same score distribution under* $\mathbf{m}^\xi$.

Proof. By anonymity and strong Pareto of R, $(i, x) \sqsupset_\xi (j, y)$ if and only if $m_i^\xi(x) > m_j^\xi(y)$ (Section 2.2). So (ES) requires $\mu\{\mathbf{x} : m_i^\xi(x_i) > t\}$ to be independent of i at every value $t = m_j^\xi(y)$, that is, at every $t \in [0, 1]$, since score functions are onto. □

Equal standing is thus a requirement of equal prospects, stated without a scale. The egalitarian-equivalent leximin ranking violates it in the economies of Theorem 2, where the agents' index distributions differ. Section 4.6 shows that equal score distributions are exactly what impartiality in expectation under every anonymous rule demands of a calibration (Theorem 6). With no calibration given, equal standing already forces percentile comparisons.

Proposition 4. *Let a social ordering function satisfy* (P)*,* (C) *and* (ES)*. In every environment* ξ*,* $(i, x) \sqsupseteq_\xi (j, y)$ *if and only if* $U_i^\xi(x) \geq U_j^\xi(y)$.

By strong Pareto, the bundles that place i above a benchmark form an upper contour set of her preference. By continuity it is open, so it consists of the bundles above some percentile. Equal standing fixes its probability at the value it has for the benchmark's owner, and percentiles are uniformly distributed. The threshold must therefore be the benchmark's percentile.

What may the ranking use beyond these comparisons? For a position (i, x) and an agent k, let

$$s_\xi(i, x \mid k) = \mu\{\mathbf{x}' : (k, x'_k) \sqsubseteq_\xi (i, x)\}$$

be the share of proposals that place k, alone, no higher than i with x.

Frequency independence (FI). *If* $s_\xi(i, x_i \mid k) = s_{\xi'}(i, x'_i \mid k)$ *and* $s_\xi(i, y_i \mid k) = s_{\xi'}(i, y'_i \mid k)$ *for all agents* i *and* k*, then* $\mathbf{x} \succeq_\xi \mathbf{y}$ *if and only if* $\mathbf{x}' \succeq_{\xi'} \mathbf{y}'$.

In words, what society may use about a position is how often the drafting process would place each person below it, by society's own comparison. The shape of indifference curves and the resource content of bundles may not enter otherwise. The condition fixes what the ranking may use, not

[12] Formally, for a permutation σ of N, let $(\sigma\mathbf{x})_{\sigma(i)} = x_i$, and let $\sigma\xi$ give agent $\sigma(i)$ the preference of agent i and replace μ by the law of $\sigma\mathbf{X}$. Anonymity requires $\mathbf{x} \succeq_\xi \mathbf{y}$ if and only if $\sigma\mathbf{x} \succeq_{\sigma\xi} \sigma\mathbf{y}$. Then $U^{\sigma\xi}(\sigma\mathbf{x}) = \sigma U^\xi(\mathbf{x})$.

how. It does not presuppose percentiles. The egalitarian-equivalent leximin ranking satisfies it: its frequencies are the distribution functions of the indices, evaluated at indices, so two situations with the same frequencies differ by one common increasing transformation of the indices, which leximin ignores. Equal standing, not frequency independence, excludes this ranking.

Theorem 3. *A social ordering function satisfies* (P)*,* (A)*,* (C)*,* (ES) *and* (FI) *if and only if there is an aggregator* R *such that, in every environment* ξ*,*

$$\mathbf{x} \succeq_\xi \mathbf{y} \quad \text{if and only if} \quad U^\xi(\mathbf{x}) \ \mathrm{R} \ U^\xi(\mathbf{y}).$$

The aggregator is unique. It is continuous if and only if the social ordering function satisfies (C*).

Such a ranking scores people by their percentiles and then applies an anonymous, Paretian rule, the same in every environment. The requirements mention neither scales nor percentiles: both are derived. The rule itself remains open. A common increasing transformation of percentiles can be moved into the rule, so no requirement on how people are compared can fix it (Section 4.4). Section 4 shows that every such ranking treats people alike in expectation, and identifies when it also does so in its actual decisions.

3.2. Two criteria

Two familiar requirements close the rule.

Hammond equity (HE). *Let* $x_k \sim_k y_k$ *for all* $k \neq i, j$*. If* $(j, y_j) \sqsubset_\xi (j, x_j) \sqsubset_\xi (i, x_i) \sqsubset_\xi (i, y_i)$*, then* $\mathbf{x} \succeq_\xi \mathbf{y}$.

A change that narrows the distance between two people, as society itself compares their positions, and leaves everyone else indifferent does not make things socially worse (Hammond, 1976).

Corollary 1. *A social ordering function satisfies* (P)*,* (A)*,* (C)*,* (ES)*,* (FI) *and* (HE) *if and only if it is* percentile leximin*:* $\mathbf{x} \succeq_\xi \mathbf{y}$ *if and only if the increasing rearrangement of* $U^\xi(\mathbf{x})$ *is lexicographically at least that of* $U^\xi(\mathbf{y})$.

Percentile leximin gives priority to the person whose bundle is worst relative to what she might have received. It is the counterpart of relative egalitarianism (Sprumont, 2013), which applies leximin to utilities normalized by each person's best and worst alternatives; here the proposal law takes the place of the two extremes.

Mixture consistency (MC). *Let* $\xi = (\succsim, \mu)$ *and* $\xi' = (\succsim, \mu')$ *be environments with the same preferences, and let* $\xi^\lambda = (\succsim, \lambda\mu + (1-\lambda)\mu')$ *with* $\lambda \in (0, 1)$*. If* $\mathbf{x} \succeq_\xi \mathbf{y}$ *and* $\mathbf{x} \succeq_{\xi'} \mathbf{y}$*, then* $\mathbf{x} \succeq_{\xi^\lambda} \mathbf{y}$.

Suppose two offices may draft the proposals, and the planner prefers $\mathbf{x}$ to $\mathbf{y}$ whichever office drafts. Mixture consistency requires the same preference when both offices draft in given proportions, or when she does not know which one will. The mixed law is again admissible. Percentiles are linear in the proposal law, $U^{\xi^\lambda} = \lambda U^\xi + (1-\lambda)U^{\xi'}$, so the requirement gives frequencies a cardinal meaning, as lotteries do for Harsanyi (1955) and Dhillon and Mertens (1999).

Theorem 4. *Let* $n \geq 2$*. A social ordering function satisfies* (P)*,* (A)*,* (C*)*,* (ES)*,* (FI) *and* (MC) *if and only if it is* percentile utilitarianism*: in every environment* ξ*,*

$$\mathbf{x} \succeq_\xi \mathbf{y} \quad \text{if and only if} \quad \sum_i U_i^\xi(x_i) \ \geq \ \sum_i U_i^\xi(y_i).$$

The sum of percentiles has a direct reading: it is the expected number of people for whom the allocation is at least as good as a randomly drafted proposal. The proof follows Harsanyi. Theorem 3 gives one aggregator for all environments, and linearity of percentiles turns mixture consistency into

a property of that aggregator: two comparisons of score vectors with the same pattern of individual preferences, each ranked one way, are ranked the same way after mixing. Mixing indifferent pairs shows that the rate at which a loss for one person is compensated by a gain for another is the same everywhere. Anonymity makes it one, and the aggregator is the sum. The argument uses environments in which every agent values the quantity of a single good; there, any interior percentile values with the right signs can be produced by choosing the proposal law.

No other rule survives. For three or more agents, mixture consistency excludes leximin, and sums of a nonlinear transformation of percentiles violate it (Online Appendix O5). With two agents, leximin satisfies mixture consistency, and full continuity excludes it.

What about frequency independence? It cannot be dropped either, and the ranking that shows it uses resources.

Remark 1. *Let e_i be agent i's egalitarian-equivalent index (Section 2.4), which depends on her preference only. Fix $\varepsilon > 0$ and rank allocations in each environment by*

$$W^{\xi}(\mathbf{x}) = \sum_i U_i^{\xi}(x_i) + \varepsilon \sum_{i<j} e_i(x_i)\, e_j(x_j).$$

This ranking satisfies (P)*,* (A)*,* (C*)*,* (ES) *and* (MC)*, but not* (FI)*.*

The second term never affects how positions compare, so equal standing cannot see it, and it does not vary with the drafts, so mixture consistency cannot either. Only frequency independence excludes it. Together with the sum of squared percentiles, which violates only (MC), and two-agent percentile leximin, which violates only (C*), the remark shows that each of (FI), (MC) and (C*) does work in Theorem 4.

Could ordinal information alone suffice? A ranking that used only how positions compare with one another would be unchanged by any common increasing transformation of percentiles (Proposition 4), and no continuous aggregator has this invariance (Proposition O1 in the Online Appendix). Ordinal information therefore leads to leximin-type rules, which are discontinuous, and a continuous ranking must use frequencies. The two criteria divide as leximin and utilitarianism do in Deschamps and Gevers (1978): equity selects one, continuity and consistency the other.

Frequency independence is the strongest requirement in Theorem 3, and it needs a defense. Within one environment it has no bite. Each percentile represents its owner's preference, so two allocations with the same percentile vector leave everyone indifferent, and strong Pareto makes them socially indifferent. The ranking in ξ is therefore already some Paretian rule R_ξ applied to percentiles. What frequency independence adds is that R_ξ is the same rule in every environment. In the welfarism theorem of d'Aspremont and Gevers (1977), this invariance across situations is derived from independence of irrelevant alternatives over an unrestricted domain. Here it is assumed. Our setting has no counterpart of that independence: a percentile depends on the whole proposal law, so a change of drafts changes how every bundle is measured. Remark 1 shows what the assumption excludes, namely rules that shift with the environment.

Three observations limit its force. First, equal standing alone already fixes how the ranking compares positions (Proposition 4); frequency independence only extends that comparison from positions to allocations. Second, it does not presuppose percentiles, since the egalitarian-equivalent leximin ranking satisfies it. Third, the alternatives have costs. A ranking that compares positions by the resource content of bundles, as reference-bundle indices do, violates equal standing in the economies of Theorem 2. A ranking that uses resources only beyond positions, as in Remark 1, passes equal standing; but its rate of exchange between two people's percentiles then depends on the quantities behind them, so two societies in which everyone has the same prospects can trade the same percentiles differently. Or a ranking may use only how positions compare with one another; by Proposition O1, it then cannot be continuous.

3.3. Percentile utilitarianism and the Borda rule

Suppose the proposal law were uniform over a finite menu of allocations. A person's percentile of her bundle in an allocation would then be the share of menu allocations that she ranks no higher, and the sum of percentiles would be the allocation's Borda score divided by the size of the menu, up to the treatment of ties. Assumption 1 excludes finite menus, because it requires atomless welfare levels. Percentile utilitarianism is thus the continuous analogue of Borda's rule, with the proposal law weighting the alternatives. The two characterizations rest on different consistency requirements. Young (1974) characterizes Borda's rule by consistency across electorates: if two groups of voters each rank an alternative first, so does their union. Mixture consistency is consistency across drafting processes, holding the people fixed, and it is what gives frequencies their cardinal meaning here.

The familiar objection to Borda's rule applies. A percentile depends on which proposals are drafted: adding drafts that are bad for i raises her percentile, and duplicating a proposal changes the weights. This dependence is the price of interpersonal comparisons built from ordinal information, not a peculiarity of percentiles. Relative utilitarianism depends on each person's best and worst alternatives in the same way (Dhillon and Mertens, 1999), and Arrow's theorem rules out nondictatorial rankings that both ignore irrelevant alternatives and use ordinal, noncomparable information only (Sen, 1970). Our criteria place the dependence in a single primitive, the drafting process, which an institution can document and audit. The dependence also creates an incentive: an office that knows the criterion could shape its drafts to move percentiles. Assumption 1 holds the law fixed, and Section 5 returns to the stability this requires.

Two further limits deserve mention. The necessity part of Theorem 4 uses a rich set of environments; on a restricted domain, for instance with preferences held fixed, it remains open. And neither criterion makes favor rates equal in every environment, as the next section shows.

4. WHAT IMPARTIAL RANKINGS GUARANTEE

Section 3 derives rankings that compare people by percentiles and leaves the rule open. Do they pass the test of Section 2, and do they treat people alike in the decisions they take? We measure this by favor rates and characterize the scales that equalize them under every anonymous rule. The answer separates what any scale can achieve from what depends on preferences and proposals.

4.1. Balance and expected balance

Theorem 1 shows that anonymity alone leaves the calibration free to favor anyone, and Section 2.4 shows that standard calibrations use this freedom. The test of the introduction excludes near-dictators; its natural strengthening requires favor rates to be equal. There are two rates, hence two requirements.

Definition 2. *A calibration* $\mathbf{m}$ *is* balanced *for the aggregator* R *if all favor rates are defined and* $f_1(\mathbf{m}, \mathrm{R}) = f_2(\mathbf{m}, \mathrm{R}) = \cdots = f_n(\mathbf{m}, \mathrm{R})$. *It is* expected-balanced *for* R *if all expected favor rates are defined and* $\bar{f}_1(\mathbf{m}, \mathrm{R}) = \bar{f}_2(\mathbf{m}, \mathrm{R}) = \cdots = \bar{f}_n(\mathbf{m}, \mathrm{R})$.

Balance requires that, when people disagree and the rule decides, the scale not let one person's preferred proposal prevail more often than another's. Expected balance requires that the scale not do so through the prospects it gives each person, whatever the proposals pair together. Both constrain the calibration, given the rule and the pair law; anonymity of the rule implies neither. With two agents, balance means that each wins half the resolved conflicts, and expected balance half the expected ones; with more, several can prefer the winner, so the common rate need not be one half.

Both criteria use only information the model contains: who prefers which proposal, and how often each conflict occurs. Weighting a conflict by interpersonal gains would require further welfare comparisons; it would not solve the measurement problem from ordinal information. Moreover, they test whether a person's interests prevail systematically under the same procedure, without judging each allocation against a separately chosen standard of material equality.

Because the proposals are random, a favor rate is also a chance: before the pair is drawn, it is the probability that a resolved conflict goes the person's favor. Rae (1969) evaluates decision rules by the same kind of probability, each voter's chance of having her way. Balance requires these chances to be equal. The concern is concrete. A person who loses nearly every resolved conflict does not stand as an equal in collective decisions, although no rule names her; her complaint is that of a permanent minority under majority rule (Guinier, 1994).

Balance also differs from voting power. We do not change one person's report while holding the others fixed, as pivotality calculations do (Dubey and Shapley, 1979), and we do not draw preference profiles from a voting culture (Gehrlein and Lepelley, 2011). We fix preferences and count how often the ranking agrees with them across the allocation comparisons the institution faces. Balance does not claim that each agreement carries the same welfare weight.

Which of the two is the right test of a welfare scale? They answer different questions. Balance judges the decisions the planner actually takes. Expected balance judges the scale, by the decisions it would produce if nothing linked one person's fortune to another's. The difference is exactly the link. For every calibration and rule,

$$f_i - \frac{1}{n}\sum_j f_j = \left[\bar{f}_i - \frac{1}{n}\sum_j \bar{f}_j\right] + \left[(f_i - \bar{f}_i) - \frac{1}{n}\sum_j (f_j - \bar{f}_j)\right]. \tag{1}$$

The first bracket is what the scale does to each person's prospects. The second is what the proposals add by pairing fortunes: a draft generous to one person is often stingy with another. Expected balance sets the first bracket to zero. Balance sets the sum to zero, and may do so by letting the scale offset a link it did not create. Is that offset impartial measurement, or a distributive correction? The two criteria divide on this question, and rule by rule neither implies the other. A scale balanced for leximin can give one person systematically lower prospects, and a scale with equal prospects can let her lose under leximin because the drafts pair her good fortune with another's bad one. Section 4.2 shows both.

Both criteria concern the impartial treatment of ordinal interests when luck is involved. A person's need for complementary goods might justify compensation, and a resource- or needs-based index could then rightly place her below others. The classic debate about who counts as worse off already makes the choice of interpersonal index substantive (Rawls, 1971; Arrow, 1973; Gibbard, 1979). Balance and expected balance are neutrality benchmarks for the case in which differences in tastes, as opposed to needs, are not grounds for priority: the numerical representation of those tastes should not make one person's preferred proposal prevail more often. The egalitarian-equivalent and money-metric calibrations are neutral toward tastes in another sense, and Theorem 2 shows that the two senses can conflict. The claim is deliberately conditional: equal favor rates do not by themselves establish equal welfare, equal influence or the right treatment of needs.

4.2. The two criteria in an example

With two agents, leximin settles a conflict by one comparison: it sides with the agent whose score in the proposal she likes less is the lower one (Proposition 6). Consider two agents who both value the quantity of one good, scored by that quantity: $m_1(x) = m_2(x) = x$. The proposal law below gives both agents the same distribution of shares, so the two score distributions coincide and this calibration is expected-balanced for every aggregator. Let the proposal law mix a uniform component on the feasible triangle with equal masses on three small squares (Figure 1). Each agent's share lies near 0.1, 0.3 or 0.6 with the same frequency. Yet when the two shares are near 0.1 and 0.3, agent 1 holds the smaller one 99.9% of the time, and likewise near 0.3 and 0.6; agent 2 holds the smaller share only near 0.1 and 0.6. With identical score distributions, agent 1 is the lower agent in 65% of proposals. Leximin protects the lower agent, and the conflicts it resolves go against her: $f_1 \approx 0.38$

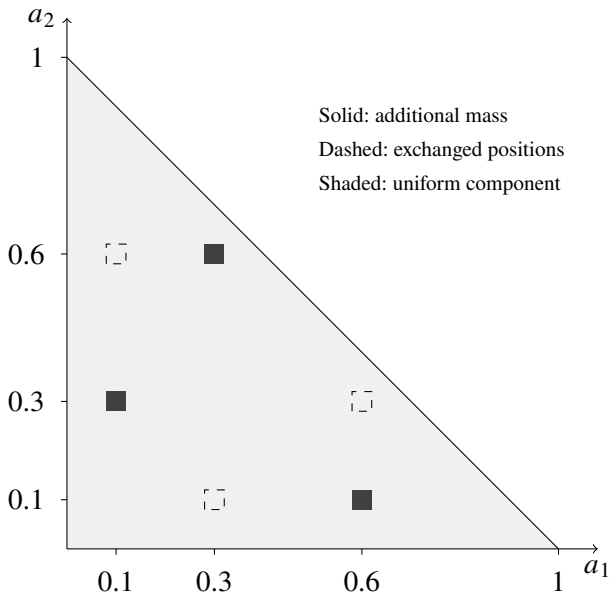


Figure 1. Identical score distributions, unequal positions. The proposal law mixes mass 0.1 uniformly on the feasible triangle with mass 0.3 on each solid square $I_1 \times I_2$, $I_2 \times I_3$ and $I_3 \times I_1$, where $I_1 = [0.08, 0.12]$, $I_2 = [0.28, 0.32]$ and $I_3 = [0.58, 0.62]$. The dashed squares receive only uniform mass. This law has a full-support density, so $\mu \otimes \mu$ satisfies Assumption 1, and the two agents' score distributions coincide; yet given sorted scores in $I_1 \times I_2$, agent 1 holds the lower one with probability 0.30032/0.30064.

under leximin, and 0.62 under leximax.[13] No common rescaling changes who holds the lower share. Balance can nonetheless be restored: raising agent 1's score to the power 1.1 gives $f_1 = \frac{1}{2}$ under leximin, at the price of unequal score distributions. This balanced calibration has an expected favor rate of 0.65.

Which calibrations satisfy each criterion once the rule is left open?

4.3. Impartiality across rules

Both criteria of Section 4.1 take the rule as given, and a planner who has settled on leximin can tune the scale to leximin, as the power transformation of Section 4.2 did. Why demand more? Because the scale is meant as a measurement and the rule as a distributive choice. Measurement puts people's welfare on a common footing; the rule decides how society trades one person's welfare against another's. A scale balanced for leximin alone has borrowed its impartiality from leximin. It loses it if the planner adopts leximax, and it does not show that the guarantee survives a change of distributive objective. Public appraisal draws the same line between appraisal and objectives (HM Treasury, 2026). A second reason is bite. For one rule and two agents, each criterion is one equation, and a scale is a function: many scales are balanced for leximin, and many are expected-balanced for it. Only the accumulation of rules can determine a scale. A third reason is the disagreement of Section 4.2. Rule by rule, balance and expected balance select different scales, and one can judge between them only by seeing what each demands when the rule is left open.

A calibration has *rule-independent balance* if it is balanced for every aggregator for which the favor rates are defined, and *rule-independent expected balance* if it is expected-balanced for every aggregator for which the expected favor rates are defined. The requirement neither averages over objectives nor asks them to choose the same proposal. It identifies what measurement can guarantee before any objective is selected. When the guarantee fails, the failure limits what scales alone can achieve; it does not show that every objective must produce an unbalanced ranking.

[13] Monte Carlo estimates from 4×10^6 independent pairs; standard errors are below 0.0005. The utilitarian rule and the Nash rule, which multiplies the scores, give rates within 0.002 of one half (Online Appendix O3).

Fix an environment and write U_i for U_i^{ξ} (Section 3.1). Every calibration is a rescaling of percentiles, agent by agent. For a calibration $\mathbf{m}$, let $\Phi_i(t) = \mu\{\mathbf{x} : m_i(x_i) \leq t\}$ be the distribution function of ν_i, the share of proposals that give i a score of at most t. Assumption 1 makes Φ_i an increasing homeomorphism of $[0, 1]$, that is, a continuous, strictly increasing bijection with a continuous inverse (Lemma O2). Thus $U_i = \Phi_i \circ m_i$ is a continuous, onto preference representation: percentiles form a calibration. Write $\mathbf{X} = (X_1, \ldots, X_n)$ for a random proposal with law μ, so that X_i is the bundle it gives agent i; then $U_i(X_i)$ is uniform on $[0, 1]$. Conversely, $m_i = \psi_i \circ U_i$ with $\psi_i = \Phi_i^{-1}$, and the distribution function of $m_i(X_i)$ is ψ_i^{-1}. Everyone has the same score distribution,

$$\mu\{\mathbf{x} : m_i(x_i) \leq t\} = \mu\{\mathbf{x} : m_j(x_j) \leq t\} \qquad (i, j \in N,\ t \in [0, 1]), \tag{2}$$

exactly when all ψ_i coincide: the calibrations with equal score distributions are the common increasing transformations of percentiles. Common ranges would not suffice. Requiring everyone's scores to run from zero to one still allows squaring only one person's scores, which keeps both endpoints and every ordinal comparison but puts her near the bottom more often. Equal distributions remove that freedom at every threshold.

4.4. When one calibration balances every rule

An anonymous rule sees the scores in a proposal, not their owners. For a score vector $\mathbf{a}$, write $\mathbf{a}_{()} = (a_{(1)}, \ldots, a_{(n)})$ for its values in increasing order. The rule compares $\mathbf{a}_{()}$ with $\mathbf{b}_{()}$; whose preference it thereby follows depends on who holds which value. The condition below requires that the sorted vector reveal nothing about this.

Definition 3. *A probability law ν on $[0, 1]^n$ is* positionally symmetric *if $(a_i, \mathbf{a}_{()})$ has the same law under ν for every agent i.*

In words, given the scores that occur together, each agent is equally likely to hold each of them: with distinct scores, every agent's rank is uniform on $\{1, \ldots, n\}$ and independent of the sorted vector. With two agents, whenever the scores are 0.2 and 0.8, each is equally likely to hold the 0.2. The one-good economy of Section 4.2 violates it: given scores near 0.1 and 0.3, agent 1 holds the 0.1. Exchangeability, which leaves the whole law unchanged under every permutation of names, implies positional symmetry (Aldous, 1985). The two coincide for two agents and differ for more.[14]

Deducing positional symmetry from balance requires enough feasible comparison profiles. For three or more agents we use the following condition. **Slack (T).** *For ν-almost every score vector $\mathbf{b}$, some feasible score vector $\mathbf{c} \in \mathbf{m}(\mathcal{X})$ has $c_i > b_i$ for every i.*

In words, almost every score vector can be improved for everyone at once. This is a feasibility restriction, not a choice of numerical units. With finitely many goods, a proposal law with a density almost surely leaves some of every good unused. Sharing out part of each remainder strictly improves every bundle, so (T) holds. The condition serves only the necessity part for $n \geq 3$; constructing percentiles and proving sufficiency do not use it.

The necessity part needs only a restricted class $\mathcal{L}$ of anonymous rules. An *upper set* of score profiles contains every profile that is coordinatewise at least as high as one of its members. For any two closed, permutation-invariant upper sets $V, W \subseteq [0, 1]^n$, compare profiles first by membership in V, then, if this test ties, by membership in W, and finally by leximin. Let $\mathcal{L}$ consist of all these rules, including leximin when $V = W = \varnothing$. A planner could, for instance, first check whether every agent reaches her 60th percentile. Such priorities preserve anonymity and strong Pareto. They serve as probes that reveal which agent holds which score, not as recommended distributive objectives.

[14] Conditional on a sorted triple of distinct scores, let only the three cyclic assignments occur, each with probability one third. Every agent then holds each position equally often, although three of the six assignments never occur.

Theorem 5. *Let $n \geq 2$, and if $n \geq 3$ assume* (T)*. For a calibration* $\mathbf{m}$ *the following are equivalent:*

(a) $\mathbf{m}$ *has rule-independent balance;*
(b) $\mathbf{m}$ *is balanced for every rule in* $\mathcal{L}$*;*
(c) the law of the score vector $\mathbf{m}(\mathbf{X})$ *is positionally symmetric;*
(d) $\mathbf{m}$ *scores every agent by her percentile on one common scale,* $m_i = \psi \circ U_i$ *for an increasing homeomorphism* ψ *of* $[0,1]$*, and the law of the percentile vector* $(U_1(X_1), \ldots, U_n(X_n))$ *is positionally symmetric.*

In particular, all calibrations with rule-independent balance make the same interpersonal comparisons: $m_i(x) \geq m_j(y)$ *if and only if* $U_i(x) \geq U_j(y)$*, for all agents* i, j *and bundles* x, y*.*

Condition (d) splits the requirement into a part the planner controls and a part it does not. Common percentile measurement is not imposed in advance: balance across rules forces it, and leaves open only the common graduation ψ. The percentile vector depends on preferences and μ alone, so the second part of (d) is a condition on the primitives. If it fails, no rescaling restores the guarantee. When both parts hold, anonymity cannot turn score configurations into different favor rates for named individuals.

The last statement is the result in its most explicit form. Agent i with bundle x is at least as well off as agent j with bundle y when x beats, for i, at least as large a share of the proposals as y beats for j. The graduation ψ cannot be determined by balance, as a matter of principle. If R is an aggregator, so is the rule R^{ψ} that ranks **a** above **b** when R ranks $(\psi(a_1), \ldots, \psi(a_n))$ above $(\psi(b_1), \ldots, \psi(b_n))$, and $(\psi \circ \mathbf{m}, \mathrm{R})$ ranks proposals as $(\mathbf{m}, \mathrm{R}^{\psi})$ does. A requirement imposed under every rule therefore leaves ψ free: it can be moved from the scale into the rule. What ψ carries is cardinal content. It changes nothing for leximin or leximax, which compare positions only, and it changes the utilitarian ranking: the sum of percentiles and the sum of their squares define two societies, equally impartial in the sense of balance. Choosing ψ is part of choosing the rule.

The proof shows why the criterion leads to a condition on the joint distribution. Under positional symmetry, each person has the same distribution of her own score together with the sorted vector. Two independent proposals then give everyone the same probability of agreeing with any anonymous comparison; unanimous improvements contribute identically for all. For the converse, pairs of threshold rules agree everywhere except on selected regions of score profiles. Balance under both rules reveals who holds the higher or lower values in those regions. The argument recovers the whole conditional score distribution, ties included, not only a few moments (Appendix D).

4.5. An economy with no balanced calibration

The second part of (d) can fail even when the resource proposals treat the agents identically. Consider two agents and two goods: for each good, draw the two agents' shares uniformly from the triangle where their sum is at most one, independently across goods; this proposal law μ is uniform on $\mathcal{X}$. Agent A values her smaller share and B her average. The pair law $\mu \otimes \mu$ satisfies Assumption 1, and μ is invariant to interchanging the agents.

Where does the asymmetry come from? From the interaction between preferences and the shared resource constraint. For A to rank her bundle very highly, both goods must be generous, because her scarcer good determines her welfare. Agent B can rank a bundle highly through different combinations of the same goods. Percentiles make high scores equally frequent for each agent separately. Feasibility can still allow one joint combination of high scores and exclude the combination with the thresholds exchanged.

Formally, the obstruction is

$$\mathbb{P}(U_A > 15/16,\ U_B > 5/6) = 0, \qquad \mathbb{P}(U_A > 5/6,\ U_B > 15/16) > 0.$$

The first event would require A to receive more than half of both goods and B more than half on average, which violates feasibility. The reversed cutoffs are attainable: giving A $3/8$ and B $5/8$ of

each good yields percentiles $3471/4096 > 5/6$ and $485/512 > 15/16$. Nearby interior allocations keep the inequalities and occur with positive probability. The percentile distribution functions behind these numbers have closed forms (Online Appendix O2).

Symmetric resource opportunities and uniform percentile marginals therefore need not give an exchangeable percentile vector, and Theorem 5 rules out a universal balanced calibration in this economy. It does not assert a large residual gap for a particular rule. Here heterogeneous preferences create the obstruction; in the one-good economy of Section 4.2, the proposal law does.

Positive cases are easy to describe. Identical preferences and an agent-symmetric proposal law give an exchangeable percentile vector, so a common percentile scale balances every rule. Identical tastes are not necessary. With two goods and two agents, mirror-image preferences also give exchangeability, provided the proposal law is invariant to exchanging agents and to exchanging goods: one agent cares about the first good exactly as the other cares about the second. Relabeling agents and goods together then interchanges the percentiles without changing their law. The restriction concerns joint welfare opportunities, not heterogeneity as such.

The threshold rules in $\mathcal{L}$ are probes. A familiar family of objectives reveals the same obstruction. For two agents, consider the rankings represented by $\varphi(a_1)+\varphi(a_2)$, where the common transformation φ is continuous and strictly increasing, from the ordinary sum to concave transformations that give priority to low scores. A calibration is balanced for every such ranking if and only if condition (d) holds (Theorem O1 in the Online Appendix).

4.6. When one calibration expected-balances every rule

Expected favor rates are computed in a society where each person's score is an independent draw from her own distribution. No feasibility constraint links the scores there, and a rule sees nothing but the n score distributions. Imposing expected balance under every rule then pins these distributions down.

Theorem 6. *Let $n \geq 2$. For a calibration* $\mathbf{m}$ *the following are equivalent:*

(a) $\mathbf{m}$ *has rule-independent expected balance;*

(b) $\mathbf{m}$ *is expected-balanced for every rule in $\mathcal{L}$;*

(c) all agents have the same score distribution, (2); equivalently, $\mathbb{E}[\varphi(m_i(X_i))]$ does not depend on i for every nondecreasing φ on $[0,1]$;

(d) $\mathbf{m}$ *scores every agent by her percentile on one common scale, $m_i = \psi \circ U_i$ for an increasing homeomorphism ψ of $[0,1]$; equivalently, $m_i(x) \geq m_j(y)$ if and only if $U_i(x) \geq U_j(y)$ for all agents i, j and bundles x, y.*

No slack condition is needed, for any number of agents. The proof reuses that of Theorem 5. Under $\bar{\nu}$ the score vector has independent coordinates with atomless full-support distributions, so every box of score vectors has positive probability and the threshold rules identify the whole law; and a product law is positionally symmetric exactly when its marginals coincide (Appendix D).

Condition (c) requires that each agent be below any given level equally often, or that her expected standing on the scale be everyone's, whatever the graduation. This is the statistical version of "same bundle, same standing", which no calibration can satisfy literally when preferences differ. Equal means alone would not do: scoring agent 1 by her percentile and agent 2 by a scale that concentrates his scores near one half gives both an expected score of one half, yet under leximin agent 1 wins every conflict in which her losing score is below one half. Requiring equal means for every graduation closes this gap and characterizes the percentile comparison exactly. The requirement can always be met, since the percentile calibration meets it.

4.7. The two requirements compared

Theorems 5 and 6 answer the question left open in Section 4.2. Rule by rule, balance and expected balance are incomparable. Imposed under every rule, they nest.

Corollary 2. *Let $n \geq 2$; in parts (b) and (c), if $n \geq 3$, assume* (T).

(a) A calibration with rule-independent expected balance always exists: the common increasing transformations of the percentile calibration have it, and no other calibration does.

(b) Rule-independent balance implies rule-independent expected balance.

(c) If some calibration has rule-independent balance, then a calibration has rule-independent balance if and only if it has rule-independent expected balance.

Part (a) restates Theorem 6. Parts (b) and (c) follow from condition (d) of the two theorems, since the second part of (d) in Theorem 5 concerns the primitives only. Expected balance thus extends balance: it gives the same answer wherever balance gives one, and an answer everywhere else. Whether the two coincide depends on preferences and proposals, not on the scale. In the economy of Section 4.5 they do not, and the percentile calibration is the answer to one requirement and the only candidate for the other.

What remains under expected balance is the second bracket of (1). Leximin and leximax compare positions only, so every calibration with rule-independent expected balance gives them the favor rates of percentiles, which depend on the primitives alone. A planner who measures impartially in the expected sense and uses leximin therefore accepts the residual favoritism that the drafts create, and one who insists on balance for leximin gives up equal prospects. The size of the residual depends on how the proposals link people's fortunes; the economy of Section 4.5 shows that it need not vanish even when the proposals treat everyone identically. An institution that has selected one rule faces an easier design problem.

That problem has a simple solution, which the Online Appendix develops (Section O8). With two agents and a rule whose strict comparisons survive small changes in scores, such as leximin, some calibration gives the first agent any prescribed favor rate, one half in particular (Theorem O2). Under leximin, raising one agent's scores to a power suffices. In a conflict, let ℓ_1 and ℓ_2 be the two agents' base scores in the proposals they like less; the balancing exponent is a median of $\log \ell_2 / \log \ell_1$ across conflicts. A planner can estimate it from ordinal preferences, base scores and proposal pairs, without any interpersonal measure of gains and losses. With three agents, a calibration balanced for leximin exists whenever each good is split among the agents by a symmetric Dirichlet law (Corollary O2).

4.8. Percentiles in the economies of Theorem 2

The test is comparative. Under a given law that treats two people identically, the egalitarian-equivalent index lets one of them win almost every conflict (Theorem 2). Do percentiles do better under the same laws? In expectation they cannot fail (Theorem 6). In actual decisions they can, because positional symmetry depends on the primitives: the two-good economy of Section 4.5, which is the case $K = 2$ of Theorem 2(b), has no calibration balanced for every rule. As goods multiply, this failure fades.

An aggregator has a *null boundary* if the set of pairs of score vectors $(\mathbf{a}, \mathbf{b})$ with $\mathbf{a}\,\mathrm{P}\,\mathbf{b}$ has a boundary of Lebesgue measure zero. A small change of scores then rarely reverses a strict comparison. Leximin, leximax, the utilitarian rule and the Nash rule have a null boundary, and so does every rule represented by $\varphi(a_1) + \varphi(a_2)$ with φ continuous and strictly increasing: on the boundary, two continuous functions of the scores are equal, as in $\min \mathbf{a} = \min \mathbf{b}$, and each such equality holds on a null set.

Theorem 7. *Let $n = 2$ and let μ_K be product laws. Let each agent either rank bundles by her smallest share or have CES preferences with equal weights and $\rho_i \in (-\frac{1}{2}, 1]$. Under the percentile calibration, $f_A \to \frac{1}{2}$ as $K \to \infty$, under every aggregator with a null boundary.*

The comparison is now made on equal terms: same economies, same laws, same rules. In the two economies of Table 1, the egalitarian-equivalent index hands almost every conflict to A under leximin and to B under leximax, and in the first economy the utilitarian and Nash rules each favor one agent as well. Percentiles give each agent half of the conflicts under all four rules. Even before

Table 1. Favor rate of agent A under the uniform product law: the egalitarian-equivalent calibration (EE) against percentiles, for the same laws and rules.

	EE				Percentiles			
K	leximin	leximax	util.	Nash	leximin	leximax	util.	Nash
A ranks by her smallest share, B by her average share								
2	0.78	0.30	0.48	0.69	0.49	0.51	0.50	0.49
4	0.96	0.09	0.41	0.80	0.49	0.51	0.50	0.49
16	1.00	0.00	0.26	0.91	0.49	0.51	0.50	0.49
64	1.00	0.00	0.14	0.95	0.50	0.50	0.50	0.50
CES with $\rho_A = 0$ *and* $\rho_B = \frac{1}{2}$								
2	0.57	0.47	0.52	0.55	0.50	0.50	0.50	0.50
4	0.63	0.42	0.52	0.58	0.50	0.50	0.50	0.50
16	0.78	0.27	0.52	0.60	0.50	0.50	0.50	0.50
64	0.94	0.08	0.52	0.61	0.50	0.50	0.50	0.50

Monte Carlo estimates from 4×10^5 pairs of drafts for each K; percentiles are computed from an independent sample of 10^6 drafts. Standard errors are below 0.002.

the limit, they stay within about one point of one half, including with two goods, where Section 4.5 shows that some rule must leave a gap. The law need not treat the agents alike. An asymmetric Q moves the typical scores of the index, and with them the identity of the near-dictator, but the theorem holds for every product law. With a density proportional to $1 + 3a$, which favors A in every good, the index hands B almost every conflict under leximin in the CES economy (A wins 7% of them with 64 goods), while percentiles give A between 49.4% and 50.3% under the four rules.

Why do percentiles succeed where the index fails? Because the percentile vector converges to an exchangeable law. When one agent is Leontief, her percentile is set by her scarcest good, and a single good carries a vanishing weight in the other agent's ranking: the two percentiles become independent. When both agents are CES, each ranks bundles by an average over goods. The central limit theorem makes the two averages jointly normal, and the percentile scale standardizes them; a standardized normal pair is exchangeable. The index does not standardize. It keeps the typical scores $c_A < c_B$ apart, and leximin reads the gap. Positional symmetry, which no scale can create, is here created by the economy itself. The bound $\rho_i > -\frac{1}{2}$ gives each average a finite variance. Below it, a few small shares dominate the average, much as in the Leontief case, and we expect the same conclusion without proving it. Simulations agree: with $\rho_A = -0.8$ and $\rho_B = \frac{1}{2}$ under the uniform law, the index lets A win more than 99.9% of leximin conflicts with 64 goods, and percentiles keep both agents within about one point of one half.

4.9. Public alternatives and reward models

So far each person receives her own bundle. Many collective choices instead select one alternative for everyone: a public project, or the response a language model gives to a prompt. The reward models used to fine-tune such models learn from pairwise comparisons. Two responses to a prompt are sampled, an annotator states which one she prefers, and a Bradley–Terry–Luce model is fitted to the labels. When annotators differ and the pairs are drawn uniformly from a finite set of responses, the fitted reward ranks responses by their Borda count over annotators: the probability that a response beats a random other one in the judgment of a random annotator (Siththaranjan et al., 2024).

Formally, let A be a compact, connected metric space of alternatives. Each agent i has a continuous weak order $\succsim_i$ on A under which some alternatives are strictly better than others. A *response law* μ on A has full support, and $\mu\{b : b \sim_i a\} = 0$ for every agent i and alternative a; the two alternatives compared are independent draws from μ, as in Assumption 1. Agent i's percentile of an alternative is

$U_i(a) = \mu\{b : b \precsim_i a\}$, the share of responses that she ranks no higher. Suppose that agent i labels a share w_i of the comparisons. Since ties have probability zero, the Borda score of a against a random response b is

$$\mathrm{BC}(a) = \sum_i w_i \, \mu\{b : a \succ_i b\} = \sum_i w_i \, U_i(a).$$

With equal shares, it is the mean percentile: the continuous analogue of the reward model's ranking is percentile utilitarianism over annotators. With unequal shares, it is a weighted sum of percentiles, the form that the argument for Theorem 4 gives without anonymity.

The guarantee in expectation carries over. A calibration now consists of continuous representations $m_i : A \to [0, 1]$ of the $\succsim_i$, each onto; they exist because A is compact, connected and separable. Score distributions, expected favor rates and rule-independent expected balance are defined as in Sections 2.3 and 4.3, with $a \sim \mu$ in place of a random proposal.

Proposition 5. *In the setting of this subsection, conditions* (a)–(d) *of Theorem 6 are equivalent for every calibration.*

Proof. The proof of Theorem 6 uses three facts about the primitives. The expected favor rates depend on the calibration only through the product $\bar{\nu}$ of the score distributions; each score distribution is atomless with full support on $[0, 1]$; and each percentile is a continuous representation onto $[0, 1]$ whose value at a random alternative is uniform. The first holds by definition. For the second, each level set of m_i is an indifference class, hence null, and the inverse image of an open subinterval of $[0, 1]$ is open and, since m_i is onto, nonempty, so it has positive probability. The distribution function Φ_i of $m_i(a)$ is therefore an increasing homeomorphism of $[0, 1]$, and $U_i = \Phi_i \circ m_i$ gives the third. □

In words, suppose several annotator groups each hold one ranking of responses, and their rewards are to be aggregated by an anonymous rule: the sum, the smallest reward (Chakraborty et al., 2024), or any other. The group scales that give every group the same prospects under every such rule are exactly the common transformations of the groups' percentiles under the response law. The choice is not idle. Rewards fitted group by group are identified only up to an additive constant for each group, so a rule such as the smallest reward depends on how these constants are set. Percentiles fix them, and the pooled estimator, through its Borda count, already uses them.

Two parts of the paper do not carry over. The converse half of Theorem 5 lowers one person's score through her own bundle, which a public alternative does not provide; and equal standing, which compares allocations that benefit one person alone, has no object. What transfers is the guarantee in expectation, together with the fact that a reward model's Borda count is a sum of percentiles (Online Appendix O7).

5. IMPLICATIONS FOR ASSESSMENT

Three data requirements follow. Percentile rankings need each person's ranking and the marginal distribution of her proposed bundles, and expected balance under every rule needs nothing more (Theorem 6). Checking whether they balance every anonymous rule also needs the joint distribution of percentile positions within proposals. Balance for a selected rule needs the distribution of comparisons the institution actually faces. Separate samples for each person answer the first question but not the second, and a few estimated favor rates speak to particular rules, not to the universal restriction in Theorem 5.

The relevant archive contains rejected as well as accepted proposals: implemented proposals have already passed through the rule and need not reveal the proposal distribution. Repeated assessment requires a stable submission process, and our model does not show that the process survives the announcement of a new criterion; an office that knows it could shape its drafts. Sampling and

reference-law errors can distort the constructed percentiles, and rare conflicts make a favor rate, which is conditional on disagreement, hard to estimate.

Nor do percentiles give an institution-independent standard. Under every proposal law allowed by Assumption 1, each person's percentile has mean one half by construction, which reveals nothing about whether two institutions generate equally good material outcomes. Comparing institutions requires a common reference or further welfare information. Likewise, ordinal rankings and an empirical proposal distribution cannot by themselves identify the disadvantages that deserve compensation.

Finally, all results rest on the independence benchmark. Reforms of a common baseline have a different pair law, even when their single-proposal marginals are unchanged, and their favor rates depend on that pair law. Our characterizations hold for independent proposals; we do not claim them for such dependent comparisons.

6. FINAL REMARKS

Measuring welfare in resources, combined with leximin, can let one taste win nearly every conflict under a given drafting process that treats everyone alike, and other rules change only which taste wins. The reason is general: an anonymous rule ignores names but needs welfare scores, which ordinal preferences leave open, and the scores alone can make the rule side with one person in nearly every conflict. Requiring that the ranking give everyone equal standing, and that it use frequencies only, leads to rankings through percentiles; equity then selects their leximin, and consistency across drafting processes their sum. These rankings treat people alike in expectation under every rule, and only percentile scales do. Whether they also treat people alike in actual decisions depends on how proposals pair people's fortunes, a fact that no scale can change. Reward models trained on random pairwise comparisons rank responses by the continuous analogue of the sum, so the guarantee in expectation, and its limit, apply to them.

The main limitation is the comparison process: our results hold for independent proposals from a fixed law, not for a reform compared with its status quo.

The results separate what impartial measurement requires from the distributive judgments it cannot settle. An institution that checks only whether its formula ignores names has checked the easy part.

A. PROOF OF NEAR-DICTATORSHIP

The proof has two parts. First, we build finite sequences of decisive score increases, without assuming that the aggregator is continuous. Second, increasing transformations concentrate each agent's score distribution near those levels, and we bound the exceptional comparisons. Lemmas O2 and O3 of the Online Appendix allow us to prescribe these marginal distributions.

A.1. The finite-grid lemma

The construction needs arbitrarily long finite sequences of decisive score increases, although the aggregator may be discontinuous. Write $[s;\mathbf{u}]_i$ for the score profile with coordinate s for agent i and other coordinates $\mathbf{u}$. Let $\mathbf{1}$ be the all-ones vector of length $n-1$ and $\max \mathbf{u}$ its largest coordinate.

Definition 4. R *satisfies* GC_K *if there are* $\mathbf{w}_- \ll \mathbf{w}_+$ *in* $[0,1]^{n-1}$*, where* $\ll$ *means strictly smaller in every coordinate, and* $\alpha_1 < \cdots < \alpha_K$ *in* $[0,1]$ *with* $[\alpha_{k+1};\mathbf{w}_-]_1 \,\mathrm{P}\, [\alpha_k;\mathbf{w}_+]_1$ *for* $k = 1,\dots,K-1$.

Lemma 1. *Every aggregator satisfies* GC_K *for every* $K \geq 1$.

Proof. Suppose by contradiction that, for some finite length, no grid exists. By anonymity, this failure occurs whichever agent we designate. We show that it creates a cycle that contradicts strong Pareto. GC_1 is trivial. Suppose GC_K fails for some $K \geq 2$. By anonymity the failure holds with any agent i in the role of agent 1: for every i, every $\mathbf{w}_- \ll \mathbf{w}_+$ and every $\alpha_1 < \cdots < \alpha_K$ there is $k < K$ with $[\alpha_k;\mathbf{w}_+]_i \,\mathrm{R}\, [\alpha_{k+1};\mathbf{w}_-]_i$.

Step 1 isolates the few score thresholds at which we cannot certify that one person's increase outweighs a small change in the others' scores. For a fixed finite grid, these exceptional intervals occupy only finitely many positions on each axis. *Step 1: negligibility.* Fix i, $\mathbf{u} \in (0,1)^{n-1}$ and $h \in (0, 1 - \max \mathbf{u}]$. Write $s' \succ_h s$ if $[s';\mathbf{u}]_i \,\mathrm{P}\, [s;\mathbf{u}+h\mathbf{1}]_i$. This local relation records a decisive increase in agent i's score against a simultaneous increase h in everyone else's. Then $s' \succ_h s$ implies $s' > s$ (else $[s;\mathbf{u}+h\mathbf{1}]_i$ dominates $[s';\mathbf{u}]_i$); $\succ_h$ is transitive; and $s_1 \geq s' \succ_h s \geq s_0$ implies $s_1 \succ_h s_0$ by strong Pareto. A chain $s_1 \prec_h s_2 \prec_h \cdots \prec_h s_K$ gives a GC_K grid, with $\mathbf{w}_- = \mathbf{u}$ and $\mathbf{w}_+ = \mathbf{u} + h\mathbf{1}$, so no chain has length K; let $\ell(s) \in \{1,\dots,K-1\}$ be the length of the longest chain ending at s, nondecreasing, so its level sets $L_j = \{\ell = j\}$ are consecutive intervals partitioning $[0,1]$, and within a level

$$[s;\mathbf{u}+h\mathbf{1}]_i \,\mathrm{R}\, [s';\mathbf{u}]_i \qquad (s < s' \text{ in the same } L_j). \tag{3}$$

Let $B_i(\mathbf{u})$ contain the boundaries between consecutive nonempty level sets of ℓ: there are at most $K-2$ such points. Empty level sets contribute no boundary. If $s < s'$ and no point of $B_i(\mathbf{u})$ lies in $[s,s']$, then s and s' lie in the same level set, since ℓ is nondecreasing, and (3) holds: a small common raise of all other agents outweighs any raise of agent i that crosses no breakpoint. The set $B_i(\mathbf{u})$ depends on h, which Step 2 fixes before choosing any point.

Step 2 chooses a shifted grid point outside every exceptional interval. The assumed failure of the grid then forces a chain of weak social comparisons around the agents. Yet the starting point of the chain strictly Pareto-dominates its endpoint. *Step 2: a cycle.* Fix an integer $M > 3n(K-2)$, $M \geq 2$, put $\delta = 1/(M+4)$, $h = \delta/n$, and $\boldsymbol{\tau}_{i+1} = \boldsymbol{\tau}_i - \delta\mathbf{e}_i + h\sum_{j\neq i}\mathbf{e}_j$ with $\boldsymbol{\tau}_1 = \mathbf{0}$ ($\mathbf{e}_j$ the unit vectors of $\mathbb{R}^n$). With this h, let D_i be the set of $\mathbf{v} \in (0,1)^n$ with $\max \mathbf{v}_{-i} \leq 1-h$ whose i-th coordinate lies at distance at most δ from a point of $B_i(\mathbf{v}_{-i})$; each axis-parallel line in direction i meets D_i in at most $K-2$ closed intervals of length 2δ, each containing at most three points of the shifted grid $G + \boldsymbol{\tau}_i$ defined below. On the grid $G = \{2\delta + \delta\mathbf{m} : \mathbf{m} \in \{1,\dots,M\}^n\}$, at most $3(K-2)M^{n-1}$ points are bad for a given i: fix the other $n-1$ grid coordinates and count the at most three points near each boundary. The total is at most $3n(K-2)M^{n-1} < M^n$. This is a finite count and requires no measurability of D_i. Thus some $\mathbf{z} \in G$ has $\mathbf{v}^{(i)} \equiv \mathbf{z} + \boldsymbol{\tau}_i \notin D_i$ for all i. All comparisons stay in their domains. Each coordinate of a shift $\boldsymbol{\tau}_i$ lies between $-\delta$ and $(n-1)h < \delta$, whereas grid coordinates lie between 3δ and $1-2\delta$. Consequently $v_i^{(i)} - \delta > 0$ and $\max \mathbf{v}_{-i}^{(i)} + h < 1$. At step i, (3) with $s = v_i^{(i)} - \delta$, $s' = v_i^{(i)}$

gives $\mathbf{v}^{(i+1)}\,\mathrm{R}\,\mathbf{v}^{(i)}$; by transitivity $\mathbf{v}^{(n+1)}\,\mathrm{R}\,\mathbf{z}$. But $\boldsymbol{\tau}_{n+1} = -(\delta - (n-1)h)\sum_i \mathbf{e}_i$ with $(n-1)h < \delta$, so $\mathbf{v}^{(n+1)}$ lies strictly below $\mathbf{z}$ in every coordinate, and strong Pareto gives $\mathbf{z}\,\mathrm{P}\,\mathbf{v}^{(n+1)}$, a contradiction. □

The contradiction shows that the strict comparisons needed for a grid exist for *every* fixed finite length. The exceptional intervals mark the places where our chosen comparisons cannot be certified. They do not say that an arbitrary aggregator has only finitely many discontinuities. Leximin, for example, may change which coordinate decides a comparison when two scores cross. The finite-grid construction avoids the relevant exceptional thresholds without assuming that the aggregator is continuous.

A.2. Calibrating the score distributions

For a calibration $\mathbf{m}$, write $M_i(\mathbf{x}) = m_i(x_i)$.

We now turn the finite grid into an actual calibration. The argument uses one auxiliary event:

$$\mathcal{A}_i(\mathbf{m}, \mathrm{R}) = \{(\mathbf{x}, \mathbf{y}) : x_i \succ_i y_i \implies \mathbf{x} \succ \mathbf{y}\}.$$

This *one-sided* event holds automatically if agent i does not prefer the first proposal. It is therefore *not* the conditional favor-rate event. To make its probability close to one, we give the designated agent scores near many separated grid rungs and keep the other agents in narrow score bands. A small full-support component keeps every score interval possible. The only important exceptions are pairs whose scores escape the bands, or that put the designated agent on the same rung in both proposals. Fix $K \geq 1$, put $\varepsilon = K^{-2}$. By Lemma 1 with $2K$ there are $\mathbf{w}_- \ll \mathbf{w}_+$ and $\gamma_1 < \cdots < \gamma_{2K}$ with

$$[\gamma_{j+1}; \mathbf{w}_-]_1\,\mathrm{P}[\gamma_j; \mathbf{w}_+]_1, \qquad j < 2K. \tag{4}$$

Let $J_k = [\gamma_{2k-1}, \gamma_{2k}]$, $J = \bigcup_k J_k$, and $B_l = [w_-^l, w_+^l]$ the band of agent $l \geq 2$. Define the atomless full-support laws

$$\nu_1 = (1-\varepsilon)\sum_{k=1}^{K} \tfrac{1}{K}\mathrm{Unif}(J_k) + \varepsilon\mathrm{Unif}[0,1], \qquad \nu_l = (1-\varepsilon)\mathrm{Unif}(B_l) + \varepsilon\mathrm{Unif}[0,1]\ (l \geq 2),$$

and let $\mathbf{m}^{(K)}$ realize them (Lemma O3). From now on $\mathbf{m}$ and M_i refer to $\mathbf{m}^{(K)}$, $\succ$ to $\succ^{\mathbf{m}^{(K)},\mathrm{R}}$, and F_1, O_1 to $F_1(\mathbf{m}^{(K)}, \mathrm{R})$, $O_1(\mathbf{m}^{(K)}, \mathrm{R})$. Let Γ be the set of pairs with $M_1(\mathbf{x}) \in J_j$, $M_1(\mathbf{y}) \in J_k$ for $j \neq k$, and $M_l(\mathbf{x}), M_l(\mathbf{y}) \in B_l$ for all $l \geq 2$. On Γ with $j > k$, strong Pareto and (4) give $\mathbf{m}(\mathbf{x})\,\mathrm{R}[\gamma_{2k+1}; \mathbf{w}_-]_1\,\mathrm{P}[\gamma_{2k}; \mathbf{w}_+]_1\,\mathrm{R}\,\mathbf{m}(\mathbf{y})$, so $\mathbf{x} \succ \mathbf{y}$, and exactly one of $x_1 \succ_1 y_1$, $y_1 \succ_1 x_1$ holds, the social ranking following it; hence $\Gamma \subseteq \mathcal{A}_1(\mathbf{m}^{(K)}, \mathrm{R})$ and on Γ, $\mathbf{x} \succ \mathbf{y} \iff x_1 \succ_1 y_1$. Finally Γ^c is covered by $2n$ escape events of probability $\leq \varepsilon$ and the same-block event of probability $\sum_k \nu_1(J_k)^2 \leq K(1/K + \varepsilon)^2$, so

$$\pi\big(\mathcal{A}_1(\mathbf{m}^{(K)}, \mathrm{R})^c\big) \leq \pi(\Gamma^c) \leq \frac{1}{K} + \frac{2n+2}{K^2} + \frac{1}{K^3} \longrightarrow 0.$$

It remains to show that $f_1(\mathbf{m}^{(K)}, \mathrm{R}) \to 1$. Let E be the set of pairs with $x_1 \succ_1 y_1$ and $y_2 \succ_2 x_2$. It does not depend on the calibration, and $c_E \equiv \pi(E) > 0$: the pair $(\mathbf{e}_1(1_\Omega), \mathbf{e}_2(1_\Omega))$ lies in E with both preferences strict, so by continuity of the preferences a nonempty relatively open set of pairs does, and π has full support. On $E \cap \Gamma$, $\mathbf{x} \succ \mathbf{y}$, so agent 1 is favored: $\pi(F_1) \geq c_E - \pi(\Gamma^c)$. If $(\mathbf{x}, \mathbf{y}) \in O_1$, then $(\mathbf{y}, \mathbf{x}) \notin \mathcal{A}_1(\mathbf{m}^{(K)}, \mathrm{R})$, and π is invariant under exchanging the two draws, so $\pi(O_1) \leq \pi(\mathcal{A}_1(\mathbf{m}^{(K)}, \mathrm{R})^c) \leq \pi(\Gamma^c)$. For large K, $c_E - \pi(\Gamma^c) > 0$, so the favor rate is defined. Discarding finitely many initial terms and reindexing gives a sequence with defined favor rates throughout. Since $t/(t+s)$ is increasing in t and decreasing in s,

$$f_1(\mathbf{m}^{(K)}, \mathrm{R}) \geq \frac{c_E - \pi(\Gamma^c)}{c_E} \longrightarrow 1.$$

Expected favor rates. The same calibrations work under $\bar{\pi}$. The law $\bar{\nu}$ has the marginals $\nu_1, \dots, \nu_n$ above and independent coordinates. Let $\bar{\Gamma}$ be the set of score pairs $(\mathbf{a}, \mathbf{b})$ with $a_1 \in J_j$, $b_1 \in J_k$ for $j \neq k$, and $a_l, b_l \in B_l$ for all $l \geq 2$. On $\bar{\Gamma}$ with $j > k$, strong Pareto and (4) give $\mathbf{a}\,\mathrm{R}[\gamma_{2k+1}; \mathbf{w}_-]_1\,\mathrm{P}[\gamma_{2k}; \mathbf{w}_+]_1\,\mathrm{R}\,\mathbf{b}$, so $\mathbf{a}\,\mathrm{P}\,\mathbf{b}$; the case $j < k$ is symmetric. On $\bar{\Gamma}$ the rule therefore follows agent 1's strict preference. The complement of $\bar{\Gamma}$ is covered by the same $2n$ escape events and the same same-block event as before. Each concerns one coordinate of one vector, or the same coordinate of both vectors, so its $\bar{\pi}$-probability is the one computed above, and $\bar{\pi}(\bar{\Gamma}^c) \leq 1/K + (2n+2)/K^2 + 1/K^3 \to 0$. Let $\bar{E}$ be the set of score pairs with $a_1 > b_1$ and $b_2 > a_2$. Under $\bar{\pi}$ the coordinates are independent and atomless, so $\bar{\pi}(\bar{E}) = \frac{1}{4}$ whatever the calibration. On $\bar{E} \cap \bar{\Gamma}$ agent 1 is favored, so $\bar{\pi}(\bar{F}_1) \geq \frac{1}{4} - \bar{\pi}(\bar{\Gamma}^c)$. If $(\mathbf{a}, \mathbf{b}) \in \bar{O}_1$, the rule does not follow agent 1's strict preference at $(\mathbf{b}, \mathbf{a})$, so $(\mathbf{b}, \mathbf{a}) \notin \bar{\Gamma}$; exchanging the two draws leaves $\bar{\pi}$ unchanged, so $\bar{\pi}(\bar{O}_1) \leq \bar{\pi}(\bar{\Gamma}^c)$. Hence

$$\bar{f}_1(\mathbf{m}^{(K)}, \mathrm{R}) \geq \frac{\frac{1}{4} - \bar{\pi}(\bar{\Gamma}^c)}{\frac{1}{4}} \longrightarrow 1. \qquad \square$$

B. PROOFS FOR STANDARD CALIBRATIONS

We first record the properties of the egalitarian-equivalent calibration that the proofs use. Write 1 for 1_Ω.

Lemma 2. *For every agent i and bundle x there is a unique $t \in [0,1]$ with $x \sim_i t\,1$, and $m_i(x) = t$ defines a calibration. Moreover $\min_\omega x(\omega) \leq m_i(x) \leq \max_\omega x(\omega)$. If i has CES preferences with equal weights and parameter ρ_i, m_i is the power mean of order ρ_i of her shares; in particular, $m_i(x) = \min_\omega x(\omega)$ if she ranks bundles by her smallest share, and $m_i(x) = \frac{1}{|\Omega|}\sum_\omega x(\omega)$ if she ranks them by her average share.*

Proof. By monotonicity, $(\min x)\,1 \precsim_i x \precsim_i (\max x)\,1$. The map $s \mapsto s\,1$ is continuous and strictly increasing for $\succ_i$, since $s'1$ exceeds $s1$ in every good when $s' > s$. Hence $\{s : s1 \succsim_i x\}$ is a closed interval $[t, 1]$ and $\{s : x \succsim_i s1\} = [0, t]$, which gives existence, uniqueness and the bounds. If $m_i(x_k) \geq t + \varepsilon$ along a sequence $x_k \to x$, then $x_k \succsim_i (t+\varepsilon)1$, so $x \succsim_i (t+\varepsilon)1 \succ_i t1 \sim_i x$ by continuity of $\succsim_i$, a contradiction; the same argument applies from below, so m_i is continuous. Since $m_i(s1) = s$, it maps onto $[0, 1]$. For CES preferences, the power mean u_i represents $\succsim_i$ and satisfies $u_i(t1) = t$, so $m_i = u_i$. $\square$

Two facts are used throughout. Under leximin, leximax, the utilitarian rule and the Nash rule, a social tie requires an equality between continuous functions of two independent proposals, a null event, so by Lemma O1 the denominators of the favor rates equal half the probability of a conflict involving A, under π as under $\bar{\pi}$. And since $t \mapsto t/(t+s)$ increases in t and decreases in s, a lower bound on $\pi(F_A)$ and an upper bound on $\pi(O_A)$ bound f_A from below.

With two agents, in a conflict where A prefers $\mathbf{x}$ and B prefers $\mathbf{y}$, let $w_A = m_A(x_A)$, $\ell_A = m_A(y_A)$, $w_B = m_B(y_B)$ and $\ell_B = m_B(x_B)$. Leximin sides with A if and only if $\ell_A < \ell_B$ (Proposition 6); leximax if and only if $w_A > w_B$, by the same argument; the utilitarian rule if and only if $w_A - \ell_A > w_B - \ell_B$; and the Nash rule if and only if $\log(w_A/\ell_A) > \log(w_B/\ell_B)$.

Proof of Proposition 1

Let $a = m_A(z)$, $b = \min_{j \neq A} m_j(z)$ and $\delta = (b-a)/3 > 0$. By continuity of the scores, choose $r > 0$ so small that $z(\omega) + r < \frac{1}{n}$ for every good and that, on $N = \{x \in X : |x(\omega) - z(\omega)| < r \text{ for all } \omega\}$, $m_A < a + \delta$ and $m_j > b - \delta$ for every $j \neq A$. Let $G = N^n$, the allocations that give every agent a bundle in N. They are feasible, since each good is distributed in a total amount below $n \cdot \frac{1}{n}$. The set G is relatively open in $\mathcal{X}$, has positive Lebesgue measure and is invariant to permutations of the agents.

Let λ be the uniform law on $\mathcal{X}$, λ_G the uniform law on G, and $\mu_k = (1-\frac{1}{k})\lambda_G + \frac{1}{k}\lambda$. Each μ_k has a full-support density on $\mathcal{X}$, so $\mu_k \otimes \mu_k$ satisfies Assumption 1 (Online Appendix O1), and it is invariant to permutations of the agents. Moreover $\mu_k(\mathcal{X} \setminus G) \le \frac{1}{k}$.

Actual favor rate. If $\mathbf{x}, \mathbf{y} \in G$, agent A's score in each proposal is below $a+\delta < b-\delta$, hence below every other agent's score in that proposal. Leximin therefore compares A's two scores first and chooses the proposal A prefers. No pair of $G \times G$ lies in O_A, and $\pi_k(O_A) \le \pi_k((G\times G)^c) \le 2\mu_k(G^c) \le \frac{2}{k}$. For a lower bound on $\pi_k(F_A)$, let $z^+ = z + \frac{r}{2}\,1 \in N$, pick $j \neq A$, and consider the allocations $\mathbf{x}$, which gives z^+ to A and z to everyone else, and $\mathbf{y}$, which gives z^+ to j and z to everyone else. Agent A strictly prefers $\mathbf{x}$ and agent j strictly prefers $\mathbf{y}$. By continuity of preferences, both strict preferences persist on a relatively open neighborhood $W \subseteq G \times G$ of this pair, with $c = \lambda_G \otimes \lambda_G(W) > 0$. On W leximin chooses the first proposal, so $W \subseteq F_A$ and $\pi_k(F_A) \ge (1-\frac{1}{k})^2 c$. Hence

$$f_A(\mathbf{m}, \mathrm{R}) \;\ge\; \frac{(1-\frac{1}{k})^2 c}{(1-\frac{1}{k})^2 c + \frac{2}{k}} \;\longrightarrow\; 1.$$

Expected favor rate. Under μ_k, each of the events $\{m_A(X_A) \ge a+\delta\}$ and $\{m_j(X_j) \le b-\delta\}$, $j \neq A$, requires $\mathbf{X} \notin G$ and has probability at most $\frac{1}{k}$. Under $\bar\pi$ the $2n$ scores are independent, so the event that both of A's scores are below $a+\delta$ and both scores of every other agent are above $b-\delta$ has probability at least $1-\frac{2n}{k}$. On it leximin follows A, so $\bar\pi(\bar O_A) \le \frac{2n}{k}$. The scores are independent and atomless, so a pair is a conflict involving A with the first proposal chosen with probability $\bar q = \frac{1}{2}(1-2^{1-n})$, and $\bar\pi(\bar F_A) = \bar q - \bar\pi(\bar O_A)$. Therefore $\bar f_A \ge 1 - \frac{2n}{k\bar q} \to 1$.

Leximax, $n=2$. On $G \times G$, A's score in her preferred proposal is below $a+\delta < b-\delta$, hence below B's score in his preferred proposal; leximax always chooses the proposal B prefers. The same argument with F_A and O_A exchanged gives $f_A \to 0$ and $\bar f_A \to 0$.

The last claim. If $m_A = m_B$ on the open box $\{z : z(\omega) < \frac{1}{2} \text{ for all } \omega\}$, the two agents rank alike all bundles in it. Otherwise some z in the box has $m_A(z) \neq m_B(z)$, and the first part applies to the agent with the lower score at z. □

Proof of Theorem 2

Let $0 < \underline{q} \le \overline{q}$ bound the density of Q on T. Fix an agent, write a_ω for her share of good ω under μ_K, and F for the distribution function of her share a under Q. The density of a at t integrates the density of Q over the other agents' shares, which range over a set of volume $(1-t)^{n-1}/(n-1)!$; it therefore lies between $\underline{q}\,(1-t)^{n-1}/(n-1)!$ and $\overline{q}/(n-1)!$. In particular $F(u) \ge \underline{q}\,(1-(1-u)^n)/n! \ge \underline{q}\,u/n!$, so that

$$\mathbb{P}\big(\min_\omega a_\omega > u\big) = (1-F(u))^K \le e^{-K\underline{q}u/n!}. \tag{5}$$

With two agents, we keep a_ω for A's share, write b_ω for B's, and $\mu_b > 0$ and $\sigma_b^2 > 0$ for the mean and the variance of b; the density bounds then read $(1-t)\underline{q}$ and $\overline{q}$, and the exponent in (5) is $K\underline{q}u/2$. Each μ_K has a full-support density on $\mathcal{X}$ and satisfies Assumption 1.

Lemma 3. *As $K \to \infty$, $m_i(\mathbf{x}) \to c_i$ in probability for each agent.*

Proof. Let s be the agent's share of a good. Its density is bounded, and bounded away from zero near 0, so $\mathbb{E}[s^\rho] < \infty$ if and only if $\rho > -1$, and $\mathbb{E}|\log s| < \infty$. For $\rho_i \in (-1, 1]$, the weak law of large numbers applied to the $s_\omega^{\rho_i}$, or to the $\log s_\omega$, and the continuity of $u \mapsto u^{1/\rho_i}$ give the claim. For finite $\rho_i \le -1$, the average of the $s_\omega^{\rho_i}$ exceeds that of the $\min(s_\omega^{\rho_i}, L)$, whose limit $\mathbb{E}[\min(s^{\rho_i}, L)]$ tends to infinity with L; hence the average tends to infinity in probability and $m_i \to 0$. For $\rho_i = -\infty$, use (5). □

The lemma decides who wins inside a conflict. It does not guarantee that conflicts remain: two nearly constant scores could, in principle, almost always move together.

Lemma 4. *Let A and B be two agents with $\rho_B > -\frac{1}{3}$. For any preference of A that depends on her shares only and is strict between two independent proposals almost surely, the probability that A and B disagree on a pair of proposals has a positive lower bound that does not depend on K.*

Proof. Let h be the increasing transformation that turns m_B into a sum: $h(b) = b^{\rho_B}$, $\log b$ or $-b^{\rho_B}$ as ρ_B is positive, zero or negative. Agent B prefers $\mathbf{x}$ to $\mathbf{y}$ if and only if $\sum_\omega \xi_\omega > 0$, where $\xi_\omega = h(b^x_\omega) - h(b^y_\omega)$; since the density of b is bounded and $\rho_B > -\frac{1}{3}$, $\mathbb{E}|h(b)|^3 < \infty$. Let $\mathcal{A}$ be the σ-field generated by A's shares in both proposals; it determines the direction $\sigma = \pm 1$ of A's preference, and the two agents disagree if and only if $\sigma \sum_\omega \xi_\omega < 0$. Given $\mathcal{A}$, the ξ_ω are independent, with means $e_\omega = \beta(a^x_\omega) - \beta(a^y_\omega)$ and variances $v_\omega = \gamma(a^x_\omega) + \gamma(a^y_\omega)$, where $\beta(a)$ and $\gamma(a)$ are the conditional mean and variance of $h(b)$ given a. Let $e = \sum_\omega e_\omega$ and $\tau^2 = \sum_\omega v_\omega$. By the Berry–Esseen inequality for independent, non-identical summands,

$$\mathbb{P}(\text{disagreement} \mid \mathcal{A}) \;\geq\; \Phi\big(-|e|/\tau\big) - \beta_K, \qquad \beta_K = \min\Big(1, \frac{C_0}{\tau^3} \sum_\omega \mathbb{E}\big[|\xi_\omega - e_\omega|^3 \;\big|\; \mathcal{A}\big]\Big),$$

with a universal constant C_0. Given a, b has a density that is positive on $(0, 1-a)$, so $\bar\gamma = \mathbb{E}[\gamma(a)] > 0$; by the strong law of large numbers, $\tau^2/K \to 2\bar\gamma$ and the sum of third moments divided by K converges to a finite constant, so $\beta_K \to 0$ almost surely and $\mathbb{E}[\beta_K] \to 0$. The e_ω are independent and centered with variance at most $2\,\mathbb{E}[h(b)^2]$, so by Chebyshev's inequality $|e| \leq M\sqrt{K}$ with probability at least $\frac{3}{4}$ for some M. On the intersection of this event with $\{\tau^2 \geq K\bar\gamma\}$, whose probability exceeds $\frac{1}{2}$ for large K, $\Phi(-|e|/\tau) \geq \Phi(-M/\sqrt{\bar\gamma})$. Hence $\liminf_K \mathbb{P}(\text{disagreement}) \geq \frac{1}{2}\Phi(-M/\sqrt{\bar\gamma}) > 0$. Since the probability of disagreement is positive for each K, its infimum is positive. □

Proof of part (a). Let $t = \frac{1}{2}(c_A + \min_{j\neq A} c_j)$. Suppose that, in both proposals of a pair, A's score is below t and every other agent's score is above t. Then A holds the lowest score in both, and leximin sides with her, since the lowest scores differ whenever her ranking is strict. Hence

$$\pi(O_A) \leq \varepsilon_K = 2\,\mathbb{P}(m_A(X_A) > t) + 2\sum_{j\neq A} \mathbb{P}(m_j(X_j) < t) \to 0$$

by Lemma 3. Take any $j \neq A$. Since at most one agent has $\rho_i \leq -\frac{1}{3}$, one of A and j has a parameter above $-\frac{1}{3}$; Lemma 4, with that agent in the role of B, shows that A and j disagree with probability at least some $q_0 > 0$, since CES and Leontief preferences are strict between independent proposals almost surely. Each such pair lies in F_A or O_A unless leximin declares a tie, and a tie has probability zero because the two proposals are independent and every score has an atomless law. Hence $\pi(F_A) + \pi(O_A) \geq q_0$, and $f_A \geq 1 - \varepsilon_K/q_0 \to 1$. For leximax, let $t = \frac{1}{2}(c_Z + \max_{j\neq Z} c_j)$. If, in both proposals, Z's score is above t and every other agent's score is below t, then Z holds the highest score in both and leximax sides with her; the same argument gives $f_Z \to 1$. Under $\bar\pi$, the agents' rankings are independent and strict almost surely, so A disagrees with any given agent with probability exactly $\frac{1}{2}$, and the same bounds apply.

Proof of part (b). Here $n = 2$. Write $U = \min_\omega a_\omega$ and $V = \frac{1}{K}\sum_\omega b_\omega$ for the scores of a proposal (Lemma 2), with indices x and y for the two proposals. By Lemma 4, with $\rho_B = 1$, the denominators are bounded below by $q_0 > 0$ for every K, so it suffices to bound the probability of the unfavorable events. Hoeffding's inequality gives $\mathbb{P}(V \leq \mu_b - \varepsilon) \leq e^{-2K\varepsilon^2}$ and $\mathbb{P}(|V_x - V_y| \geq \varepsilon) \leq 2e^{-K\varepsilon^2/2}$.

Leximin. With $t = \mu_b/2$, the proof of part (a), together with (5) and Hoeffding's inequality, gives $\pi(O_A) \leq 2e^{-K\underline{q}\mu_b/4} + 2e^{-K\mu_b^2/2}$.

Utilitarian rule. It sides with A only if $|U_x - U_y| \geq |V_x - V_y|$. Since $|U_x - U_y| \leq \max(U_x, U_y)$, (5) with $u = M/K$ gives

$$\mathbb{P}\big(|U_x - U_y| \geq |V_x - V_y|\big) \leq 2e^{-\underline{q}M/2} + \mathbb{P}\big(\sqrt{K}|V_x - V_y| \leq M/\sqrt{K}\big).$$

The variable $\sqrt{K}(V_x - V_y)$ is a normalized sum of K independent, centered, bounded terms with variance $2\sigma_b^2$; by the Berry–Esseen inequality, the last probability is at most $M/(\sigma_b\sqrt{\pi K}) + C_1/\sqrt{K}$. Taking $M = (2/\underline{q}) \log K$ gives the bound.

Nash rule. It rules against A only if $|\log U_x - \log U_y| \le |\log V_x - \log V_y|$. Outside an event of probability at most $2e^{-K\mu_b^2/2}$, both V_x and V_y exceed $\mu_b/2$, and then $|\log V_x - \log V_y| \le 2|V_x - V_y|/\mu_b$. For $s = M/\sqrt{K}$, $\mathbb{P}(2|V_x - V_y|/\mu_b > s) \le 2e^{-M^2\mu_b^2/8}$. The density of $\log U$ at $\log u$ is $uKf(u)(1 - F(u))^{K-1}$, where $f \le \overline{q}$ is the density of a; by (5) it is at most $\overline{q}\, Ku\, e^{-(K-1)\underline{q}u/2} \le G = 4\overline{q}/(e\underline{q})$ for $K \ge 2$. The density of the difference of two independent copies of $\log U$ is bounded by the same constant, so $\mathbb{P}(|\log U_x - \log U_y| \le s) \le 2Gs$. Taking $M = \sqrt{8 \log K}/\mu_b$ gives the bound.

Under $\bar{\pi}$, each bound involves the two scores of one agent only, or separate events for the two agents, so it holds unchanged, with denominators bounded below by $\frac{1}{4}$. □

Proof of Theorem 7

Write U_A and U_B for the two agents' percentiles of a random proposal under μ_K, and ν_K for the law of (U_A, U_B). Both coordinates are uniform on $[0, 1]$. Since each U_i represents $\gtrsim_i$, a pair of proposals is in F_A or O_A exactly when the pair of their percentile vectors, two independent draws from ν_K, is in

$$E^+ = \{(\mathbf{a}, \mathbf{b}) : \mathbf{a}\,\mathrm{P}\,\mathbf{b},\ a_A > b_A,\ b_B > a_B\} \quad \text{or} \quad E^- = \{(\mathbf{a}, \mathbf{b}) : \mathbf{a}\,\mathrm{P}\,\mathbf{b},\ a_A < b_A,\ a_B > b_B\}.$$

Step 1: from the limit law to favor rates. Suppose that ν_K converges weakly to a law ν that has a positive density on $(0, 1)^2$ and is exchangeable. Then $\nu_K \otimes \nu_K$ converges weakly to $\nu \otimes \nu$, which has a positive density on $(0, 1)^4$. The boundaries of E^+ and E^- lie in the boundary of $\{\mathbf{a}\,\mathrm{P}\,\mathbf{b}\}$ together with $\{a_A = b_A\}$ and $\{a_B = b_B\}$, a Lebesgue-null set for an aggregator with a null boundary. By the portmanteau theorem, $\nu_K \otimes \nu_K(E^\pm) \to \nu \otimes \nu(E^\pm)$. Exchanging the two coordinates in both vectors preserves $\nu \otimes \nu$ and, by anonymity, the relation P, and it maps E^+ onto E^-; hence $\nu \otimes \nu(E^+) = \nu \otimes \nu(E^-) =: L$. Exchanging the two vectors also preserves $\nu \otimes \nu$, so the conflicts that the aggregator resolves have probability $4L$. Social ties are null: if $\mathbf{a}\,\mathrm{I}\,\mathbf{b}$ with $\mathbf{a} \in (0, 1)^2$, then $(\mathbf{a} + \varepsilon\mathbf{1})\,\mathrm{P}\,\mathbf{b}$ for small $\varepsilon > 0$ by strong Pareto, so $(\mathbf{a}, \mathbf{b})$ lies on the boundary of $\{\mathbf{a}\,\mathrm{P}\,\mathbf{b}\}$. Conflicts have positive probability under a positive density, so $L > 0$, and $f_A = \nu_K \otimes \nu_K(E^+)/\nu_K \otimes \nu_K(E^+ \cup E^-) \to \frac{1}{2}$.

It remains to identify the limit. For $\rho \in (-\frac{1}{2}, 1]$ let $\phi_\rho(s) = s^\rho/\rho$ if $\rho \ne 0$ and $\phi_0(s) = \log s$. A CES agent with parameter ρ_i ranks bundles by $S_i = \frac{1}{K}\sum_\omega \phi_{\rho_i}(s_\omega)$, where s_ω is her share of good ω, since her power mean is an increasing function of S_i. The density of a share is bounded by $\overline{q}$, so $\mathbb{E}[\phi_\rho(s)^2] < \infty$ exactly when $\rho > -\frac{1}{2}$; let e_i and $\sigma_i^2 > 0$ be the mean and the variance of $\phi_{\rho_i}(s)$, $W_i = \sqrt{K}(S_i - e_i)/\sigma_i$, and $F_{i,K}$ the distribution function of S_i. By the central limit theorem and the continuity of Φ, the standard normal distribution function, $F_{i,K}(e_i + \sigma_i w/\sqrt{K}) \to \Phi(w)$ uniformly in w. Hence

$$U_i = F_{i,K}(S_i) = \Phi(W_i) + \varepsilon_K, \qquad |\varepsilon_K| \le \sup_w \left|F_{i,K}(e_i + \sigma_i w/\sqrt{K}) - \Phi(w)\right| \to 0. \tag{6}$$

Step 2: two CES agents. The vectors $(\phi_{\rho_A}(a_\omega), \phi_{\rho_B}(b_\omega))$ are independent and identically distributed with finite variances, so (W_A, W_B) converges in law to a standard bivariate normal pair with correlation $r = \mathrm{corr}(\phi_{\rho_A}(a), \phi_{\rho_B}(b))$. Since (a, b) has a density on T, the two functions are not affinely related, and $|r| < 1$. By (6), ν_K converges to the law of $(\Phi(Z_A), \Phi(Z_B))$ for this normal pair, which is exchangeable and has a positive density on $(0, 1)^2$.

Step 3: one Leontief agent. Let A rank bundles by her smallest share and B be CES with $\rho_B \in (-\frac{1}{2}, 1]$; write $Z_\omega = \phi_{\rho_B}(b_\omega)$, $e = e_B$, $\sigma = \sigma_B$. We show that $\mathbb{P}(U_A \le s, U_B \le t) \to st$ for all $s, t \in (0, 1)$, so that ν is the uniform law on $[0, 1]^2$. Choose τ_K with $\mathbb{P}(\min_\omega a_\omega \le \tau_K) = s$, that is, $F(\tau_K) = p_K = 1 - (1 - s)^{1/K}$; then $Kp_K \to -\log(1 - s)$. Let N be the number of goods with $a_\omega \le \tau_K$, a binomial variable with parameters K and p_K, so that $\{U_A \le s\} = \{N \ge 1\}$ and $\mathbb{P}(N > M) \le 2 \log \frac{1}{1-s}/M$ for large K.

Fix $j \geq 1$ and condition on the set of goods with $a_\omega \leq \tau_K$ being $\{1, \dots, j\}$. The pairs (a_ω, b_ω) remain independent, with law $Q(\cdot \mid a \leq \tau_K)$ for $\omega \leq j$ and $Q_K^+ = Q(\cdot \mid a > \tau_K)$ for $\omega > j$. Given $a \leq \tau$, the density of b is at most $\overline{q}\tau/F(\tau) \leq 2\overline{q}/\underline{q}$, so $\mathbb{E}[Z^2 \mid a \leq \tau_K] \leq C$ for a constant C; the first j terms of $\sum_\omega Z_\omega$ are therefore of order one. Under Q_K^+, the density of Z relative to its law under Q is at most $1/(1-p_K) \leq 2$, and its mean and variance differ from e and σ^2 by at most a constant times p_K. Hence the Lindeberg condition holds for the terms $\omega > j$, and

$$W_B = \frac{1}{\sigma\sqrt{K}}\Big[\sum_{\omega \leq j} Z_\omega + \sum_{\omega > j}\big(Z_\omega - \mathbb{E}_{Q_K^+} Z\big) + (K-j)\,\mathbb{E}_{Q_K^+} Z - Ke\Big]$$

converges in law to a standard normal variable: the first and last terms in the bracket are of order one. By (6), $\mathbb{P}(U_B \leq t \mid N = j) \to t$, by exchangeability of the goods for every set of j goods. Therefore

$$\Big|\mathbb{P}(U_A \leq s,\, U_B \leq t) - st\Big| = \Big|\sum_{j \geq 1} \mathbb{P}(N=j)\big(\mathbb{P}(U_B \leq t \mid N=j) - t\big)\Big| \leq \max_{j \leq M}\big|\mathbb{P}(U_B \leq t \mid N=j) - t\big| + \mathbb{P}(N > M),$$

which tends to zero by letting $K \to \infty$ and then $M \to \infty$. The case of a Leontief B and a CES A is symmetric.

Step 4: two Leontief agents. With τ_K as above, choose τ'_K with $\mathbb{P}(b \leq \tau'_K) = p'_K = 1 - (1-t)^{1/K}$. Then

$$\mathbb{P}(U_A > s,\, U_B > t) = \big(1 - p_K - p'_K + \mathbb{P}(a \leq \tau_K,\, b \leq \tau'_K)\big)^K \to (1-s)(1-t),$$

since $\mathbb{P}(a \leq \tau_K, b \leq \tau'_K) \leq \overline{q}\,\tau_K\tau'_K$ and $\tau_K, \tau'_K \leq 2p_K/\underline{q}$ are of order $1/K$. With uniform marginals, $\mathbb{P}(U_A \leq s, U_B \leq t) \to st$, and ν is again uniform on $[0,1]^2$. In each case Step 1 applies. □

C. PROOFS FOR SOCIAL ORDERING FUNCTIONS

Throughout, a social ordering function and an environment $\xi = (\gtrsim, \mu)$ are fixed when not stated otherwise, and we write U_i for U_i^ξ. By Lemma O2, each U_i is a continuous representation of $\gtrsim_i$ onto $[0,1]$, and $U_i(X_i)$ is uniform when $\mathbf{X}$ has law μ. The map $U^\xi : X^n \to [0,1]^n$ is onto.

Proof of Proposition 4

Fix a position (j,y) with $s = U_j(y)$, an agent i, and let $A = \{x \in X : (i,x) \sqsupset_\xi (j,y)\}$ and $B = X \setminus A$. By (P), A is an upper set for $\gtrsim_i$: if $x' \in A$ and $x \gtrsim_i x'$, then $\mathbf{e}_i(x) \succeq_\xi \mathbf{e}_i(x') \succ_\xi \mathbf{e}_j(y)$. Hence B is a lower set, and it is closed by (C). Suppose $B \neq \varnothing$. Since X is compact and U_i continuous, U_i attains its maximum t_0 on B at some $z \in B$. Every x with $U_i(x) \leq t_0$ satisfies $x \lesssim_i z$ and lies in B, and every $x \in B$ has $U_i(x) \leq t_0$. Thus $A = \{x : U_i(x) > t_0\}$, and $\mu\{\mathbf{x} : x_i \in A\} = 1 - t_0$ because $U_i(X_i)$ is uniform.

Apply (ES) with $i' = j$. For agent j the set is $\{x : \mathbf{e}_j(x) \succ_\xi \mathbf{e}_j(y)\} = \{x : x \succ_j y\} = \{x : U_j(x) > s\}$ by (P), and its probability is $1 - s$. Hence $t_0 = s$ and $A = \{x : U_i(x) > s\}$. If $B = \varnothing$, then $A = X$ has probability one, so $s = 0$; then $y \sim_j 0_\Omega$, and $\mathbf{e}_j(y) \sim_\xi \mathbf{e}_j(0_\Omega) = \mathbf{e}_i(0_\Omega)$ contradicts $0_\Omega \in A$. Therefore $(i,x) \sqsupset_\xi (j,y)$ if and only if $U_i(x) > U_j(y)$, for all positions. Exchanging the two positions, $(i,x) \sqsupseteq_\xi (j,y)$ if and only if not $(j,y) \sqsupset_\xi (i,x)$, that is, if and only if $U_i(x) \geq U_j(y)$. □

Proof of Theorem 3

Necessity. By Proposition 4, $s_\xi(i,x \mid k) = \mu\{\mathbf{x}' : U_k(x'_k) \leq U_i(x)\} = U_i(x)$ for every k, since $U_k(X_k)$ is uniform. Frequency independence therefore requires that $U^\xi(\mathbf{x}) = U^{\xi'}(\mathbf{x}')$ and $U^\xi(\mathbf{y}) = U^{\xi'}(\mathbf{y}')$ imply the same ranking. Define $\mathbf{a}\,\mathrm{R}\,\mathbf{b}$ if $\mathbf{x} \succeq_\xi \mathbf{y}$ for some $\xi, \mathbf{x}, \mathbf{y}$ with $U^\xi(\mathbf{x}) = \mathbf{a}$ and $U^\xi(\mathbf{y}) = \mathbf{b}$. Such triples exist because U^ξ is onto, and frequency independence makes the definition independent of the triple. Completeness and transitivity pass to R from any single $\succeq_\xi$ through preimages. If $\mathbf{a} \geq \mathbf{b}$ and $\mathbf{a} \neq \mathbf{b}$, preimages satisfy $x_i \gtrsim_i y_i$ for all i with one strict preference, so (P) gives $\mathbf{a}\,\mathrm{P}\,\mathbf{b}$. Anonymity of R follows from (A) and $U^{\sigma\xi}(\sigma\mathbf{x}) = \sigma U^\xi(\mathbf{x})$. Uniqueness is immediate.

Sufficiency. Strong Pareto holds because each U_i^ξ represents $\succsim_i$ and R is strongly Paretian; anonymity by the permutation identity. Anonymity and strong Pareto of R rank $t\mathbf{e}_i$ strictly above $s\mathbf{e}_j$ when $t > s$ and as indifferent when $t = s$. Hence $\{x : (i,x) \sqsupset_\xi (j,y)\} = \{x : U_i(x) > U_j(y)\}$: it is open, which is (C), and its probability is $1 - U_j(y)$ for every i, which is (ES). Finally $s_\xi(i, x \mid k) = U_i^\xi(x)$, so equal frequencies give equal percentile vectors and the same ranking: (FI).

Continuity. If R is continuous, every $\succsim_\xi$ is, because U^ξ is continuous. Conversely, the map $t \mapsto U_i^\xi(t1_\Omega)$ is continuous and strictly increasing from 0 to 1, because raising every coordinate is strictly preferred; let τ_i be its inverse. The map $\mathbf{a} \mapsto (\tau_i(a_i)1_\Omega)_i$ is a continuous right inverse of U^ξ, and the contour sets of R are its preimages of the closed contour sets of $\succsim_\xi$. □

Proof of Corollary 1

Percentile leximin satisfies the first five conditions by Theorem 3, and (HE) because leximin on $[0,1]^n$ satisfies Hammond equity and, by Proposition 4, the hypothesis of (HE) is a statement about percentiles.

Conversely, let R be the aggregator of Theorem 3. If $\mathbf{a}, \mathbf{b} \in [0,1]^n$ agree outside $\{i,j\}$ and $b_j < a_j < a_i < b_i$, choose ξ and $\mathbf{x}, \mathbf{y}$ with $U^\xi(\mathbf{x}) = \mathbf{a}$, $U^\xi(\mathbf{y}) = \mathbf{b}$ and $x_k = y_k$ for $k \neq i, j$. By Proposition 4, the hypothesis of (HE) holds, so $\mathbf{a}\,\mathrm{R}\,\mathbf{b}$.

By anonymity it suffices to compare sorted vectors $\mathbf{a}, \mathbf{b}$. If they are equal they are indifferent. Otherwise let m be the first index with $a_m \neq b_m$, say $a_m > b_m$; we show $\mathbf{a}\,\mathrm{P}\,\mathbf{b}$. Let $L = \{k > m : b_k > a_k\}$. If L is empty, $\mathbf{b} \leq \mathbf{a}$ coordinatewise with $b_m < a_m$, and strong Pareto concludes. Otherwise choose $\delta \in (0, a_m - b_m)$ and levels $b_m = t_0 < t_1 < \cdots < t_{|L|} = a_m - \delta$. Starting from $\mathbf{c}^0 = \mathbf{b}$, at step ℓ raise coordinate m from $t_{\ell-1}$ to t_ℓ and lower the ℓth element k of L from b_k to a_k. Since $\mathbf{a}$ is sorted, $a_k \geq a_m > t_\ell$, so $t_{\ell-1} < t_\ell < a_k < b_k$, and the translated (HE) gives $\mathbf{c}^\ell\,\mathrm{R}\,\mathbf{c}^{\ell-1}$. The final vector $\mathbf{c}$ satisfies $\mathbf{c} \leq \mathbf{a}$ and $c_m < a_m$, so strong Pareto gives $\mathbf{a}\,\mathrm{P}\,\mathbf{c}\,\mathrm{R}\,\mathbf{b}$. Hence R is leximin. □

Proof of Theorem 4

The proof needs one construction, which shows that the percentile values of two allocations can be set freely, subject only to individual preferences.

Lemma 5. *Let* $\mathbf{a}, \mathbf{b}, \mathbf{a}', \mathbf{b}' \in (0,1)^n$ *satisfy* $\operatorname{sign}(a_i - b_i) = \operatorname{sign}(a_i' - b_i')$ *for every* i. *There are a preference profile* $\succsim$, *allocations* $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ *and environments* $\xi = (\succsim, \mu)$, $\xi' = (\succsim, \mu')$ *with* $U^\xi(\mathbf{x}) = \mathbf{a}$, $U^\xi(\mathbf{y}) = \mathbf{b}$, $U^{\xi'}(\mathbf{x}) = \mathbf{a}'$ *and* $U^{\xi'}(\mathbf{y}) = \mathbf{b}'$.

Proof. Fix a good ω_1 and let every agent rank bundles by their quantity of ω_1. These preferences are continuous and monotone. Choose numbers $p_i, q_i \in (0, \frac{1}{n})$ ordered as a_i and b_i are, and let x_i and y_i contain p_i and q_i of good ω_1 and nothing else; both allocations are feasible. Fix a proposal law μ_0 under which every good is split by an independent symmetric Dirichlet law, as in Section O8.3 of the Online Appendix;[15] it has full support, and each agent's share of ω_1 has a continuous distribution function H_i. For $\eta \in (0,1)$, let ν draw the agents' shares of ω_1 independently, agent i's from a positive continuous density g_i on $[0, \frac{1}{n}]$, and give nobody any other good; let $\mu = (1-\eta)\nu + \eta\mu_0$. Then μ has full support, each welfare level is atomless, and $U_i^\xi(x_i) = (1-\eta)G_i(p_i) + \eta H_i(p_i)$, where G_i is the distribution function of g_i. For η small enough, the targets $G_i(p_i) = (a_i - \eta H_i(p_i))/(1-\eta)$ and $G_i(q_i) = (b_i - \eta H_i(q_i))/(1-\eta)$ lie in $(0,1)$ and are ordered as p_i and q_i are (equal when $p_i = q_i$), and a positive continuous density g_i meets them. The same construction with $\mathbf{a}'$ and $\mathbf{b}'$ gives μ'. □

Sufficiency. Theorem 3 gives the first five conditions, since the sum is a continuous aggregator. For (MC), let $\mathbf{a} = U^\xi(\mathbf{x})$, $\mathbf{b} = U^\xi(\mathbf{y})$ and $\mathbf{a}', \mathbf{b}'$ the same under ξ'. The function $D(\mathbf{a}, \mathbf{b}) = \sum_i (a_i - b_i)$ is linear and $U^{\xi^\lambda} = \lambda U^\xi + (1-\lambda)U^{\xi'}$, so D under ξ^λ is $\lambda D(\mathbf{a}, \mathbf{b}) + (1-\lambda)D(\mathbf{a}', \mathbf{b}') \geq 0$.

[15] For infinite Ω, take the product over goods, a Radon measure as in the footnote to Assumption 1.

Necessity. Let R be the continuous aggregator of Theorem 3. Call two pairs $(\mathbf{a}, \mathbf{b})$ and $(\mathbf{a}', \mathbf{b}')$ *sign-consistent* if $\operatorname{sign}(a_i - b_i) = \operatorname{sign}(a'_i - b'_i)$ for every i. By Lemma 5, linearity of percentiles in the proposal law, and (MC), for sign-consistent interior pairs,

$$\mathbf{a}\,\mathrm{R}\,\mathbf{b} \text{ and } \mathbf{a}'\,\mathrm{R}\,\mathbf{b}' \implies \lambda\mathbf{a} + (1-\lambda)\mathbf{a}' \;\mathrm{R}\; \lambda\mathbf{b} + (1-\lambda)\mathbf{b}' \qquad (\lambda \in (0,1)). \tag{7}$$

Write $\mathbf{e}_i$ also for the ith unit vector of $\mathbb{R}^n$ and $\mathbf{1}$ for the vector of ones.

Step 0. For $\mathbf{a} \in [0,1]^n$, the sets $\{t : t\mathbf{1}\,\mathrm{R}\,\mathbf{a}\}$ and $\{t : \mathbf{a}\,\mathrm{R}\,t\mathbf{1}\}$ are closed intervals covering $[0,1]$, of the forms $[t^*, 1]$ and $[0, t^*]$ by strong Pareto; they share exactly one point $V(\mathbf{a})$. Then $\mathbf{a}\,\mathrm{R}\,\mathbf{b}$ if and only if $V(\mathbf{a}) \geq V(\mathbf{b})$, and V is continuous, strictly increasing in each coordinate, symmetric, and satisfies $V(t\mathbf{1}) = t$.

Step 1: indifferent pairs mix into indifferent pairs. If two sign-consistent interior pairs are each indifferent, (7) applied to them, and to the reversed pairs, which are again sign-consistent, shows that their mixture is indifferent.

Step 2: linear compensation. Fix $i \neq j$ and $\mathbf{a} \in (0,1)^n$. A loss $s > 0$ for i is *compensable* if some $r \in (0, 1 - a_j)$ satisfies $V(\mathbf{a} - s\mathbf{e}_i + r\mathbf{e}_j) = V(\mathbf{a})$, with $\mathbf{a} - s\mathbf{e}_i$ interior; r is then unique, and we write $r = \rho_{\mathbf{a}}(s)$. Small losses are compensable, by continuity and the intermediate value theorem, and compensable losses form an interval starting at 0. For two compensable losses s, s', the pairs $(\mathbf{a}, \mathbf{a} - s\mathbf{e}_i + \rho_{\mathbf{a}}(s)\mathbf{e}_j)$ and $(\mathbf{a}, \mathbf{a} - s'\mathbf{e}_i + \rho_{\mathbf{a}}(s')\mathbf{e}_j)$ are interior, indifferent and sign-consistent. By Step 1 and uniqueness, $\rho_{\mathbf{a}}(\lambda s + (1-\lambda)s') = \lambda\rho_{\mathbf{a}}(s) + (1-\lambda)\rho_{\mathbf{a}}(s')$: $\rho_{\mathbf{a}}$ is affine. It tends to 0 at 0, since otherwise continuity would give $V(\mathbf{a} + \varepsilon\mathbf{e}_j) \leq V(\mathbf{a})$ for some $\varepsilon > 0$, against strong Pareto. Hence $\rho_{\mathbf{a}}(s) = \kappa_{ij}(\mathbf{a})s$ with $\kappa_{ij}(\mathbf{a}) > 0$.

Step 3: the rate is constant. Let $\mathbf{a}, \mathbf{a}'$ be interior and $\bar{\mathbf{a}} = \lambda\mathbf{a} + (1-\lambda)\mathbf{a}'$. For small $s, s' > 0$, Step 1 applied to the pairs $(\mathbf{a}, \mathbf{a} - s\mathbf{e}_i + \kappa_{ij}(\mathbf{a})s\mathbf{e}_j)$ and $(\mathbf{a}', \mathbf{a}' - s'\mathbf{e}_i + \kappa_{ij}(\mathbf{a}')s'\mathbf{e}_j)$ shows that at $\bar{\mathbf{a}}$ the loss $\lambda s + (1-\lambda)s'$ is compensated by $\lambda\kappa_{ij}(\mathbf{a})s + (1-\lambda)\kappa_{ij}(\mathbf{a}')s'$. By Step 2 at $\bar{\mathbf{a}}$,

$$\kappa_{ij}(\bar{\mathbf{a}})\big(\lambda s + (1-\lambda)s'\big) = \lambda\kappa_{ij}(\mathbf{a})s + (1-\lambda)\kappa_{ij}(\mathbf{a}')s'.$$

Letting $s' \to 0$ and then $s \to 0$ gives $\kappa_{ij}(\mathbf{a}) = \kappa_{ij}(\bar{\mathbf{a}}) = \kappa_{ij}(\mathbf{a}')$. So κ_{ij} is a constant.

Step 4: the rate is one. From $\mathbf{b} = \mathbf{a} - s\mathbf{e}_i + \kappa_{ij}s\mathbf{e}_j$, compensate the loss $\kappa_{ij}s$ for j by a gain $\kappa_{ji}\kappa_{ij}s$ for i: $V(\mathbf{a} + (\kappa_{ji}\kappa_{ij} - 1)s\mathbf{e}_i) = V(\mathbf{a})$, so $\kappa_{ij}\kappa_{ji} = 1$ by strict monotonicity. Symmetry of V under the transposition of i and j turns a compensation of a loss s for i by $\kappa_{ij}s$ for j into a compensation of a loss s for j by $\kappa_{ij}s$ for i, so $\kappa_{ji} = \kappa_{ij}$. Hence $\kappa_{ij} = 1$.

Step 5: V is the mean. For interior $\mathbf{a}$ let $m = \min(a_i, 1 - a_j)$ and $S = \{s \in [0, m) : V(\mathbf{a} - s\mathbf{e}_i + s\mathbf{e}_j) = V(\mathbf{a})\}$. By Steps 2–4, S contains a neighborhood of 0, and the same steps at any point $\mathbf{a} - s\mathbf{e}_i + s\mathbf{e}_j$ with $s \in S$, which is interior, make S open; it is closed in $[0, m)$ by continuity. So $S = [0, m)$, and by continuity V is invariant under every transfer between two coordinates that stays in $[0,1]^n$. Transfers from coordinates above the mean to coordinates below it reach $(\frac{1}{n}\sum_i a_i)\mathbf{1}$ in at most n steps. Hence $V(\mathbf{a}) = \frac{1}{n}\sum_i a_i$, and R is the sum. □

Anonymity enters only in Step 4. Without it, Steps 1–3 still give constant rates with $\kappa_{ik} = \kappa_{ij}\kappa_{jk}$, hence $\kappa_{ij} = w_i/w_j$ for positive weights, and V is a function of a weighted sum of percentiles.

D. PROOF OF THE BALANCE CHARACTERIZATION

The characterization turns on a question that marginal welfare distributions cannot answer: given the scores in a proposal, which agent holds each score? We first describe the probability law that records this information, and express favorable social comparisons in terms of it. If the law is the same for every agent, every anonymous aggregator is balanced. For the converse, we use anonymous threshold rules to reveal differences in who occupies particular score positions. Throughout, we keep apart the distribution of one person's score and the joint law that records who holds each position.

Fix a calibration and draw two proposals independently. Write their score vectors as $\mathbf{a} = \mathbf{m}(\mathbf{x})$ and $\mathbf{b} = \mathbf{m}(\mathbf{y})$; both have joint law ν, and cross-proposal score ties are null. We write $\mathbf{a}\,\mathrm{P}\,\mathbf{b}$ if the

aggregator strictly prefers the first vector, and $\mathbf{a} \gg \mathbf{b}$ if every agent's score is higher in the first vector. The feasible score vectors form $\mathcal{I} = \mathbf{m}(\mathcal{X})$.

An anonymous aggregator sees only the scores, not their owners. For example, the ordered profile $(0.2, 0.8)$ does not reveal whether agent 1 or agent 2 has score 0.2. Let Q be the distribution of the ordered vector $\mathbf{a}_{()}$. Sorting is continuous, so these are Borel random vectors. For each agent i, let λ_i be the joint law of $(a_i, \mathbf{a}_{()})$. Unlike the marginal law of a_i, λ_i records how often agent i holds each score *in each configuration of all agents' scores*. Positional symmetry is exactly $\lambda_1 = \cdots = \lambda_n$.

Favor rates through λ_i. A favorable decision for agent i occurs when the social ranking selects the proposal she prefers and at least one other agent disagrees. First count every strict social comparison that agrees with agent i, then subtract unanimous improvements, which every agent favors. Under strong Pareto, a pair lies in F_i iff $\mathbf{a}\,\mathrm{P}\,\mathbf{b}$, $a_i > b_i$ and $b_j > a_j$ for some j; the last condition fails only when $\mathbf{a} \gg \mathbf{b}$. A pair lies in O_i iff $\mathbf{a}\,\mathrm{P}\,\mathbf{b}$, $a_i < b_i$ and $a_j > b_j$ for some j. Here the last condition is automatic: otherwise strong Pareto would rank $\mathbf{b}$ above $\mathbf{a}$. Hence

$$\pi(F_i) = \mathbb{P}(\mathbf{a}\,\mathrm{P}\,\mathbf{b},\ a_i > b_i) - \mathbb{P}(\mathbf{a} \gg \mathbf{b}), \qquad \pi(O_i) = \mathbb{P}(\mathbf{a}\,\mathrm{P}\,\mathbf{b},\ a_i < b_i). \tag{8}$$

An anonymous aggregator compares only the ordered vectors $\mathbf{a}_{()}$ and $\mathbf{b}_{()}$. Since the proposals are independent, $\lambda_i \otimes \lambda_i$ therefore determines the two agent-specific probabilities in (8). The unanimous-improvement probability is identical for all agents. If social ties are null, then $\pi(F_i) + \pi(O_i) = q_n$ for each i by Lemma O1; consequently balance is equivalent to requiring that $G_i(\mathrm{R}) = \mathbb{P}(\mathbf{a}\,\mathrm{P}\,\mathbf{b},\ a_i > b_i)$ be the same for every agent.

The threshold-priority testing class. We need to test whether two agents hold different positions in otherwise identical score profiles. To do so, consider a rule that first asks whether a proposal meets one welfare threshold, then a second threshold, and finally uses leximin when the threshold tests are tied. The thresholds depend on scores, never on agents' names. A set of scores is an *upper set* if raising one or more scores cannot move a profile out of it; this property makes threshold priority consistent with strong Pareto.

Formally, for closed symmetric upper sets $V, W \subseteq [0,1]^n$, let $\mathrm{R}_{V,W}$ rank $\mathbf{a}$ above $\mathbf{b}$ if $\mathbb{1}_V(\mathbf{a}) > \mathbb{1}_V(\mathbf{b})$; or, if these indicators are equal, if $\mathbb{1}_W(\mathbf{a}) > \mathbb{1}_W(\mathbf{b})$; or, if both are equal, if leximin ranks $\mathbf{a}$ above $\mathbf{b}$. Each $\mathrm{R}_{V,W}$ is an aggregator: anonymity holds because V, W do not distinguish agents, and strong Pareto holds because they are upper sets and leximin breaks the remaining ties. Let $\mathcal{L}$ be this class, including ordinary leximin as $\mathrm{R}_{\varnothing,\varnothing}$. These rules are tools for proving the necessity of positional symmetry, not proposed social objectives.

Proof of Theorem 5

The proof separates an easy implication from a difficult one. If each agent has the same joint distribution of her own score and the ordered score profile, an anonymous aggregator cannot favor one agent more often than another. Conversely, we must show that balance across all the threshold rules forces those joint distributions to coincide. We do this by choosing thresholds that detect whether a named agent is among the lowest scorers in a narrowly specified profile.

We compare four conditions: (a) balance under every anonymous aggregator; (b) balance under $\mathcal{L}$; (c) positional symmetry of the calibrated score law ν; and (d) common percentile calibration together with positional symmetry of the law of the percentile vector. We prove (c)⇒(a)⇒(b)⇒(c), followed by (c)⇔(d). The necessity argument identifies score positions first for gaps below the largest feasible common score, then for higher gaps using condition (T). It must recover the whole joint distribution, not just a few summary statistics.

The joint law ν has full support on $\mathcal{I} = \mathbf{m}(\mathcal{X})$: the inverse image of any nonempty relatively open subset of $\mathcal{I}$ is a nonempty relatively open subset of $\mathcal{X}$. Non-unanimity also has positive probability,

since allocations close to $\mathbf{e}_i(1_\Omega)$ and $\mathbf{e}_j(1_\Omega)$, for $i \neq j$, give a strict disagreement. Consequently, a rule with null social ties has $q_n > 0$.

(c)⇒(a). If $\lambda_1 = \cdots = \lambda_n$, (8) makes $\pi(F_i)$ and $\pi(O_i)$ independent of i, under every aggregator.

(a)⇒(b). Under $\mathrm{R}_{V,W}$, a social tie requires $\mathbf{a}_{()} = \mathbf{b}_{()}$, which has probability zero, so the favor rates are defined.

(b)⇒(c). Two rules. We want a threshold test that reveals which agent prefers one kind of score profile to another. Compare $\mathrm{R}_{V,W}$ with the rule $\mathrm{R}_{W,V}$ that reverses the order of the same two thresholds. These rules disagree only when one proposal belongs to $V \setminus W$ and the other to $W \setminus V$. Both have null social ties, so balance under both rules implies that $G_i(\mathrm{R}_{V,W}) - G_i(\mathrm{R}_{W,V})$ does not depend on i. Independence gives

$$G_i(\mathrm{R}_{V,W}) - G_i(\mathrm{R}_{W,V}) = \nu(V \setminus W)\,\nu(W \setminus V)\,(2p_i - 1),$$

where $p_i = \mathbb{P}(a_i > b_i \mid \mathbf{a} \in V \setminus W,\ \mathbf{b} \in W \setminus V)$. Thus the probability that agent i prefers the first region to the second is identical across agents whenever both regions have positive probability.

Boxes. We now make those two regions small enough to reveal the ownership of particular positions. For a threshold vector $s \in \mathbb{R}^n$, let $H(s) = \{\mathbf{c} \in [0,1]^n : c_{(k)} \geq s_k \text{ for all } k\}$, a closed symmetric upper set; s need not be sorted. A small box in the space of ordered scores restricts each order statistic to a short interval. For any lower corner s and side lengths $h \in (0,1]^n$, write

$$K(s,h) = H(s) \setminus \textstyle\bigcup_k H(s + h_k e_k) = \{\mathbf{c} : c_{(k)} \in [s_k, s_k + h_k) \text{ for all } k\};$$

for a scalar h we write $K(s,h)$ for $K(s,(h,\ldots,h))$. Choose $K_\alpha = K(\alpha,h)$ and $K_\beta = K(\beta,h')$ such that $K_\alpha \cap H(\beta) = K_\beta \cap H(\alpha) = \varnothing$. Let Z be the union of the $2n$ sets $H(\cdot)$ removed in their definitions, and put $V = Z \cup H(\alpha)$ and $W = Z \cup H(\beta)$. Then $V \setminus W = K_\alpha$ and $W \setminus V = K_\beta$. The two-rule comparison therefore yields, whenever $\nu(K_\alpha)\nu(K_\beta) > 0$,

$$\mathbb{P}(a_i > b_i \mid \mathbf{a} \in K_\alpha,\ \mathbf{b} \in K_\beta) \text{ does not depend on } i. \tag{9}$$

The left side reports whether a named agent tends to be above the other proposal's score when the *ordered* profiles lie in the specified boxes.

Target. What must the tests recover? Suppose a score profile has a gap between its mth and $(m+1)$st lowest scores. The first m places belong to some m agents. Conditional on the ordered profile, positional symmetry requires every named agent to be among those m agents with probability m/n. Define $L_{i,m}$ as the event $a_i \leq a_{(m)} < a_{(m+1)}$: agent i is among the m lowest scorers, with a gap above them. Conditional on $\mathbf{a}_{()}$, the probabilities of all such events determine the law of a_i, and their sum over agents is m.

For each m, define the finite measure $\kappa_{i,m}(E) = \mathbb{P}(\mathbf{a}_{()} \in E,\ L_{i,m})$ on the space of ordered profiles. These measures determine λ_i even when scores within a proposal are tied. Indeed, for every Borel set E of ordered profiles and every $t \in [0,1]$,

$$\lambda_i([0,t] \times E) = Q(E \cap \{s_n \leq t\}) + \sum_{m=1}^{n-1} \kappa_{i,m}(E \cap \{s_m \leq t < s_{m+1}\}).$$

If t separates two distinct scores, exactly one summand applies; if all scores are at most t, the first term applies. Thus equality of the $\kappa_{i,m}$ across agents implies equality of the λ_i. This argument does not assume that scores within one proposal are distinct.

Fix m and abbreviate $\kappa_{i,m}$ to κ_i. It suffices to identify these measures locally on $\{s : s_m < s_{m+1}\}$, outside which they vanish. We use the following measure-determination argument. Intersect half-open dyadic coordinate cubes with the ordered region. Whenever every sufficiently small such cube in a neighborhood has the same κ_i-mass for every i, finite-measure uniqueness gives equality on every

Borel subset of that neighborhood: the cubes and their refinements generate its relative Borel sets. Each cube is of the form $K(s,h)$ above, interpreted in ordered coordinates; no sorting of its lower corner is required. A countable cover by these neighborhoods then gives equality on their union. Finally, $\kappa_i \ll Q$, so a Q-null exceptional set can be discarded. The comparisons below establish exactly these local equalities.

The easiest test uses a profile in which all agents have almost the same score. Let $c^* = \max\{c : (c,\dots,c) \in \mathcal{I}\}$ be the largest feasible common score. By Lemma O4, $\mathcal{I}$ is downward comprehensive. Each score is positive almost surely, since its distribution is atomless; hence an attainable vector has all coordinates positive, and $c^* > 0$. Compactness of $\mathcal{I}$ attains the maximum defining c^*. Every vector satisfies $a_{(1)} \le c^*$ by downward comprehensiveness, with equality only if some $a_i = c^*$, a null event. Thus $a_{(1)} < c^*$ almost surely, and every diagonal vector $(t,\dots,t)$ with $0 < t < c^*$ is feasible. These profiles supply comparison boxes that separate lower from higher scores.

Step 1: gaps below c^*. Consider a narrow box K of ordered profiles with a gap between positions m and $m+1$. If that gap straddles some $t < c^*$, compare K with a box near $(t,\dots,t)$. The agent prefers the almost-equal proposal exactly when she holds one of the m lower scores in the other proposal. Thus (9) identifies the probability of the event $L_{i,m}$.

Formally, let $K = K(\beta, h')$ and choose $t < c^*$ such that $\beta_m + h'_m \le t < \beta_{m+1}$. Take $h > 0$ with $t + h < \min\{\beta_{m+1}, c^*\}$ and put $K_\alpha = [t, t+h)^n$. Points of K_α have their $(m+1)$st order statistic below β_{m+1}, so they do not belong to $H(\beta)$. A point of K has its mth lowest coordinate below t, so its minimum is below t and it does not belong to $H(t,\dots,t)$. The open cube $(t, t+h)^n$ contains a feasible diagonal point; hence $\nu(K_\alpha) > 0$ by full support. For $\mathbf{a} \in K_\alpha$ and $\mathbf{b} \in K$, $a_i > b_i$ iff $b_i < t$, that is, iff $\mathbf{b} \in L_{i,m}$. By (9), $\kappa_i(K)$ does not depend on i; if $\nu(K) = 0$, this is immediate. Since exactly m agents satisfy $L_{i,m}$ on K, the common mass is $(m/n)\nu(K)$. Near each point with $s_m < \min(s_{m+1}, c^*)$, choose one t in that strict gap and then one sufficiently small h. The same reference box separates every sufficiently small target box in a neighborhood of the point. The local measure-determination argument therefore applies. The argument applies to every gap index m. For $n = 2$, the only gap is $m = 1$, and $a_{(1)} < c^*$ almost surely, so nothing further remains to identify.

Step 2: gaps above c^* ($n \ge 3$). A diagonal comparison cannot distinguish the lower and higher groups when their gap lies entirely above c^*. Condition (T) supplies a different comparison profile: it gives a feasible vector strictly above the target vector in every coordinate. We use that room to construct a proposal with $n-1$ low scores and one high score. The low scores reveal ownership of the very lowest positions, already identified in Step 1; the remaining part of the comparison then reveals ownership of the higher gap.

Let $\mathbf{b}$ satisfy (T) with $\mathbf{c} \gg \mathbf{b}$ in $\mathcal{I}$, and put $\beta = \mathbf{b}_{()}$ with $\beta_1 < c^* \le \beta_m < \beta_{m+1}$ (recall that $\beta_1 < c^*$ almost surely). Let r be the number of coordinates of β equal to β_1; then $r < m \le n-1$. Set $\varepsilon = \min_k(c_k - b_k) > 0$. Choose t with $\beta_1 < t < \min(\beta_{r+1}, c^*, \beta_1 + \varepsilon)$, choose u with $\beta_m < u < \beta_{m+1}$. Choose $h > 0$ so that

$$t + h < \min\{\beta_{r+1}, c^*, u, \beta_1 + \varepsilon\}, \qquad u + h < \beta_{m+1}.$$

Let $\alpha = (t,\dots,t,u)$, with $n-1$ copies of t, and $K_\alpha = K(\alpha, h)$. Choose a neighborhood of β so that every ordered profile in it has its first r coordinates below t and its remaining coordinates above $t+h$, and its first m coordinates below u and its remaining coordinates above $u+h$. For every sufficiently small target box $K = K(\beta', h')$ in this neighborhood, $K_\alpha \cap H(\beta') = \varnothing$: the $(r+1)$st coordinate of a profile in K_α is below β'_{r+1}. Also $K \cap H(\alpha) = \varnothing$, since profiles in K have minimum below t. Give the value $u + h/2$ to an agent holding the largest coordinate of $\mathbf{b}$ and $t + h/2$ to every other agent. Each low coordinate is below $\beta_1 + \varepsilon \le c_k$; the high coordinate is below $\beta_{m+1} \le \beta_n < c_k$ for its recipient. The constructed vector therefore lies strictly below $\mathbf{c}$ and in $(0,1)^n$. Comprehensiveness puts an open box around it in $\mathcal{I}$, and the box can be chosen inside K_α. Thus $\nu(K_\alpha) > 0$. This construction uses the actual owners of $\mathbf{b}$'s scores; it does not assume that the calibrated feasible set is invariant under

permutations.

For $\mathbf{a} \in K_\alpha$ and $\mathbf{b}' \in K$, the event $a_i > b'_i$ occurs either when a_i is one of the $n-1$ low coordinates and agent i is among the r lowest of $\mathbf{b}'$, or when a_i is the top coordinate and agent i is among the m lowest of $\mathbf{b}'$. Step 1, applied to K at gap r (threshold t) and to K_α at gap $n-1$ (a threshold between $t+h$ and $\min(u, c^*)$), shows that agent i belongs to the first group with probability r/n and holds the top coordinate of $\mathbf{a}$ with probability $1/n$. Independence therefore gives

$$\mathbb{P}(a_i > b'_i \mid \mathbf{a} \in K_\alpha,\ \mathbf{b}' \in K) = \frac{n-1}{n} \cdot \frac{r}{n} + \frac{1}{n} \frac{\kappa_i(K)}{\nu(K)},$$

and (9) shows that $\kappa_i(K)$ does not depend on i. Zero-mass target boxes again require no comparison. The same reference box works for all sufficiently small boxes in the chosen neighborhood, so the local measure-determination argument gives equality there.

To justify coverage, take the union of the neighborhoods just constructed over all feasible $\mathbf{b}$ satisfying (T), $\beta_1 < c^*$, and the specified gap. This is open in the ordered-profile space. Its inverse image under sorting contains every such $\mathbf{b}$, so (T) makes the uncovered part of the higher-gap region Q-null. A countable subcover exists because the ordered-profile space is a subspace of $\mathbb{R}^n$. Together with Step 1, this proves equality of the $\kappa_{i,m}$ for every m, hence equality of the λ_i by the displayed recovery formula.

(c)⇔(d). Once all the owner-and-profile laws agree, the remaining restriction on calibration follows from marginal distributions. Under positional symmetry, the conditional distribution of an agent's score is the same for every agent. Summing these distributions over agents counts each entry of the ordered profile once, including repeated entries. Thus the law of a_i is $\int \frac{1}{n} \sum_k \delta_{s_k}\, dQ(s)$, independently of i. Writing $m_i = \psi_i \circ U_i$, where each $U_i(x_i)$ is uniform, gives the distribution function ψ_i^{-1} for $m_i(x_i)$. Equal marginals force $\psi_1 = \cdots = \psi_n = \psi$. Applying the same increasing transformation ψ to all coordinates preserves their order, so ν is positionally symmetric exactly when the law of the percentile vector $(U_1(X_1), \ldots, U_n(X_n))$ is. The reverse argument proves the converse. Finally, if $m_i = \psi \circ U_i$ for all i, then $m_i(x) \geq m_j(y)$ iff $\psi(U_i(x)) \geq \psi(U_j(y))$ iff $U_i(x) \geq U_j(y)$, which is the last statement of the theorem. □

Proof of Theorem 6

Replace ν by $\bar{\nu}$ and π by $\bar{\pi}$ in the argument above: the score vectors $\mathbf{a}, \mathbf{b}$ are now independent draws from $\bar{\nu}$, agent i prefers $\mathbf{a}$ when $a_i > b_i$, and $\bar{F}_i, \bar{O}_i$ replace F_i, O_i. The identities (8) hold verbatim, since their derivation uses only strong Pareto and the definitions of the two events. Each marginal ν_i is atomless with full support on $[0,1]$ (Lemma O2) and the coordinates are independent, so any two coordinates of $(\mathbf{a}, \mathbf{b})$ coincide with probability zero, and every nonempty open box of $[0,1]^n$ has positive $\bar{\nu}$-mass. A conflict occurs unless one vector dominates the other coordinatewise, so its probability is $1 - 2^{1-n} > 0$; for a rule with null social ties, exchanging the two draws gives $\bar{\pi}(\bar{F}_i) + \bar{\pi}(\bar{O}_i) = \frac{1}{2}(1 - 2^{1-n})$ for every i, and expected balance under such a rule is equivalent to $\bar{\pi}(\mathbf{a} \,\mathrm{P}\, \mathbf{b},\ a_i > b_i)$ not depending on i.

(c)⇒(a). If the ν_i coincide, $\bar{\nu}$ is a product of identical laws, hence exchangeable and positionally symmetric, and (8) makes $\bar{\pi}(\bar{F}_i)$ and $\bar{\pi}(\bar{O}_i)$ independent of i under every aggregator.

(a)⇒(b). Under $\mathrm{R}_{V,W}$ a social tie requires $\mathbf{a}_{()} = \mathbf{b}_{()}$, a null event, so the expected favor rates are defined.

(b)⇒(c). The two-rule comparison and (9) hold under $\bar{\pi}$ with the same proof. The support of $\bar{\nu}$ is $[0,1]^n$, which is compact and downward comprehensive, so it plays the role of $\mathcal{I}$ and $c^* = 1$: every diagonal vector $(t, \ldots, t)$ with $0 < t < 1$ lies in the support, and $a_{(1)} < 1$ almost surely. Step 1 therefore applies to every gap index m, and Step 2 and condition (T) are not needed. Hence $\bar{\nu}$ is positionally symmetric. As in the proof of (c)⇔(d) above, the law of a_i is then $\int \frac{1}{n} \sum_k \delta_{s_k}\, dQ(s)$ for every i, so the ν_i coincide. The reformulation with nondecreasing φ is immediate: equal laws give equal expectations, and $\varphi = \mathbb{1}_{(t,1]}$ recovers (2).

(c)⇔(d). Section 4.3 shows that (2) holds iff $\psi_1 = \cdots = \psi_n$. If $m_i = \psi \circ U_i$ for all i, then $m_i(x) \geq m_j(y)$ iff $U_i(x) \geq U_j(y)$. Conversely, suppose these comparisons agree for all i, j, x, y. Given $u \in [0,1]$, choose bundles x, y with $U_i(x) = U_j(y) = u$, which exist because percentiles are onto; then $m_i(x) \geq m_j(y)$ and $m_j(y) \geq m_i(x)$, so $\psi_i(u) = \psi_j(u)$. □

E. LEXIMIN IN A TWO-AGENT CONFLICT

Under two-agent leximin, a conflict reduces to one simple comparison. Suppose agent 1 prefers $\mathbf{x}$ and agent 2 prefers $\mathbf{y}$. Leximin chooses $\mathbf{x}$ exactly when agent 1's score *in* $\mathbf{y}$, the proposal she likes less, is below agent 2's score *in* $\mathbf{x}$, the proposal he likes less. To put both scores on one scale, convert agent 1's score with the inverse of agent 2's increasing transformation. The resulting function $g = \psi_2^{-1} \circ \psi_1$ contains all the relative-scale information that leximin uses.

Proposition 6. *Let $n = 2$ and* R *be leximin, and write a calibration as $m_i = \psi_i \circ U_i$. Up to a π-null set, on a conflict with $x_1 \succ_1 y_1$ and $y_2 \succ_2 x_2$,*

$$\mathbf{x} \succ \mathbf{y} \iff m_2(x_2) > m_1(y_1) \iff g\big(U_1(y_1)\big) < U_2(x_2), \qquad g \equiv \psi_2^{-1} \circ \psi_1,$$

and, on a conflict with $y_1 \succ_1 x_1$ and $x_2 \succ_2 y_2$, $\mathbf{x} \succ \mathbf{y} \iff m_1(x_1) > m_2(y_2)$. The social preference depends on the calibration only through the increasing homeomorphism g, the scale conversion *between the agents' percentiles, and agent* 1 *wins a conflict exactly when her score for her less-preferred proposal is below agent* 2*'s corresponding score.*

The preceding comparison reduces the two agents' calibrations to one increasing homeomorphism $g = \psi_2^{-1} \circ \psi_1$. A common rescaling does not change g, because the transformation cancels between ψ_2^{-1} and ψ_1. For conflicts in which agent 1 prefers $\mathbf{x}$ and agent 2 prefers $\mathbf{y}$, the favor rate is therefore

$$f_1(\mathbf{m}, \mathrm{R}) = \pi\big\{g(U_1(y_1)) < U_2(x_2) \,\big|\, x_1 \succ_1 y_1,\ y_2 \succ_2 x_2\big\}.$$

It depends on the calibration only through g, and counts the conflicts in which leximin chooses agent 1's preferred proposal. Lowering g pointwise lowers agent 1's score for her *less-preferred* proposal. Leximin then protects her more often, and f_1 weakly increases; we do not claim strict monotonicity. This one-dimensional comparison makes the power calibration of Theorem O2 in the Online Appendix constructive.

Proof of Proposition 6. Leximin compares the lower score under each proposal. In a conflict, one of those lower scores belongs to the agent who likes that proposal less. This observation reduces the choice to one inequality between the two agents. Consider the first orientation and put $a = m_1(x_1)$, $b = m_2(x_2)$, $c = m_1(y_1)$, $d = m_2(y_2)$. The conflict gives $a > c$ and $d > b$, and leximin compares $\min(a, b)$ with $\min(c, d)$; a tie has probability zero. If $b > c$, then $\min(a, b) > c \geq \min(c, d)$, so $\mathbf{x} \succ \mathbf{y}$. If $b < c$, then $\min(a, b) = b < \min(c, d)$, so $\mathbf{y} \succ \mathbf{x}$. Hence $\mathbf{x} \succ \mathbf{y} \iff m_2(x_2) > m_1(y_1)$. Substituting $m_i = \psi_i \circ U_i$ and applying ψ_2^{-1} gives the percentile form. The second orientation is the first with the labels of the two allocations exchanged. In the first orientation agent 1 prefers $\mathbf{x}$, so she is favored exactly when $m_1(y_1) < m_2(x_2)$. The compared scores belong to the agents' less-preferred proposals, and she wins when hers is lower. Only g enters, and a common relabeling $\psi_i \mapsto \varphi \circ \psi_i$ leaves g unchanged. □

REFERENCES

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII, 1983*, volume 1117 of *Lecture Notes in Mathematics*, pages 1–198. Springer. doi:10.1007/BFb0099421.

Arrow, K. J. (1973). Some ordinalist-utilitarian notes on Rawls's theory of justice. *Journal of Philosophy*, 70(9):245–263. doi:10.2307/2025006.

Babichenko, Y., Feldman, M., Holzman, R., and Narayan, V. V. (2024). Fair division via quantile shares. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC 2024)*, pages 1235–1246. doi:10.1145/3618260.3649728.

Blackorby, C. and Donaldson, D. (1988). Money metric utility: A harmless normalization? *Journal of Economic Theory*, 46(1):120–129.

Chakraborty, S., Qiu, J., Yuan, H., Koppel, A., Manocha, D., Huang, F., Bedi, A. S., and Wang, M. (2024). MaxMin-RLHF: Alignment with diverse human preferences. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 6116–6135. PMLR.

Conitzer, V., Freedman, R., Heitzig, J., Holliday, W. H., Jacobs, B. M., Lambert, N., Mossé, M., Pacuit, E., Russell, S., Schoelkopf, H., Tewolde, E., and Zwicker, W. S. (2024). Position: Social choice should guide AI alignment in dealing with diverse human feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 9346–9360. PMLR.

d'Aspremont, C. and Gevers, L. (1977). Equity and the informational basis of collective choice. *Review of Economic Studies*, 44(2):199–209. doi:10.2307/2297061.

Decancq, K., Fleurbaey, M., and Schokkaert, E. (2015). Happiness, equivalent incomes and respect for individual preferences. *Economica*, 82(s1):1082–1106.

Deschamps, R. and Gevers, L. (1978). Leximin and utilitarian rules: A joint characterization. *Journal of Economic Theory*, 17(2):143–163.

Dhillon, A. and Mertens, J.-F. (1999). Relative utilitarianism. *Econometrica*, 67(3):471–498. doi:10.1111/1468-0262.00033.

Dubey, P. and Shapley, L. S. (1979). Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research*, 4(2):99–131. doi:10.1287/moor.4.2.99.

Fleurbaey, M. and Blanchet, D. (2013). *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford University Press, Oxford.

Fleurbaey, M. and Maniquet, F. (2011). *A Theory of Fairness and Social Welfare*. Cambridge University Press, Cambridge. doi:10.1017/CBO9780511851971.

Fleurbaey, M. and Zuber, S. (2021). Fair utilitarianism. *American Economic Journal: Microeconomics*, 13(2):370–401. doi:10.1257/mic.20170234.

Ge, L., Halpern, D., Micha, E., Procaccia, A. D., Shapira, I., Vorobeychik, Y., and Wu, J. (2024). Axioms for AI alignment from human feedback. In *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, pages 80439–80465. doi:10.52202/079017-2557.

Gehrlein, W. V. and Lepelley, D. (2011). *Voting Paradoxes and Group Coherence: The Condorcet Efficiency of Voting Rules*. Studies in Choice and Welfare. Springer, Berlin. doi:10.1007/978-3-642-03107-6.

Gibbard, A. (1979). Disparate goods and Rawls' difference principle: A social choice theoretic treatment. *Theory and Decision*, 11(3):267–288. doi:10.1007/BF00126381.

Guinier, L. (1994). *The Tyranny of the Majority: Fundamental Fairness in Representative Democracy*. Free Press, New York.

Hammond, P. J. (1976). Equity, Arrow's conditions, and Rawls' difference principle. *Econometrica*, 44(4):793–804.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4):309–321. doi:10.1086/257678.

HM Treasury (2026). The green book (2026). UK Government guidance on appraisal. Official online guidance, updated 28 October 2026.

Kalai, G. (2002). A Fourier-theoretic perspective on the Condorcet paradox and Arrow's theorem. *Advances in Applied Mathematics*, 29(3):412–426. doi:10.1016/S0196-8858(02)00023-4.

Karni, E. and Weymark, J. A. (2024). Impartiality and relative utilitarianism. *Social Choice and Welfare*, 63(1):1–18. doi:10.1007/s00355-024-01518-1.

Laruelle, A. and Valenciano, F. (2005). Assessing success and decisiveness in voting situations. *Social Choice and Welfare*, 24(1):171–197. doi:10.1007/s00355-003-0298-7.

MacAskill, W. (2016). Normative uncertainty as a voting problem. *Mind*, 125(500):967–1004. doi:10.1093/mind/fzv169.

MacAskill, W., Cotton-Barratt, O., and Ord, T. (2020). Statistical normalization methods in interpersonal and intertheoretic comparisons. *Journal of Philosophy*, 117(2):61–95. doi:10.5840/jphil202011725.

Mossel, E. (2012). A quantitative Arrow theorem. *Probability Theory and Related Fields*, 154(1–2):49–88. doi:10.1007/s00440-011-0362-7.

Pazner, E. A. and Schmeidler, D. (1978). Egalitarian equivalent allocations: A new concept of economic equity. *Quarterly Journal of Economics*, 92(4):671–687. doi:10.2307/1883182.

Piacquadio, P. G. (2017). A fairness justification of utilitarianism. *Econometrica*, 85(4):1261–1276. doi:10.3982/ECTA14151.

Rae, D. W. (1969). Decision-rules and individual values in constitutional choice. *American Political Science Review*, 63(1):40–56. doi:10.2307/1954283.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press, Cambridge, MA. doi:10.2307/j.ctvjf9z6v.

Roberts, K. W. S. (1980). Interpersonal comparability and social choice theory. *Review of Economic Studies*, 47(2):421–439. doi:10.2307/2297002.

Samuelson, P. A. (1974). Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. *Journal of Economic Literature*, 12(4):1255–1289.

Sen, A. K. (1970). *Collective Choice and Social Welfare*. Holden-Day, San Francisco. Online edition.

Sen, A. K. (1977). On weights and measures: Informational constraints in social welfare analysis. *Econometrica*, 45(7):1539–1572. doi:10.2307/1913949.

Siththaranjan, A., Laidlaw, C., and Hadfield-Menell, D. (2024). Distributional preference learning: Understanding and accounting for hidden context in RLHF. In *The Twelfth International Conference on Learning Representations (ICLR 2024)*. arXiv:2312.08358.

Sprumont, Y. (2013). On relative egalitarianism. *Social Choice and Welfare*, 40(4):1015–1032. doi:10.1007/s00355-012-0653-7.

Sunstein, C. R. (2024). The economic constitution of the united states. *Journal of Economic Perspectives*, 38(2):25–42. doi:10.1257/jep.38.2.25.

Young, H. P. (1974). An axiomatization of Borda's rule. *Journal of Economic Theory*, 9(1):43–52.