Multidimensional welfare rankings*

Stergios Athanassoglou†

November 2013

Abstract

Social well-being is intrinsically multidimensional. Welfare indices attempting to reduce this complexity to a unique measure abound in many areas of economics and public policy. Ranking alternatives based on such measures depends, sometimes critically, on how the different dimensions of welfare are weighted. In this paper, a theoretical framework is presented that yields a set of consensus rankings in the presence of such weight imprecision. The main idea is to consider a vector of weights as an imaginary voter submitting preferences over alternatives in the form of an ordered list. With this voting construct in mind, a rule for aggregating the preferences of many plausible choices of weights, suitably weighted by the importance attached to them, is proposed. An axiomatic characterization of the rule is provided, and its computational implementation is developed. An analytic solution is derived for an interesting special case of the model corresponding to generalized weighted means and the ϵ-contamination framework of Bayesian statistics. The model is applied to the Academic Ranking of World Universities index of Shanghai University, a popular composite index measuring academic excellence.

Keywords: multidimensional welfare, social choice, voting, Kemeny's rule, graph theory, ϵ-contamination

JEL classifications: D71, D72, I31, C61

*I am grateful to Jim Lawrence for helpful and clarifying remarks and to Michaela Saisana for introducing me to this interesting area of research. I also would like to thank Andrea Saltelli for his useful comments. The views expressed herein are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

†European Commission Joint Research Center, Econometrics and Applied Statistics Unit, athanassoglou@gmail.com.

1 Introduction

Many aspects of social well-being are intrinsically multidimensional. Development, poverty, inequality, education: these are all concepts that depend on a number of different criteria and cannot be captured by simple quantitative measures ([35, 36, 38, 3, 5, 10, 2], among many others). Despite the inherent complexity of these settings, there is still a need to compare and eventually order possible alternatives on the basis of the multidimensional information they involve. Welfare indices attempt to accomplish this task by integrating the various dimensions of well-being into a single, one-dimensional measure. This is generally achieved by assigning weights to the different dimensions and, in some fashion, aggregating over them.

It should be intuitively clear that the chosen weights can have a major effect on composite welfare scores and therefore on the final ranking of the various alternatives. Their choice is laden with complex philosophical and operational dilemmas (see section A.7 in Foster and Sen [17], Ravallion [31, 32], and Decanq and Lugo [12]), which means that they need to be assigned in a systematic, transparent, and judicious fashion. From a practical standpoint, many different methods for doing so have been proposed, including principal component and factor analyses, data envelopment, public opinion polls, budget allocation, analytic hierarchy processes, and expert consultation, among others. The interested reader is referred to [12, 29] for a comprehensive survey.

Despite the wealth of available techniques to determine welfare index weights, their determination remains controversial. Deep disagreements regarding the United Nations' Human Development Index (HDI), a very popular measure of national wellbeing, are emblematic of this point [31, 30]. Indeed, there is frequently no one "right" way to set them and we are often justified, if not compelled to, consider the effect of many different weights at once. Such an analysis would serve two goals: (a) to examine how robust a given ranking of alternatives is to changes in weights, and (b) to determine a compromise ranking that is in some sense "optimal" in the presence of weight imprecision.

Computational work in gauging the robustness of composite measures of welfare with respect to the choice of weights primarily focused on Monte Carlo simulation (Saisana et al. [33], OECD and JRC [29]). These approaches assessed the importance of weights in the context of a broader uncertainty and sensitivity analysis of given indices. Of more relevance to the present work, Duclos et al. [13] studied the robustness of multidimensional poverty comparisons in a nonparametric setting. They established an analytic criterion for determining whether a (pairwise) poverty comparison is robust to the choice of aggregation procedure and poverty line within a wide class of poverty indices. Moreover, they provided statistical tests that apply their insights to empirical settings. In follow-up work, Duclos et al. [14] developed a similar framework for multidimensional inequality comparisons. Pinar et al. [30] focused on the HDI index and used ideas from stochastic dominance to determine the set of weights that results in best-case performance for a particular country. They used these weights to test the robustness of the HDI index over time. Anderson et al. [4] adopted a nonparametric approach, by imposing monotonicity and quasiconcavity on the welfare function and deriving upper and lower bounds on welfare levels for the various alternatives/agents. They, too, applied their framework to the HDI. Alternatively, Foster et al. [16] studied linear composite indices of welfare and adopted a parametric structure for weight imprecision based on the ϵ-contamination model of Bayesian analysis (Berger and Berliner [8]). In their setting, a pairwise comparison between alternatives was defined to be robust with respect to a given level of weight imprecision if *all* of the weight vectors corresponding to it (i.e., this level of weight imprecision) produce composite scores that maintain the same relative ranking. In their analysis, the key quantity of interest became the maximum level of imprecision at which the pairwise comparison remains robust.

Contribution. I take a fundamentally different approach. An important feature of the analytical framework I propose is that, unlike that of previous contributions [13, 14, 4, 16, 30], it can efficiently produce complete rankings in high-dimensional settings of multiple alternatives and indicators (e.g., see Section 6). In my model, which I describe in Section 2, welfare is modeled via a general function u, whose arguments are normalized achievement vectors (corresponding to performance across the different dimensions of welfare) and weight vectors, reflecting the weight attached to each dimension of welfare. Each vector of weights is itself weighted via a nontrivial and Lebesgue integrable *importance* function f. This function captures beliefs and/or tastes regarding the correct set of weights to use.[1] Subsequently, in Section 3, I use the theory of social choice to propose a consensus ranking of the alternatives given this weight imprecision. Viewing a vector of weights as an abstract voter who expresses his/her preferences over alternatives via the welfare function u, an electorate is constructed by considering each weight vector $\boldsymbol{w}$

[1] That both u and f are model primitives is reminiscent of certain settings of decision-making under ambiguity, where the existence of a utility function and a second-order probability distribution over a set of Bayesian priors are assumed (Klibanoff et al. [21]). Still, it should be mentioned that these assumptions are not crucial and that the model is general enough to accommodate fully nonparametric analytical frameworks (see Section 2).

in the simplex and introducing $f(\boldsymbol{w})$ (possibly irrational) copies of itself having identical preferences. This operation results in a continuum of voters $\mathcal{E}(f)$ of finite Lebesgue measure. Next, I propose Kemeny's rule (Kemeny [20]) as a way of resolving voter disagreement and computing a consensus ranking, given the preferences of this electorate of weight vectors $\mathcal{E}(f)$. Well-established in the social choice and voting literatures [25, 42], Kemeny's rule produces a ranking (referred to as "Kemeny-optimal") that minimizes the sum of pairwise rank disagreements with respect to stated voter preferences. Drawing from the results of Young [40] and Young and Levenglick [43], an axiomatic characterization of the rule is provided, on the basis of five intuitive properties. Yet, it should be noted that the appeal of the proposed rule may extend beyond axiomatic considerations since, under certain statistical frameworks, Kemeny's rule can be interpreted as providing the maximum likelihood estimate of the objectively "correct" ranking [41, 11].

In Section 4, I explore the computational implementation of Kemeny's rule to the present context. This is because, despite its many virtues, Kemeny's rule suffers from one very serious drawback: the computation of a Kemeny-optimal ranking is NP-hard [7], even when the number of criteria is just four [15]. The main problem arises from the potential presence of so-called *Condorcet cycles*, implying intransitive majority (pairwise) preferences. Yet, viewing the problem under a graph theoretic lens, it may be possible to reduce it to a series of smaller and more tractable problems. The main insight is to use a seminal algorithm of Tarjan [37] to identify the strongly connected components of a suitably-defined directed graph, which in turn contain maximal-size Condorcet cycles of the underlying social choice problem. If the length of these cycles is small enough, brute-force enumeration can resolve them efficiently. If not, it is necessary to use approximation algorithms and I implement the ones that provide the best approximation guarantees available, due to Van Zuylen and Williamson [44].

In Section 5, I analytically work out an interesting special case of the model corresponding to generalized weighted means, a family of welfare functions that are very popular in practice [29, 16, 39], and the ϵ-contamination setting of Foster et al. [16]. Using results from polyhedral geometry (Lawrence [24]) it is possible to present a closed-form formula for the proportion of weights ranking one alternative over another. This formula facilitates computations significantly, as Kemeny's rule can be applied without the need to compute these proportions numerically. Moreover, it is possible to prove that these proportions are monotonic in ϵ, the magnitude of weight imprecision.

In Section 6, I apply the special case of the model presented in Section 5 to Shanghai

Jiao Tong University's Academic Ranking of World Universities (ARWU).[2] The ARWU index is a widely-used composite index measuring academic excellence. It linearly aggregates (normalized) academic performance across 6 dimensions to produce university rankings. These rankings receive a lot of media attention; in some cases they may even drive political discourse on national university systems [34]. Nonetheless, it has been shown that they can be very sensitive to seemingly ad-hoc conventions, including the choice of weights [34]. The controversial nature of this index, in combination with its relatively high dimensionality (100 countries, 6 indicators), render it a good application area for the model. The analysis does indeed show how introducing even moderate amounts of weight imprecision may have a very significant effect on a university's ARWU rank.

Connections to Knightian uncertainty. On a final note, it is worth mentioning the possible connections of this work to the decision-theoretic literature on Knightian uncertainty. The examination of welfare indices under weight imprecision is, mathematically if not conceptually, analogous to certain settings of decision-making under ambiguity (see Gilboa and Marinacci [18]). Uncertain weight vectors over dimensions of well-being can be recast as uncertain Bayesian priors over (a finite collection of) states of nature. Correspondingly, the analogy of Bayesian priors as voters, expressing their preferences over a finite set of actions can once again be evoked and ideas from voting theory can be used in this context, too. As a result, the work presented here may be of relevance to this vibrant field of economic theory.

Paper outline. The structure of the paper is as follows. Section 2 introduces the formal model. Section 3 applies Kemeny's rule to our context and discusses the normative implications of this choice. Section 4 focuses on computational issues and shows how the determination of Kemeny-optimal rankings can be efficiently tackled. Section 5 derives an analytic solution for a compelling special case of the model corresponding to generalized weighted means and ϵ-contamination. Section 6 implements the proposed procedure to the 2013 ARWU Index, while Section 7 provides conclusions and directions for future research. All mathematical proofs and figures are collected in the Appendix.

[2] www.shanghairanking.com

2 Model Description

Consider a set of *alternatives* $\mathcal{A}$ indexed by $a = 1, 2, ..., A$ and a set of *indicators* $\mathcal{I}$ indexed by $i = 1, 2, ..., I$. Let $x_{ai} \in [0, 1]$ denote alternative a's normalized value of indicator i, and $\boldsymbol{x_a} \in [0, 1]^I$ its *achievement* vector collecting all such information[3] (all vectors are taken to be column vectors). $\boldsymbol{X}_{\mathcal{A}} \subset [0, 1]^{I \times A}$ denotes an achievement matrix collecting the achievement vectors of alternatives in $\mathcal{A}$. That is, the (i, j)'th entry of $\boldsymbol{X}_{\mathcal{A}}$ is $x_{a_j i}$.

Assume that performance across indicators is weighted by a vector $\boldsymbol{w}$ belonging in $\Delta^{I-1} = \{w \in \Re^I : \quad \boldsymbol{w} \geq \boldsymbol{0}, \quad \sum_{i=1}^{I} w_i = 1\}$, the $(I-1)$-dimensional simplex. Here, w_i represents the weight given to indicator i. The *welfare* corresponding to achievement vector $\boldsymbol{x}$ and $\boldsymbol{w}$ is given by a general function

$$u(\boldsymbol{x}, \boldsymbol{w}) : \Re^I \times \Delta^{I-1} \mapsto \Re. \tag{1}$$

The welfare function is purposefully left general in order to accommodate many different multidimensional phenomena from poverty [13, 2], to inequality [17, 38], to development [3].

Now, define an *importance function* f on the simplex Δ^{I-1}, satisfying $f(\boldsymbol{w}) \geq 0$ for all $\boldsymbol{w} \in \Delta^{I-1}$ and $0 < \int_{\Delta^{I-1}} f(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} < +\infty$. For instance, if $f(\boldsymbol{w}) = 1$ if and only if $\boldsymbol{w} \in W$ for some $W \subset \Delta^{I-1}$ having non-negligible Lebesgue measure in $\Re^{I-1}$ and is zero everywhere else, then this implies that all $\boldsymbol{w} \in W$ are assigned equal importance, while all $\boldsymbol{w} \notin W$ are not considered at all. Similarly to u, the importance function f is a model input and its determination will be case-specific.[4]

Given a welfare function u and achievement matrix $\boldsymbol{X}_{\mathcal{A}}$, define for all $(a_i, a_j) \in \mathcal{A} \times \mathcal{A}$ the following subset of Δ^{I-1}:

$$W_{a_i a_j}^{\boldsymbol{X}_{\mathcal{A}}, u} = \left\{ \boldsymbol{w} \in \Delta^{I-1} : \; u(\boldsymbol{x}_{\boldsymbol{a_i}}, \boldsymbol{w}) \geq u(\boldsymbol{x}_{\boldsymbol{a_j}}, \boldsymbol{w}) \right\}. \tag{2}$$

Now define a *profile* L to be a triplet $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$, and let $\boldsymbol{\mathcal{L}}$ denote the space of all profiles. Given a profile L, define the *election matrix* $\boldsymbol{Y}^L$:

$$Y_{ij}^L = \int_{W_{a_i a_j}^{\boldsymbol{X}_{\mathcal{A}}, u}} f(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} - \int_{W_{a_j a_i}^{\boldsymbol{X}_{\mathcal{A}}, u}} f(\boldsymbol{w}) \mathrm{d}\boldsymbol{w}, \;\; \forall i, j \in \{1, 2, ..., A\}. \tag{3}$$

[3]For simplicity, I choose the normalized values to lie in [0,1], though any other interval $[x_{\min}, x_{\max}]$ would also work.

[4]For example, in the case of the HDI, u is a weighted geometric mean, while f could be set in the following manner: ask each country c to provide its importance function f_c on Δ^2 and then set $f = \sum_{c=1}^{C} f_c$ (where C is the total number of countries).

The entry Y^L_{ij} measures the net majority of weights resulting in higher welfare for alternative a_i than a_j, as measured by welfare function u applied to achievement vectors $\boldsymbol{x}_{\boldsymbol{a_i}}$ and $\boldsymbol{x}_{\boldsymbol{a_j}}$, appropriately weighted by the importance function f.

Similarly, we define the *proportion* matrix $\boldsymbol{V}^L$, where

$$V^L_{ij} = \frac{\int_{W^{\mathbf{x}_A,u}_{a_i a_j}} f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}}{\int_{\Delta^{I-1}} f(\boldsymbol{w})\mathrm{d}\boldsymbol{w}}, \quad \forall i,j \in \{1,2,...,A\},. \tag{4}$$

representing the proportion of weights in Δ^{I-1} resulting in weakly higher welfare for alternative a_i than a_j, as measured by welfare function u applied to achievement vectors $\boldsymbol{x}_{\boldsymbol{a_i}}$ and $\boldsymbol{x}_{\boldsymbol{a_j}}$, appropriately weighted by the importance function f.

In general, the integrals in Eqs. (3)-(4) will need to be tackled numerically. Nevertheless, we will see that analytic solutions are also possible for some compelling special cases of u and f (see Section 5).

Discrete importance functions. Before proceeding, I wish to note that the model can be straightforwardly extended to account for discrete importance functions f on a finite (or countable) subset of weights belonging in Δ^{I-1}. In this case, the integrals in Eqs. (3)-(4) would become summations over the relevant subset of Δ^{I-1} and all of the results in Sections 3 and 4 would naturally extend.

Embedding a completely nonparametric framework. On a final note, the model I have introduced can accommodate a completely nonparametric framework of analysis, in which there is no need to choose a functional form for the welfare function u. I briefly explain how. Consider the following discrete model instance with:

$$u(\boldsymbol{x},\boldsymbol{w}) = \sum_{i=1}^{I} w_i x_i,$$

and (here, $\boldsymbol{e}_i$ denotes the I-dimensional i'th standard basis vector)

$$f(\boldsymbol{w}) \;=\; \begin{cases} f_i \geq 0 & \text{if } \boldsymbol{w} = \boldsymbol{e}_i \in \{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_I\}, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where $\sum_{i=1}^{I} f_i > 0$. In this setting, the importance function f can be re-interpreted as the importance placed on each *indicator* i, as opposed to each vector of *weights* $\boldsymbol{w}$. Correspondingly, profiles featuring such u-f combinations lead to election matrices whose entries reflect the net majority of *indicators* favoring one alternative over another, appropriately weighted by their (i.e., the indicators') assumed importance.

3 Weight imprecision and social choice theory

3.1 Weight Vectors as Abstract Voters

Given a profile $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$, suppose we think of a vector of weights $\boldsymbol{w} \in \Delta^{I-1}$ as an imaginary voter who expresses their preferences over the alternatives in $\mathcal{A}$ via the application of welfare function u on the achievement vectors in $\boldsymbol{X}_{\mathcal{A}}$. That is, voter $\boldsymbol{w}$ (weakly) prefers an alternative a_i over a_j if and only if $u(\boldsymbol{x}_{\boldsymbol{a_i}}, \boldsymbol{w}) > u(\boldsymbol{x}_{\boldsymbol{a_j}}, \boldsymbol{w})$ ($u(\boldsymbol{x}_{\boldsymbol{a_i}}, \boldsymbol{w}) \geq u(\boldsymbol{x}_{\boldsymbol{a_j}}, \boldsymbol{w})$). Thus, voter $\boldsymbol{w}$'s preferences will be expressed as a possibly partial ordering of the alternatives.

Subsequently, construct an abstract electorate of voters by considering each $\boldsymbol{w} \in \Delta^{I-1}$ and introducing $f(\boldsymbol{w})$ (possibly irrational) copies of itself. Thus, the greater $f(\boldsymbol{w})$ is, the more voters holding $\boldsymbol{w}$'s preferences are introduced. This operation results in a continuum of voters $\mathcal{E}(f)$ of finite measure, since

$$0 < \int_{\mathcal{E}(f)} \mathrm{d}v = \int_{\Delta^{I-1}} f(\boldsymbol{w}) \mathrm{d}w < +\infty.$$

With the electorate $\mathcal{E}(f)$ in mind, the quantity $Y^L_{a_i a_j}$ of Eq. (4) defines the *net majority* of voters within $\mathcal{E}(f)$ preferring alternative a_i to a_j, under welfare function u applied to $\boldsymbol{x}_{\boldsymbol{a_i}}$ and $\boldsymbol{x}_{\boldsymbol{a_j}}$. Thus, all information on the pairwise preferences of the electorate $\mathcal{E}(f)$ over the set of alternatives $\mathcal{A}$ is succinctly summarized by the matrix $\boldsymbol{Y}^L$.

Given that it is unlikely that all weight vectors will yield the same ordered lists of alternatives, there will not be a ranking of $\mathcal{A}$ that is consistent with the preferences of all weights belonging to Δ^{I-1}. Thus, the question arises: In view of this inconclusiveness, what would constitute a "good" way of aggregating the preferences of all weight vectors, suitably weighted by the importance a decision maker places on them?

3.2 Kemeny's Rule

Variations of the above question have concerned philosophers and social scientists since the work of Condorcet and Borda in the 18th century. To state the problem formally, assume a finite set of voters $\mathcal{N} = \{1, 2, ..., N\}$ and alternatives $\mathcal{A} = \{a_1, a_2, ..., a_A\}$. A *ranking* R is a bijective map from $\mathcal{A}$ to $\{1, 2, ..., A\}$, where $R(a)$ is interpreted as the rank of alternative a. Let $\mathcal{R}_{\mathcal{A}}$ denote the set of all rankings of $\mathcal{A}$. A (voting) *rule* maps a set of N input rankings, representing voters' ordinal preferences, to a set of nonempty subsets of $\mathcal{R}_{\mathcal{A}}$, representing the socially optimal consensus ranking(s). The objective becomes that of selecting a voting rule that is, in some sense, "optimal".

In an important paper, Kemeny [20] introduced a rule which I will refer to as *Kemeny's rule*.[5] Given a set of individual rankings, Kemeny's rule produces a ranking (referred to as "Kemeny-optimal") that minimizes the sum of pairwise disagreements with respect to voter preferences. More formally, the *Kendall-τ distance* between two rankings R_1 and R_2, denoted by $\tau(R_1, R_2)$, is defined as the number of pairs (a_i, a_j) such that $R_1(a_i) > R_1(a_j)$ and $R_2(a_i) < R_2(a_j)$. Hence, $\tau(R_1, R_2)$ counts the number of (pairwise) relative rank disagreements between R_1 and R_2. Given a set of rankings $\mathcal{S}$ (corresponding to the ranked preferences of the set of voters), a *Kemeny-optimal* ranking is a ranking K that minimizes the function $\sum_{S \in \mathcal{S}} \tau(\cdot, S)$ over the set of rankings $\mathcal{R}_{\mathcal{A}}$, i.e., $K = \arg\min_{R \in \mathcal{R}_{\mathcal{A}}} \sum_{S \in \mathcal{S}} \tau(R, S)$. In the realm of social choice, Kemeny's rule is widely recognized as a very appealing method for aggregating individual preferences [25, 42].

Beyond its intuitiveness, Kemeny's rule rests on strong axiomatic foundations, which I briefly discuss. A voting rule is *anonymous* if it treats all voters alike, while it is *neutral* if it treats all alternatives alike. That is, the identity of a voter or alternative do not play a role in the consensus ranking. A rule satisfies *reinforcement* if, whenever two distinct groups of voters separately reach a common set of consensus rankings, then this set is also agreed upon when the two groups are merged. A rule satisfies the *Condorcet criterion* if, whenever there is an alternative that, when compared with every other, is preferred by a majority of voters, then this alternative is ranked first. A rule is *Condorcet* if it satisfies a variation of the Condorcet criterion, suitably interpreted for preference functions (see Young and Levenglick [43]). A voting rule is *unanimous* if, whenever all voters rank one alternative over another, then so does the consensus ranking. Finally, define an *interval* of a ranking to be a subset of alternatives that are ranked in succession, i.e., that appear "together" in the ranking without gaps. A rule satisfies *local independence of irrelevant alternatives (LIIA)* if the ranking of alternatives within each interval remains fixed if we ignore alternatives outside of it (see Young [41, 42]). For instance, this means that the ranking of the alternatives at the top of the consensus ranking is unaffected if we remove those in the bottom and re-apply the rule to the restricted problem. As a result, LIIA implies that the voting rule cannot be manipulated by adding unrealistically bad or good alternatives to the mix. In an important paper, Young and Levenglick [43] proved that Kemeny's rule is the only voting rule that is *Condorcet* and satisfies *neutrality* and

[5] In the literature it is also sometimes referred to as the "maximum likelihood" or the "Condorcet" rule. As Young noted in [41], the Marquis de Condorcet was the first to propose this way of deciding an election, even if he did not work out the formal details.

reinforcement. What is more, this characterization can be used to establish that Kemeny's rule is the unique voting rule satisfying *anonymity, neutrality, unanimity, reinforcement, and LIIA* (see Young [41, 42]).

Yet, the appeal of the Kemeny rule extends beyond purely normative considerations. This is because Young [41] showed that, under a particular kind of statistical framework, Kemeny's rule can be characterized using maximum likelihood estimation. That is, assuming the existence of an objectively "correct" ranking of which voter preferences provide a particular kind of noisy signal, Kemeny's rule provides the ranking which maximizes the associated likelihood function.[6] This finding has proven to be robust to generalizations of Young's basic model (Conitzer [11]), suggesting the existence of an epistemic basis behind the adoption of Kemeny's rule.

Application to multidimensional welfare indices. I now proceed to apply the conceptual apparatus of Kemeny's rule to rankings based on multidimensional welfare indices. Fix the set of alternatives $\mathcal{A}$ and set of indicators $\mathcal{I}$. Given a profile $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$ and a weight vector $\boldsymbol{w} \in \Delta^{I-1}$, let $R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}},u}$ denote the *partial* ranking of alternatives in $\mathcal{A}$ that this vector of weights implies, via the application of the welfare function u to $\boldsymbol{w}$ and the achievement matrix $\boldsymbol{X}_{\mathcal{A}}$. That is, for all pairs of alternatives (a_i, a_j), $R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}},u}(a_i) \leq R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}},u}(a_j) \Leftrightarrow u(\boldsymbol{x}_{\boldsymbol{a_i}}, \boldsymbol{w}) \geq u(\boldsymbol{x}_{\boldsymbol{a_j}}, \boldsymbol{w})$. Note that this allows for ties between alternatives, hence the fact that $R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}},u}$ is a *partial* ranking.

In the usual formulation of Kemeny's rule, the set of voters is finite so the sum $\sum_{S \in \mathcal{S}} \tau(\cdot, S)$ is well-defined. In our context, recall how we constructed the electorate $\mathcal{E}(f)$: each $\boldsymbol{w}$ in Δ^{I-1} corresponds to $f(\boldsymbol{w})$ voters, each having the partial input ranking $R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}},u}$. Thus, given a profile $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$, the Kemeny-optimal set of rankings K^L is given by:

$$K^L = \underset{R \in \mathcal{R}_{\mathcal{A}}}{\arg\min} \left\{ \int_{\Delta^{I-1}} \tau\left(R, R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}},u}\right) f(\boldsymbol{w}) d\boldsymbol{w} \right\}. \tag{6}$$

Note that while input rankings $R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}},u}$ are allowed to be partial, K^L still has to be a set of full rankings. This generalization of Kemeny's rule to allow for partial input rankings is consistent with other contributions, such as Ailon [1] and Van Zuylen and Williamson [44].

To apply formula (6) to our context, it is necessary to identify the partial rankings $R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}},u}$ for at least all $\boldsymbol{w} \in \Delta^{I-1}$ satisfying $f(\boldsymbol{w}) \neq 0$. However, this challenging task can be sidestepped, since to calculate a Kemeny-optimal ranking we only need the results

[6] Interestingly, under the same statistical framework, when the objective is to determine just the first-place alternative and not the entire ranking, it is Borda's rule [40] that provides the maximum likelihood estimate (Young [41]).

of all *pairwise* comparisons between elements in $\mathcal{A}$, given by matrix $\boldsymbol{Y}^L$. Thus, letting $(a_i, a_j) \in \mathcal{A} \times \mathcal{A}$ denote ordered pairs of alternatives in $\mathcal{A}$ and $\mathbf{1}\{\cdot\}$ the indicator function, we may write, for all $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$,

$$K(L) \equiv K^L = \underset{R \in \mathcal{R}_{\mathcal{A}}}{\arg\min} \sum_{(a_i, a_j) \in \mathcal{A} \times \mathcal{A}} \mathbf{1}\{R(a_i) < R(a_j)\} Y_{ji}^L. \tag{7}$$

Assuming that a set of $M \geq 1$ rankings $\{R_1^L, R_2^L, ..., R_M^L\}$ attains the minimum in Eq (7), K^L will be expressed as an $A \times M$ matrix, where entry K_{im}^L equals $R_m^L(a_i)$.

As a final note, I briefly comment on the applicability of partial input rankings within the context of the model. Consider the restricted domain of profiles $\tilde{\mathcal{L}}$ where

$$\tilde{\mathcal{L}} = \left\{ L \in \mathcal{L} : \;\; V_{ij}^L + V_{ji}^L = 1, \;\; \forall (a_i, a_j) \in \mathcal{A} \times \mathcal{A} \text{ s.t. } i \neq j \right\}. \tag{8}$$

The domain restriction $\tilde{\mathcal{L}}$ is equivalent to imposing that the set of weights resulting in equal welfare for any pair of alternatives will, upon being appropriately weighted by the importance function, have negligible Lebesgue measure in $\Re^{I-1}$. As a result, weight vectors resulting in ties between alternatives can be ignored, as those vectors are, in a relative sense, too "few" to have an effect on the selection of the consensus rankings. Indeed, V_{ij}^L can be interpreted as the proportion of weights strictly favoring a_i over a_j given profile L.

The above restriction will be applicable for widely-used profile classes (see [12] and Section 5.1), as the following Proposition suggests.

Proposition 1 *Consider the class of welfare functions $\mathcal{U}^*$, where $u \in \mathcal{U}^*$ if and only if it can be expressed as either (a) $u(\boldsymbol{x}, \boldsymbol{w}) = g\left(\sum_{i=1}^{I} w_i u_i(x_i)\right)$, or (b) $u(\boldsymbol{x}, \boldsymbol{w}) = g\left(\prod_{i=1}^{I} u_i(x_i)^{w_i}\right)$, where u_i are positive functions that are strictly monotone on $[0, 1]$, and g is an invertible real-valued function. Then, for any profile L satisfying $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$ with $u \in \mathcal{U}^*$ and $\boldsymbol{x}_{\boldsymbol{a_i}} \neq \boldsymbol{x}_{\boldsymbol{a_j}}$ for all $i \neq j$, we have $L \in \tilde{\mathcal{L}}$.*

Proof. See Appendix.

3.3 An Axiomatic Characterization

In this section, we explore the normative implications of adopting rule K. This requires adapting to our setting the properties discussed in Section 3.2.

Define a *rule* ϕ to be a function from $\mathcal{L}$ to the set of nonempty subsets of $\mathcal{R}_A$. For a profile $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$ and a ranking R, let

$$n_R^L = \int_{\Delta^{I-1}} \mathbf{1}\left\{R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}}, u} = R\right\} f(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \tag{9}$$

denote the f-weighted Lebesgue measure of the set of weights in Δ^{I-1} having ranking R, given welfare function u and achievement matrix $\boldsymbol{X}_{\mathcal{A}}$.

Definition 1 *A rule ϕ satisfies* **anonymity** *if for all profiles L, ϕ depends only on the numbers n_R^L for each $R \in \mathcal{R}_A$.*

Anonymity ensures that no weight vector is given special treatment in the underlying social choice problem. Instead, the ordinal preferences of all weights $\boldsymbol{w}$ receive consideration according to their importance $f(\boldsymbol{w})$. As a result, cardinal information on the *intensity*, or magnitude, of an alternative's dominance over another is not allowed to directly affect the consensus ranking. To illustrate this point more fully, consider the following example. Suppose we have two sets of weights W_1 and W_2 and a set of two alternatives $\mathcal{A} = \{a_1, a_2\}$ such that:

$$\int_{W_1} f(\boldsymbol{w})\mathrm{d}\boldsymbol{w} = \int_{W_2} f(\boldsymbol{w})\mathrm{d}\boldsymbol{w},$$
$$u(\boldsymbol{x_{a_1}}, \boldsymbol{w}^1) = 1000 > u(\boldsymbol{x_{a_2}}, \boldsymbol{w}^1) = 1, \;\; \forall \boldsymbol{w}^1 \in W_1,$$
$$u(\boldsymbol{x_{a_1}}, \boldsymbol{w}^2) = 2 > u(\boldsymbol{x_{a_2}}, \boldsymbol{w}^2) = 1, \;\; \forall \boldsymbol{w}^2 \in W_2.$$

What anonymity implies is that the sets of weights W_1 and W_2 will play the *exact same* role in the determination of the consensus rankings. Furthermore, it ensures that, all else being equal, we could change $u(\boldsymbol{x_{a_1}}, \boldsymbol{w}^1)$ from 1000 to 1.1 for all $\boldsymbol{w}^1 \in W_1$, and the consensus rankings would remain unaffected.

Definition 2 *A rule ϕ is* **neutral** *if for all profiles $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$ and permutations σ of alternatives in $\mathcal{A}$, $\phi(\boldsymbol{X}_{\sigma\circ\mathcal{A}}, f, u) = \sigma \circ \phi(\boldsymbol{X}_{\mathcal{A}}, f, u)$.*

Neutrality imposes that the identity of alternatives not matter to the rank they receive. Instead, the rule is symmetric in its treatment of alternatives: if their identities were permuted, then the consensus rankings would be similarly permuted.

Definition 3 *A rule ϕ satisfies* **reinforcement** *if, for all profiles $L_1 = (\boldsymbol{X}_{\mathcal{A}}, f_1, u)$ and $L_2 = (\boldsymbol{X}_{\mathcal{A}}, f_2, u)$ such that f_1 and f_2 have non-overlapping supports in Δ^{I-1},*

$$\phi(L_1) \cap \phi(L_2) \neq \emptyset \Rightarrow \phi(\boldsymbol{X}_{\mathcal{A}}, f_1 + f_2, u) = \phi(L_1) \cap \phi(L_2).$$

Reinforcement sets the following requirement: if, *ceteris paribus*, two importance functions having non-overlapping support separately lead to two sets of consensus rankings whose intersection is nonempty, then applying the rule to the profile having the sum of the importance functions should acknowledge and reinforce this agreement.

Reinforcement imposes a degree of consistency to the way a rule aggregates individual preferences. Take for instance the case of the HDI whose indicator set involves three criteria: (1) health (2) education, and (3) income. Suppose two groups of countries, say Africa and Europe, have completely differing opinions regarding the correct weighting scheme of the three dimensions of the HDI. In particular, African countries claim that the only viable weight vectors are those that assign the greatest weight to health, the second greatest to income, and the third greatest to education, while no indicator can receive a weight that exceeds 0.5. All weight vectors satisfying the above restriction are equally valid, while all others should not be considered at all. European countries have similarly structured preferences, except that they claim that highest weight must always be given to education, second-highest to health, and third-highest to income. The above value judgments can be encoded in importance functions that are equal to 1 in the relevant subsets of Δ^2 and zero everywhere else. Suppose now that the United Nations chooses a method of ranking countries that, when considering the opinions of Africa and Europe *separately* leads to the same consensus ranking. In that case, it is eminently desirable to additionally require that the UN's method, when considering the preferences of Africa and Europe *jointly*, not disturb their pre-existing consensus.

Definition 4 *A rule ϕ is* **extended-Condorcet** *if, for all profiles L, if there exists a partition $(\mathcal{A}_1, \mathcal{A}_2)$ of $\mathcal{A}$ such that for all $(a_i, a_j) \in \mathcal{A}_1 \times \mathcal{A}_2$ we have $Y_{ij}^L > 0$, then $R(a_i) < R(a_j)$ for all $(a_i, a_j) \in \mathcal{A}_1 \times \mathcal{A}_2$, for all $R \in \phi(L)$.*

Thus, a rule is extended-Condorcet if it obeys the majority wishes of the electorate $\mathcal{E}(f)$, whenever these do not imply circular ambiguities majority pairwise preferences. This property generalizes the concept of the Condorcet criterion to settings in which the objective is to choose an entire consensus ranking, not just the winning alternative.[7]

Definition 5 *A rule ϕ is* **unanimous** *if, for all profiles L, if there exists a ranking R such that for all pairs $(a_i, a_j) \in \mathcal{A} \times \mathcal{A}$, $R(a_i) < R(a_j) \Rightarrow \{V_{ij}^L = 1$ and $V_{ji}^L = 0\}$, then $\phi(L) = R$.*

Unanimity requires that, if all but a negligible set of weights submit the same ranking of the alternatives, then the rule should respect this complete consensus. It is a standard property in many problems in social choice.

[7]It should be noted that this property is not the same as the Condorcet property of Young and Levenglick's [43] characterization. The latter is a little more intricate.

Definition 6 *A rule ϕ satisfies* **local independence of irrelevant alternatives (LIIA)** *if, for all profiles $(\boldsymbol{X}_\mathcal{A}, f, u)$, for all $R \in \phi(\boldsymbol{X}_\mathcal{A}, f, u)$, for every interval B of R, $R_B \in \phi(\boldsymbol{X}_{\mathcal{A}_B}, f, u)$ where $\boldsymbol{X}_{\mathcal{A}_B}$ represents $\boldsymbol{X}_\mathcal{A}$ restricted to alternatives in B, and R_B represents R's relative ordering of alternatives in B; moreover, all rankings in $\phi(\boldsymbol{X}_{\mathcal{A}_B}, f, u)$ can be generated in this way.*

LIIA is a little more subtle than the previous properties. But, some reflection also suggests that it is a desirable stabilizing property for a rule to satisfy. It implies that the ordering of alternatives that are ranked "together" (i.e., consecutively) in a consensus ranking should not change if we apply the rule to the restricted problem that focuses just on these alternatives and ignores all others. Usually such intervals correspond to meaningful categories of alternatives. To take an example from the empirical application I will pursue in Section 6, suppose that a composite index is measuring academic excellence and has ranked the 100 best universities in the world. We would prefer the relative ordering of the top 20 (representing, say, Tier 1 institutions or something to that effect), to remain unchanged if we, for instance, re-apply the rule by ignoring those universities ranked 91-100, 51-100, or even the entire 21-100 for that matter. The same argument would apply for the relative ordering of Tier 2 institutions, or indeed any other meaningful interval within the consensus ranking. For more information on LIIA the interested reader can refer to Young [42].

Part (ii) of the following theorem is a consequence of the results of Young and Levenglick [43] and Young [40, 41] adapted to our setting.

Theorem 1 *Consider the rule K of Eq. (7).*

(i) On the domain $\mathcal{L}$, K satisfies anonymity, neutrality, reinforcement, extended-Condorcet, unanimity, and LIIA.

(ii) Let $\mathcal{Y}^Q$ denote the set of rational skew-symmetric matrices whose rows and columns are indexed by the elements of $\mathcal{A}$. On the restricted domain $\mathcal{L}^Q = \{L \in \mathcal{L} : \ \boldsymbol{Y}^L \in \mathcal{Y}^Q\}$, K uniquely satisfies anonymity, neutrality, reinforcement, unanimity, and LIIA.

Proof. See Appendix. ■

Theorem 1 makes clear a set of five reasonable properties that characterize the rule K. In addition, K satisfies the extended-Condorcet criterion. Ultimately, if one finds these properties compelling, then they may also look favorably upon K as a way of dealing with weight imprecision in multidimensional welfare indices. In any case, the normative basis of K is made evident.

4 Computational issues

Eq. (7) summarizes the Kemeny-optimal ranking we wish to find. We would at this point be done, if not for the fact that the associated optimization problem is NP-hard [7], even when the number of indicators is just four [15].

The main difficulty arises from the potential presence of Condorcet cycles, which imply intransitive majority pairwise preferences. Formally, a set of alternatives $\{a_{i_1}, a_{i_2}, ..., a_{i_M}\}$ forms a *Condorcet cycle* if $\min\{Y^L_{12}, Y^L_{23}, ..., Y^L_{M-1M}, Y^L_{M1}\} \geq 0$. If no such cycles existed, finding a unique Kemeny-optimal ranking would be a straightforward affair (see Theorem 2 below). Thus, it is important to identify and, in some fashion, resolve these cycles.

Graph-theoretic preliminaries. In this section, I briefly introduce some relevant graph-theoretic concepts that will be useful later. A *directed graph* G is a pair $G = (\mathcal{V}, E)$, where $\mathcal{V}$ is a set of *vertices* and E a set of ordered pairs of vertices, referred to as *edges*.

Given two subsets $\mathcal{V}_1$, $\mathcal{V}_2$ of $\mathcal{V}$ such that $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$, let

$$in(\mathcal{V}_1, \mathcal{V}_2) = \left|\{e \in E : e = (u, v), \;\; u \in \mathcal{V}_2, \; v \in \mathcal{V}_1\}\right| . \tag{10}$$

Thus Eq. (10) denotes the number of edges leading to vertices in $\mathcal{V}_1$, originating from vertices in $\mathcal{V}_2$.

A *subgraph* $G' = (\mathcal{V}', E')$ of a directed graph $G = (\mathcal{V}, E)$ is a graph such that $\mathcal{V}' \subset \mathcal{V}$ and $E' = \{(v, u) \in E : \;\; v, u \in \mathcal{V}'\}$. A directed graph is called *strongly connected* if there is a path from each vertex to every other vertex (in both directions). The *strongly connected components* of a directed graph G are its maximal-size strongly connected subgraphs.

***L*-Dominance graphs and Kemeny rankings.** Given a profile L, define the directed graph $G^L = \left(\mathcal{A}, E^L\right)$, whose vertices correspond to alternatives in $\mathcal{A}$ and whose edges satisfy $E^L = \left\{(a_i, a_j) \in \mathcal{A} \times \mathcal{A} : \;\; Y^L_{a_i a_j} \geq 0, \;\; i \neq j\right\}$. Thus, a pair (a_i, a_j) belongs to G^L's edge set if and only if a weak net majority of the voters in the electorate $\mathcal{E}(f)$ assign to a_i a higher welfare than a_j, as measured by u applied to $\boldsymbol{x}_{\boldsymbol{a_i}}$ and $\boldsymbol{x}_{\boldsymbol{a_j}}$ – with no self-edges (a_i, a_i) allowed. We refer to G^L as an *L-dominance* graph, as it captures the above dominance relationship between all pairs of alternatives. Clearly, a cycle in G^L corresponds to a Condorcet cycle in the underlying social choice problem. Moreover, alternatives forming such a cycle must, by definition, belong to the same strongly connected component of G^L. As a result, if no Condorcet cycles existed, then all strongly connected components of G^L would be singletons.

The following Theorem exploits the special structure of G^L to connect it to the Kemeny-optimal ranking K^L of Eq. (7).

Theorem 2 *Consider a profile* $L = (\boldsymbol{X}_{\mathcal{A}}, f, u) \in \boldsymbol{\mathcal{L}}$ *and its* L*-dominance graph* $G^L = \left(\mathcal{A}, E^L\right)$*. Let* $G' = (\mathcal{A}', E')$ *denote a strongly connected component of* G^L *and define the profile* $L_{\mathcal{A}'} = (\boldsymbol{X}_{\mathcal{A}'}, f, u)$*, where* $\boldsymbol{X}_{\mathcal{A}'}$ *is the restriction of* $\boldsymbol{X}_{\mathcal{A}}$ *to alternatives in* $\mathcal{A}'$*. We have the following holding (here* $K^L_{\mathcal{A}'}$ *is the restriction of* K^L *to alternatives in* $\mathcal{A}'$*):*

$$K^L_{\mathcal{A}'} = \frac{in\,(\mathcal{A}', \mathcal{A}\backslash\mathcal{A}')}{|\mathcal{A}'|} \cdot \boldsymbol{e} + K^{L_{\mathcal{A}'}}, \tag{11}$$

where $\boldsymbol{e}$ *is a matrix of ones of appropriate dimension.*

Proof. See Appendix. ■

Example 1. Theorem 2 is best illustrated with an example. Consider a profile L with six alternatives leading to the following election matrix:

$$\boldsymbol{Y}^L = \begin{bmatrix} 0 & -2 & -3.5 & .3 & 1 & .7 \\ 2 & 0 & -1 & 1 & -1.2 & .6 \\ 3.5 & 1 & 0 & 2 & 3 & 1.5 \\ -.3 & -1 & -2 & 0 & -4 & 0 \\ -1 & 1.2 & -3 & 4 & 0 & 1.5 \\ -.7 & -1 & -1.5 & 0 & -1.5 & 0 \end{bmatrix}$$

The L-dominance graph corresponding to election matrix $\boldsymbol{Y}^L$ is seen in Figure 1. It has the following three strongly connected components:

$$G_1 = (\{a_3\}, \emptyset)\,,\;\; G_2 = (\{a_1, a_2, a_5\},\; \{(a_1, a_5), (a_5, a_2), (a_2, a_1)\})\,, G_3 = (\{a_4, a_6\},\; \{(a_4, a_6), (a_6, a_4)\})\,,$$

implying the respective (sub)profiles $L_{\mathcal{A}_1}, L_{\mathcal{A}_2}, L_{\mathcal{A}_3}$. It is immediate to see that, among voters in $\mathcal{E}(f)$, alternative a_3 enjoys a net majority over all other alternatives. Furthermore, alternatives in $\mathcal{A}_2$ form a Condorcet cycle, since a net majority of voters in $\mathcal{E}(f)$ prefer a_1 to a_5, a_5 to a_2, and a_2 to a_1. Moreover, these alternatives are all preferred by a majority of voters in $\mathcal{E}(f)$ to those in $\mathcal{A}_3$. Conversely, alternatives in $\mathcal{A}_3$ receive exactly equal support from voters in $\mathcal{E}(f)$, and lose all pairwise contests against all other alternatives.

Applying Theorem 2 we see that, in any Kemeny-optimal ranking, a_3 will be first, while the ranks of (a_1, a_2, a_5) will range between 2 and 4, and those of (a_4, a_6) between 5 and 6. By inspection, it is easy to optimally resolve the Condorcet cycle of graph G_2 and

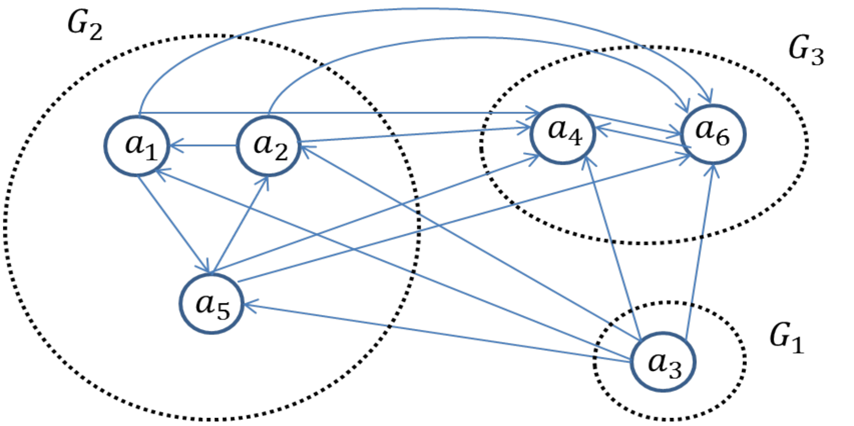

Figure 1: L-dominance graph of Example 1. Strongly connected components in dashed ovals.

determine the precise Kemeny-optimal ranks of a_1, a_2 and a_5: $K^{L_{\mathcal{A}_2}}(a_5) = 1,\ K^{L_{\mathcal{A}_2}}(a_2) = 2,\ K^{L_{\mathcal{A}_2}}(a_1) = 3$. Indeed, we obtain:

$$K^L_{a_3} = [0] + [1] = [1],$$

$$K^L_{a_1,a_2,a_5} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix},$$

$$K^L_{a_4,a_6} = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 6 \\ 6 & 5 \end{bmatrix}.$$

Putting $K^L_{a_3}$, $K^L_{a_1,a_2,a_5}$ and $K^L_{a_4,a_6}$ together we obtain the two Kemeny-optimal rankings of profile L:

$$K^L = \begin{bmatrix} 4 & 4 \\ 3 & 3 \\ 1 & 1 \\ 5 & 6 \\ 2 & 2 \\ 6 & 5 \end{bmatrix}.$$

Computational implementation. The strongly connected components of the graph G^L can be efficiently identified with Tarjan's algorithm [37]. Thus, Theorem 2 suggests that problem (7) may be efficiently reduced to a series of smaller sub-problems, corresponding to the strongly connected components of graph G^L.

The question now becomes: how do we deal with an individual strongly connected component $G' = (\mathcal{A}', E')$? If $|\mathcal{A}'|$ is small enough (for a regular PC running Matlab this means less than or equal to 10), then $K^{L'}$ can be determined via brute-force enumeration within a few minutes. Otherwise, for practicality, it is preferable to resort to polynomial-time approximation algorithms.[8] Here, we distinguish between two cases. If the profile L' satisfies restriction (8), then, the problem is one of *full rank aggregation*, and we may employ a 4/3-approximation algorithm due to Van Zuylen and Williamson (VZW) [44] (see Corollary 5.3 and DerandFASLP-Pivot in Figure 1 of [44]). This provides the best approximation guarantee (of 4/3) currently available. Conversely, if the profile L' does not satisfy restriction (8), then the problem is one of *partial rank aggregation*, and we employ the combinatorial algorithm of VZW outlined in Theorem 2.3 of [44], which in turn builds on a randomized algorithm of Ailon [1]. This has an approximation guarantee of 8/5.

The above observations are summarized in the following algorithm:

Algorithm 1. **Input:** $L = (\boldsymbol{X}_{\mathcal{A}}, f, u)$**,** M

1. Compute the election matrix Y^L.
2. Construct graph G^L.
3. Run Tarjan's algorithm on G^L.
4. For every strongly connected component $G' = (\mathcal{A}', E')$ of G^L:

   4a. If $|\mathcal{A}'| \leq M$, determine $K^{L_{\mathcal{A}'}}$ via enumeration.

   4b. If $|\mathcal{A}'| > M$ and $L' \in \tilde{\mathcal{L}}$, approximate $K^{L_{\mathcal{A}'}}$ by Corollary 5.3 in VZW [44].

   4c. If $|\mathcal{A}'| > M$ and $L' \notin \tilde{\mathcal{L}}$, approximate $K^{L_{\mathcal{A}'}}$ by Theorem 2.3 in VZW [44].

5. Apply Eq. (11) to all strongly connected components of G^L. Where applicable, approximate $K^{L_{\mathcal{A}'}}$ with the VZW algorithm output.

Step 1 will usually need to be tackled numerically, via Monte Carlo methods, though some profile spaces may admit analytic solutions (see Section 5). When a strongly connected component of G^L is large (e.g., having more than 20-25 vertices) and step 4b is applicable, then, to speed up the running time of the relevant VZW approximation algorithm, it is advisable to use a specialized LP-solver to quickly solve the linear programming relaxation of the problem, as the latter is a subroutine of the main algorithm. Conversely, the VZW algorithm for step 4c is a combinatorial algorithm that does not

[8] Clearly, the determination of what are considered "big" and "small" values for $|\mathcal{A}'|$ will depend on computing power and memory.

require linear programming. Finally, in cases where approximation algorithms are employed, the output ranking(s) of Algorithm 1 can be potentially improved by applying to them a local search heuristic known as "local Kemenization" (Dwork et al. [15]).

5 A Special Case of the Model

5.1 Generalized weighted means

A family of welfare functions that is particularly popular in many policy contexts are known as generalized weighted means [3, 29, 12, 16]. Parameterized by $\gamma \in \Re$, they are denoted by u^γ and satisfy

$$u^\gamma(\boldsymbol{x}, \boldsymbol{w}) = \begin{cases} \left(\sum_{i=1}^I w_i x_i^\gamma\right)^{\frac{1}{\gamma}} & \gamma \neq 0, \\ \prod_{i=1}^I x_i^{w_i} & \gamma = 0. \end{cases} \tag{12}$$

When $\gamma = 1$ we recover the weighted arithmetic mean, whereas $\gamma = 0$ or $\gamma = -1$ imply a weighted geometric and harmonic mean, respectively. As $\gamma \to +\infty(-\infty)$, $u^\gamma(\boldsymbol{x}, \boldsymbol{w})$ converges to the maximum (minimum) coordinate of $\boldsymbol{x}$.

I note a simple corollary of the definition of Proposition 1.

Corollary 1 *The family of generalized weighted means (12) belongs to the class $\mathcal{U}^*$ discussed in Proposition 1.*

As a result, we do not need to account for partial input rankings.

5.2 Weight imprecision via ϵ-contamination

Assume we are given an initial vector of weights $\bar{\boldsymbol{w}} \in \Delta^{I-1}$, representing a preliminary and tentative choice of weighting scheme. Now, suppose that we are willing to grant equal consideration to weights deviating from $\bar{\boldsymbol{w}}$ belonging in the set $W^\epsilon_{\bar{\boldsymbol{w}}}$, where[9]

$$W^\epsilon_{\bar{\boldsymbol{w}}} \equiv W^\epsilon = (1-\epsilon)\bar{\boldsymbol{w}} + \epsilon\Delta^{I-1} = \left\{ \boldsymbol{w} \in \Re^I : \ \boldsymbol{w} \geq (1-\epsilon)\bar{\boldsymbol{w}}, \ \sum_{i=1}^I w_i = 1 \right\}. \tag{13}$$

Here, the parameter $\epsilon \in [0, 1]$ measures the imprecision associated with the initial vector of weights $\bar{\boldsymbol{w}}$. If $\epsilon = 0$, then we are completely confident in our choice of $\bar{\boldsymbol{w}}$, while if $\epsilon = 1$ we assign no special status to $\bar{\boldsymbol{w}}$ and consider all possible weight vectors equally valid. Originally developed in Bayesian analysis, this way of parameterizing probabilistic imprecision is known as ϵ-contamination (the interested reader may consult Berger and

[9] In what follows I will suppress dependence of the results on $\bar{\boldsymbol{w}}$ to avoid very cumbersome notation.

Berliner [8] for additional information). The ϵ-contamination parametric structure has been studied in the decision-theoretic literature on Knightian uncertainty ([27, 22], among many others), which however places more emphasis on the normative foundations and behavioral implications of such belief imprecision. Application areas have included job search [26] and optimal investment [28].

First introduced by Foster et al. [16] in the context of composite indices of welfare, this way of parameterizing imprecision over $\boldsymbol{w}$ is equivalent to the following importance function f^ϵ:

$$f^\epsilon(\boldsymbol{w}) \;=\; \begin{cases} \frac{1}{Vol(W^\epsilon)} & \boldsymbol{w} \in W^\epsilon, \\ 0 & otherwise, \end{cases} \tag{14}$$

where Vol denotes volume in $\Re^{I-1}$.

For a pair of alternatives $(a_i, a_j) \in \mathcal{A} \times \mathcal{A}$ the set $W^{u^\gamma}_{a_i a_j}$ is equal to

$$W^{u^\gamma}_{a_i a_j} = \left\{ \boldsymbol{w} \in \Delta^{I-1} : \;\; u^\gamma(\boldsymbol{x}_{\boldsymbol{a_i}}, \boldsymbol{w}) \geq u^\gamma(\boldsymbol{x}_{\boldsymbol{a_j}}, \boldsymbol{w}) \right\}.$$

Now, we introduce the difference vector $\boldsymbol{d}^\gamma_{\boldsymbol{a_i a_j}}$, abbreviated for simplicity by $\boldsymbol{d}^\gamma$, where, for all $k = 1, 2, ..., I$:

$$d^\gamma_k \;=\; \begin{cases} x^\gamma_{a_{ik}} - x^\gamma_{a_{jk}}, & \text{if } \gamma > 0, \\ x^\gamma_{a_{jk}} - x^\gamma_{a_{ik}}, & \text{if } \gamma < 0, \\ \log\left(x_{a_{ik}}\right) - \log\left(x_{a_{jk}}\right), & \text{if } \gamma = 0. \end{cases} \tag{15}$$

It is straightforward to see that the set $W^{u^\gamma}_{a_i a_j}$ can be simplified into the following polytope[10]

$$W^{u^\gamma}_{a_i a_j} = \left\{ \boldsymbol{w} \in \Delta^{I-1} : \;\; (\boldsymbol{d}^\gamma)' \, \boldsymbol{w} \geq 0 \right\}. \tag{16}$$

Subsequently, define the set $W^{\epsilon,\gamma}_{a_i a_j}$ as the intersection of W^ϵ and $W^{u^\gamma}_{a_i a_j}$, so that

$$W^{\epsilon,\gamma}_{a_i a_j} = \left\{ \boldsymbol{w} \in \Re^I : \;\; \boldsymbol{w} \geq (1-\epsilon)\bar{\boldsymbol{w}}, \;\; \sum_{i=1}^{I} w_i = 1, \;\; (\boldsymbol{d}^\gamma)' \, \boldsymbol{w} \geq 0 \right\}. \tag{17}$$

Then, applying Eqs. (3)-(4) to profiles satisfying $L = (\boldsymbol{X}_\mathcal{A}, u^\gamma, f^\epsilon)$ for a choice of $\epsilon \in (0, 1]$ and $\gamma \in \Re$, we may write:

$$V^L_{ij} \equiv V^{\epsilon,\gamma}_{ij} = \frac{Vol\left(W^{\epsilon,\gamma}_{a_i a_j}\right)}{Vol(W^\epsilon)}, \tag{18}$$

$$Y^L_{ij} \equiv Y^{\epsilon,\gamma}_{ij} = V^{\epsilon,\gamma}_{ij} - V^{\epsilon,\gamma}_{ji}. \tag{19}$$

[10] This follows by the following equivalence

$$u^\gamma(\boldsymbol{x}_{\boldsymbol{a_i}}, \boldsymbol{w}) \geq u^\gamma(\boldsymbol{x}_{\boldsymbol{a_j}}, \boldsymbol{w}) \quad \Leftrightarrow \quad \begin{cases} \sum_{i=1}^I w_i x^\gamma_{a_i i} \geq \sum_{i=1}^I w_i x^\gamma_{a_j i}, & \text{if } \gamma > 0, \\ \sum_{i=1}^I w_i x^\gamma_{a_i i} \leq \sum_{i=1}^I w_i x^\gamma_{a_j i} & \text{if } \gamma < 0, \\ \sum_{i=1}^I w_i \log\left(x_{a_j i}\right) \geq \sum_{i=1}^I w_i \log\left(x_{a_j i}\right) & \text{if } \gamma = 0. \end{cases}$$

We focus on the more intuitive quantity of $V_{ij}^{\epsilon,\gamma}$. By Corollary 1 we know that, unless $\boldsymbol{x_{a_i}} = \boldsymbol{x_{a_j}}$, $V_{ij}^{\epsilon,\gamma} + V_{ji}^{\epsilon,\gamma} = 1$. Now, basic geometric reasoning allows us to establish an unambiguous monotonicity property of $V_{ij}^{\epsilon,\gamma}$ with respect to the level of imprecision ϵ.

Theorem 3 *Consider a pair of alternatives (a_i, a_j) and their difference vector $\boldsymbol{d}^\gamma$ of Eq. (15).*

(i) Suppose $(\boldsymbol{d}^\gamma)' \bar{\boldsymbol{w}} \neq 0$. Then, $V_{ij}^{\epsilon,\gamma}$ is monotonic in $\epsilon \in (0,1]$. It is weakly increasing (decreasing) in ϵ if

$$(\boldsymbol{d}^\gamma)'\bar{\boldsymbol{w}} < (>)0.$$

(ii) Suppose $(\boldsymbol{d}^\gamma)' \bar{\boldsymbol{w}} = 0$. Then, $V_{ij}^{\epsilon,\gamma}$ is constant in $\epsilon \in (0,1]$.

Proof. See Appendix. ■

Thus, when $(\boldsymbol{d}^\gamma)' \bar{\boldsymbol{w}} \neq 0$, and a_i and a_j do not yield identical welfare under the initial vector of weights $\bar{\boldsymbol{w}}$, Theorem 3 establishes that the proportion of weights favoring one alternative over another varies monotonically in the imprecision ϵ attached to $\bar{\boldsymbol{w}}$. The direction of the relationship depends on the comparison of alternatives a_i and a_j under initial weights $\bar{\boldsymbol{w}}$. It is decreasing if a_i initially dominates a_j and increasing otherwise. Conversely, when the initial weights yield identical composite scores for a_i and a_j, then $V_{ij}^{\epsilon,\gamma}$ remains constant as we vary ϵ.

The next step is to find a simple way to calculate the ratios $V_{ij}^{\epsilon,\gamma}$. To this end, I need to first express polytope $W_{a_i a_j}^{\epsilon,\gamma}$ as a system of linear inequalities. Define the function

$$D^{\epsilon,\gamma} = -\frac{1-\epsilon}{\epsilon} (\boldsymbol{d}^\gamma)' \bar{\boldsymbol{w}}, \tag{20}$$

and suppose that there exists at least one indicator $i^* \in \mathcal{I}$ such that $d_{i^*}^\gamma \geq D^{\epsilon,\gamma}$. If such an i^* does not exist, then we may immediately conclude that $V_{ij}^{\epsilon,\gamma} = 0$ (see proof of Theorem 3). Define $\boldsymbol{w^*} \in \Re^{I-1}$ and $\boldsymbol{d}^{\gamma *} \in \Re^{I-1}$ as the restriction of vectors $\boldsymbol{w}$ and $\boldsymbol{d}^\gamma$ to indicators in $\mathcal{I} \backslash \{i^*\} \equiv \mathcal{I}^*$. Consider the following polytope $W_{a_i a_j}^{\epsilon,\gamma *}$ (here $\boldsymbol{e}$ denotes a vector of all ones of dimension $I-1$)

$$W_{a_i a_j}^{\epsilon,\gamma *} = \left\{ \boldsymbol{w^*} \in \Re^{I-1} : \ \boldsymbol{w^*} \geq 0, \ \sum_{i \in \mathcal{I}^*} w_i^* \leq 1, \ (\boldsymbol{d}^{\gamma,*} - d_{i^*}^\gamma \boldsymbol{e})' \boldsymbol{w^*} + d_{i^*}^\gamma \geq D^{\epsilon,\gamma} \right\}. \tag{21}$$

Polytope $W_{a_i a_j}^{\epsilon,\gamma *}$ is obtained upon performing a sequence of simple affine transformations to polytope $W_{a_i a_j}^{\epsilon,\gamma}$ (see the proof of Theorems 3 and 4). Using basic results from linear algebra (Lang [23]) I arrive at the following Theorem.

Theorem 4 *Let $i^* \in \mathcal{I}$ such that $d_{i^*}^{\gamma} \geq D^{\epsilon,\gamma}$, and consider polytope $W_{a_i a_j}^{\epsilon,\gamma,*}$ given by (21). $V_{ij}^{\epsilon,\gamma}$ defined in Eq. (18) satisfies*

$$V_{ij}^{\epsilon,\gamma} = (I-1)!\, Vol\left(W_{a_i a_j}^{\epsilon,\gamma *}\right)$$

Proof. See Appendix. ■

In light of Theorem 4, the main challenge now lies in calculating the volume of polytope $W_{a_i a_j}^{\epsilon,\gamma *}$. To this end, I make the following assumption.

Assumption 1 *There does not exist $i \in \mathcal{I}^*$ such that $\boldsymbol{d}_i^{\gamma *} = D^{\epsilon,\gamma}$.*

Assumption 1 ensures that polytope $W_{a_i a_j}^{\epsilon,\gamma *}$ is simple, i.e., that all of its vertices are nondegenerate. Primal nondegeneracy is a desirable property in linear programming as it facilitates the application of the simplex method (see Chapter 2 in Bertsimas and Tsitsiklis [6]). This enables the direct use of Lawrence's [24] formula for computing the volume of a polytope.

Remark 1. Assumption 1 is not strictly necessary for the implications of the analysis (i.e., the following Proposition 2) to hold. As Lawrence himself notes in his paper's conclusion [24], his method can be extended to non-simple polytopes using standard linear programming techniques (see Bueler et al. [9] for further discussions and applications). I choose to impose Assumption 1 because it can be always easily satisfied by a slight perturbation of $\bar{\boldsymbol{w}}$ or ϵ, while simplifying computations.

Proposition 2 *Suppose Assumption 1 holds. Then polytope $W_{a_i a_j}^{\epsilon,\gamma *}$ is simple. Moreover it has $O(I^2)$ vertices and $O(I^3)$ edges that can be readily identified (Eqs. (V1)-(V4) and (E1)-(E8) in Appendix). Using this information we can construct a vector $\boldsymbol{c} \in \Re^{I-1}$ such that the function $h(\boldsymbol{w}^*) = \boldsymbol{c}'\boldsymbol{w}^*$ is non-constant on each edge of $W_{a_i a_j}^{\epsilon,\gamma *}$ (Eq. (33) in Appendix). As a result, the volume of polytope $W_{a_i a_j}^{\epsilon,\gamma *}$ can be computed efficiently using the formula in Theorem 1 of Lawrence [24] (Eq. (34) in Appendix).*

Proof. See Appendix. ■

Thus, by Proposition 2 we have an efficient method of computing $V_{ij}^{\epsilon,\gamma}$ for any two alternatives a_i and a_j, and choice $\epsilon \in (0,1]$ and $\gamma \in \Re$. Examples 2 and 3 illustrate the two different cases of Theorem 3.

Example 2. One of the strengths of the proposed framework is that it sheds light on subtle differences among alternatives. This is illustrated by the following example:

$$\boldsymbol{x}_{\boldsymbol{a_i}} = (0.1, 0.2, 0.3, 0.4)', \quad \boldsymbol{x}_{\boldsymbol{a_j}} = (0.4, 0.3, 0.2, 0.15)', \quad \bar{\boldsymbol{w}} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)', \quad \gamma = 1.$$

Looking at just the ordinal dimension of the indicator data, we see that both a_i and a_j dominate in exactly two dimensions. Moreover, the difference in the composite scores under zero imprecision is relatively small: 0.2500 for a_i and 0.2625 for a_j. This may lead us to think that the two alternatives are roughly equal, and remain all the more so when we take weight imprecision into account. However, this is not the case. Indeed, we see that in reality a_j fares significantly better than a_i if we allow for weight imprecision (and our model makes quantitatively precise to what degree this is so), especially if we only wish to consider small levels of ϵ:

$$V_{ij}^{0.1,1} = 0.090, \;\; V_{ij}^{0.25,1} = 0.312, \;\; V_{ij}^{0.5,1} = 0.408, \;\; V_{ij}^{1,1} = 0.458.$$

Example 3. I provide an example of the situation discussed in part (ii) of Theorem 3. Consider:

$$\boldsymbol{x}_{\boldsymbol{a}_i} = (0.1, 0.2, 0.35, 0.4, 0.5)', \;\; \boldsymbol{x}_{\boldsymbol{a}_j} = (0.35, 0.4, 0.5, 0.1, 0.2)', \;\; \bar{\boldsymbol{w}} = \left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)', \;\; \gamma = 1.$$

Indeed, $\boldsymbol{x}_{\boldsymbol{a}_j}$ is obtained by simply permuting the elements of $\boldsymbol{x}_{\boldsymbol{a}_i}$. This instance yields

$$\bar{\boldsymbol{w}}'\boldsymbol{x}_{\boldsymbol{a}_i} = \bar{\boldsymbol{w}}'\boldsymbol{x}_{\boldsymbol{a}_j}, \;\; V_{ij}^{\epsilon,1} = 0.478, \;\; \forall \epsilon \in (0, 1].$$

This example further suggests how the *distribution* of the achievement scores across indicators can be very important in determining the dominance relation between alternatives.

How could ϵ be set? I close Section 5 by offering a few thoughts on the calibration of the parameter ϵ. From a general statistical standpoint, ϵ may be interpreted as the amount of error attached to the prior $\bar{\boldsymbol{w}}$ (Berger and Berliner [8]). In the context of composite indices of welfare, the choice of ϵ is largely subjective and should, in theory, be set by the policy makers who will actually use the composite index. Nevertheless, the simple structure of ϵ-contamination may inform this process by shedding light on the implications of different choices. First, it is clear that restricting weights to lie in W^ϵ places a lower bound of $(1-\epsilon)$ on all ratios $\frac{w_i}{\bar{w}_i}$, implying the following upper and lower bounds on w_i:

$$\begin{aligned} & \frac{w_i}{\bar{w}_i} \geq 1 - \epsilon, \;\; \forall i \in \mathcal{I}, \\ \Leftrightarrow \;\; & w_i \in \left[\bar{w}_i - \epsilon \bar{w}_i, \bar{w}_i + \epsilon(1 - \bar{w}_i)\right], \;\; \forall i \in \mathcal{I}. \end{aligned} \tag{22}$$

Less obviously, a choice of ϵ implies the following relationship between the volumes (in $(I-1)$-dimensional space) of W^ϵ and the standard $(I-1)$-simplex Δ^{I-1}:

$$\frac{Vol\,(W^\epsilon)}{Vol\,(\Delta^{I-1})} = \epsilon^{I-1}. \tag{23}$$

Eq. (23) can be easily established and the relation it implies is independent of the vector $\bar{\boldsymbol{w}}$ (see Eqs. (26)-(27) in the proof of Theorem 4). It could possibly serve as a guide for policy makers who wish to "cover" a target percentage of all possible vectors of weights, while respecting the bounds of Eq. (22).

6 Numerical Application

In this section, I apply Algorithm 1 to Shanghai University's Academic Ranking of World Universities (ARWU), a popular composite index measuring research excellence in academic institutions.[11] The ARWU Index measures academic excellence through the following six criteria: (1) number of alumni winning Nobel prizes or Fields medals, (2) number of faculty winning Nobel prizes or Fields medals; (3) highly-cited researchers in 21 broad categories; (4) papers published in Nature or Science; (5) papers indexed in Science Citation Index-expanded and Social Science Citation Index; (6) per capita academic performance of an institution. Indicator 1 is meant to capture quality of education, indicators 2 and 3 the quality of the faculty, 4 and 5 research output, and 6 academic performance. For each indicator, the highest scoring institution is assigned a score of 100, and other institutions are calculated as a percentage of the top score.

University scores are computed via a simple linear average, that is $u(\boldsymbol{x}, \boldsymbol{w}) = \sum_{i=1}^{6} w_i x_i$. Hence, the relevant welfare function is given by Eq. (12) with $\gamma = 1$, i.e., u^1. The following vector of weights is employed: $\boldsymbol{w}' = (.1, .2, .2, .2, .2, .1)$, and university rankings are based on these composite scores. Despite its increasing influence and popularity, the ARWU index has been criticized on many grounds, including its non-robustness to changes in weights (Saisana et al. [34]).

I proceed to apply the analysis of Section 5 and consider imprecision over the ARWU index weights via the $\epsilon-$contamination framework of Eq. (13) with $\bar{\boldsymbol{w}}$ equal to the vector of weights used by the developers of the ARWU index. Focusing on the top-100 universities reported in the 2013 ARWU rankings, denoted by $\mathcal{A}_{100}$, I consider three values for ϵ that reflect plausible departures from the initial weights $\bar{\boldsymbol{w}}$, namely $\epsilon \in \{1/6, 1/3, 1/2\}$. When $\epsilon = 0$ and there is no imprecision over weights, the (partial) ranking of the 100 first universities is denoted by K^0. Unlike the designers of the ARWU index, I do not round university scores to 1 decimal point. This means that K^0 has fewer ties than the partial ranking reported on the ARWU website.

I implement Algorithm 1 in Matlab with a choice of $M = 10$. Step 1 is performed

[11] See www.shanghairanking.com

using the insights of Lawrence's formula via Proposition 2.[12] In the case of Step 3, I use a Matlab implementation of Tarjan's algorithm due to David F. Gleich of Stanford University (Copyright).[13] For Step 4b, I coded the Van Zuylen and Williamson algorithm myself (Step 4c was not applicable, due to Corollary 1).

The second column of Table 1 shows the number of non-singleton strongly connected components for the corresponding L-dominance graphs, while the third a vector indicating the number of universities that each such strongly connected component includes. We see that as ϵ grows and we enlarge the set of weights W^ϵ under consideration, the Condorcet paradox becomes more and more pronounced. As the size of the strongly connected components is never greater than $M = 10$, Step 4a of Algorithm 1 was always applicable. Thus, we can definitely conclude that Algorithm 1's output are the Kemeny-optimal rankings of the three problem instances.

ϵ	Number	Sizes
1/6	1	5
1/3	3	(4,8,3)
1/2	6	(4,9,3,6,5,3)

Table 1: 2013 ARWU rankings: Number and sizes of non-singleton strongly connected components for L-dominance graphs G^L, where $L = (\boldsymbol{X}_{\mathcal{A}_{100}}, f^\epsilon, u^1)$.

For convenience, denote by K^ϵ the Kemeny-optimal ranking of universities in U_{100} when applying the method for different values of ϵ. Figure 2 shows the initial ARWU scores for the top-100 universities, as well as their rankings under K^0 and the Kemeny-optimal rankings K^ϵ for $\epsilon \in \{1/6, 1/3, 1/2\}$. Figure 3 depicts $K^0(a) - K^\epsilon(a)$ for all $a \in \mathcal{A}_{100}$ and $\epsilon \in \{1/6, 1/3, 1/2\}$. We see that these differences grow as we increase ϵ, and are much more pronounced for universities in the 51-100 range. There are moreover a handful of really substantial swings in rankings. For instance, the Ecole Normale Superieure-Paris was ranked 71st in the official 2013 ARWU ranking, whereas its Kemeny-optimal ranks for $\epsilon = 1/6, 1/3, 1/2$ are are 62, 54, and 49, respectively. This discrepancy seems to be due to its extremely good performance in Indicator 6.

On a final note, to test the efficacy of the VZW approximation algorithm, I reran Algorithm 1 using $M = 3$ and compared the output rankings to the known Kemeny-optimal ones. The VZW rankings for $\epsilon = 1/6$ and $1/3$ are completely unchanged. When

[12] For stability, where necessary I truncate the constant C in Eq. (33) to not exceed 10.

[13] This code can be found at http://www.mathworks.it/matlabcentral/fileexchange/24134-gaimc-graph-algorithms-in-matlab-code/content/gaimc/scomponents.m

$\epsilon = 1/2$, the only difference between the VZW and $K^{1/2}$ rankings is that the University of Western Australia (alternative a_{91}) and Case Western University (alternative a_{99}) switch their Kemeny-optimal ranks of 91 and 92 respectively. Thus the VZW ranking is not Kemeny-optimal, since $V^{\frac{1}{2},1}_{91,99} = .5087$ and flipping the two ranks results in an improved objective function value. (Note, however, that this error could have been corrected by using the local Kemenization heuristic of Dwork et al. [15]).

7 Directions for Future Research

Judgments based on composite indices of welfare depend, sometimes critically, on how different dimensions of performance are weighted. As there is frequently no single "right" way to assign such weights, it is important to take this imprecision into account in a systematic and transparent manner. In this paper I have drawn from the theory of social choice to present a procedure for determining a ranking of the relevant alternatives that is normatively compelling and statistically interpretable. Special attention to issues of practicality and computational tractability has also been given. The applicability of the proposed framework was illustrated through a numerical example based on Shanghai University's ARWU index.

This work suggests fruitful avenues for future research. An immediate one involves investigating whether the characterization of Theorem 1 can be extended to the full profile domain $\mathcal{L}$. In particular, this would entail extending the results of Young and Levenglick [43] to real-valued election matrices. Alternatively, one could introduce random noise to achievement vectors and attempt to use the statistical interpretation of Kemeny's rule as a maximum likelihood estimate [41] to derive significance tests along the lines of Duclos et al. [13]. Finally, one could also undertake more complex empirical applications of the model, like for instance those corresponding to discontinuous welfare functions that are often found in poverty indices (e.g. [13, 2]).

Broader connections with decision-theoretic models of Knightian uncertainty could also be explored. This is because, as mentioned in the introduction, the examination of welfare indices under weight imprecision is, mathematically if not conceptually, analogous to certain settings of decision-making under ambiguity (see Gilboa and Marinacci [18]).[14] Perhaps the work presented here, with its connections to classical voting theory, could be of relevance to this vibrant field of economic theory.

[14]The conceptual difference is that in the former we have uncertain weight vectors over dimensions of well-being, while in the latter uncertain Bayesian priors over states of nature.

Appendix

A1: Proofs

Proposition 1. Consider case (a) first. Since g is invertible,

$$W_{a_i a_j}^{\boldsymbol{X}_{\mathcal{A}},u} \cap W_{a_j a_i}^{\boldsymbol{X}_{\mathcal{A}},u} = \left\{ \boldsymbol{w} \in \Delta^{I-1} : \ \sum_{i=1}^{I} w_i u_i(x_{a_i i}) = \sum_{i=1}^{I} w_i u_i(x_{a_j i}) \right\}.$$

As the functions u_i are strictly monotone, unless $\boldsymbol{x}_{\boldsymbol{a_i}} = \boldsymbol{x}_{\boldsymbol{a_j}}$ the above set will be a polytope of dimension $I-2$. Thus, $W_{a_i a_j}^{\boldsymbol{X}_{\mathcal{A}},u} \cap W_{a_j a_i}^{\boldsymbol{X}_{\mathcal{A}},u}$ will have zero Lebesgue measure in Δ^{I-1}. The result follows.

Case (b) is exactly similar, with the only difference that the intersection of $W_{a_i a_j}^{\boldsymbol{X}_{\mathcal{A}},u}$ and $W_{a_j a_i}^{\boldsymbol{X}_{\mathcal{A}},u}$ reduces to:

$$W_{a_i a_j}^{\boldsymbol{X}_{\mathcal{A}},u} \cap W_{a_j a_i}^{\boldsymbol{X}_{\mathcal{A}},u} = \left\{ \boldsymbol{w} \in \Delta^{I-1} : \ \sum_{i=1}^{I} w_i \log\left(u_i(x_{a_i i})\right) = \sum_{i=1}^{I} w_i \log\left(u_i(x_{a_j i})\right) \right\}.$$

■

Theorem 1. (i) By Eqs. (4) and (7), it is clear that K satisfies all six properties.
(ii) First, we prove the following lemma that will be needed later on.

Lemma 1 *For all subsets $\mathcal{L}'$ of $\tilde{\mathcal{L}}$, define the set $\mathcal{Y}^{\mathcal{L}'} = \{\boldsymbol{Y}^L : \ L \in \mathcal{L}'\}$. Recall restriction (8) and let $\tilde{\mathcal{L}}^Q = \{L \in \tilde{\mathcal{L}} : \ Y^L \in \mathcal{Y}^Q\}$. We have $\mathcal{Y}^{\tilde{\mathcal{L}}^Q} = \mathcal{Y}^Q = \mathcal{Y}^{\mathcal{L}^Q}$.*

Proof. First of all, it is clear by the definition of Eq. (3) that

$$\mathcal{Y}^{\tilde{\mathcal{L}}^Q} \subset \mathcal{Y}^Q.$$

Now, consider the welfare function $u^*(\boldsymbol{x}, \boldsymbol{w}) = \sum_{i=1}^{I} w_i x_i$. Moreover, consider the achievement matrix $\boldsymbol{X}_{\mathcal{A}}^* = \boldsymbol{I}_A$ where $\boldsymbol{I}_A$ is the $A \times A$ identity matrix.

For each $R \in \mathcal{R}_{\mathcal{A}}$, define the set $W_R^* = \{\boldsymbol{w} \in \Delta^{I-1} : R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}}^*,u^*} = R\}$. Letting $\hat{a}_{R(i)} = \{a \in \mathcal{A} : R(a) = i\}$, it is easy to see that $W_R^* = \{\boldsymbol{w} \in \Delta^{I-1} : w_{\hat{a}_{R(1)}} > w_{\hat{a}_{R(2)}} > ... > w_{\hat{a}_{R(A)}}\}$.
The sets W_R^* are symmetric in Δ^{I-1} and mutually exclusive. Moreover, u^* satisfies the assumption of Proposition 1. Indeed, we have

$$\int_{W_R^*} \mathrm{d}\boldsymbol{w} = \int_{\Delta^{I-1}} 1\left\{ R_{\boldsymbol{w}}^{\boldsymbol{X}_{\mathcal{A}}^*,u^*} = R \right\} \mathrm{d}\boldsymbol{w} = \frac{1}{A!}, \ \ \forall R \in \mathcal{R}_{\mathcal{A}}. \tag{24}$$

Now, writing $\mathcal{R}_{\mathcal{A}} = \{R_1, R_2, ..., R_{A!}\}$, consider the family of importance functions $\mathcal{F}^*$, where (Q_+^I is the set of non-negative rational I-dimensional vectors)

$$\mathcal{F}^* = \left\{ f : \Delta^{I-1} \mapsto \Re_+ : \ \ f(\boldsymbol{w}) = z_i, \ \forall \boldsymbol{w} \in W_{R_i}^*, \forall i = 1, 2, ..., A!, \ \ \boldsymbol{z} \in Q_+^I, \ \sum_{i=1}^{A!} z_i > 0 \right\}.$$

Define the set of profiles $\mathcal{L}^* = \{L \in \mathcal{L} : \; \boldsymbol{X}_\mathcal{A} = \boldsymbol{X}^*_\mathcal{A}, \; f \in \mathcal{F}^*, \; u = u^*\}$. By Eq. (24) and the structure of the importance functions in $\mathcal{F}^*$, we may conclude

$$\mathcal{Y}^{\mathcal{L}^*} = \mathcal{Y}^Q.$$

Thus, since $\mathcal{L}^* \subset \tilde{\mathcal{L}}^Q$, we have $\mathcal{Y}^Q \subset \mathcal{Y}^{\tilde{\mathcal{L}}^Q}$ implying $\mathcal{Y}^Q = \mathcal{Y}^{\tilde{\mathcal{L}}^Q}$. That $\mathcal{Y}^Q = \mathcal{Y}^{\mathcal{L}^Q}$ follows from the fact that $\mathcal{Y}^{\mathcal{L}^Q} \subset \mathcal{Y}^Q$ and $\mathcal{Y}^Q = \mathcal{Y}^{\tilde{\mathcal{L}}^Q} \subset \mathcal{Y}^{\mathcal{L}^Q}$. ■

Let us now consider a rule ϕ satisfying the stated properties and recall the constructed electorate $\mathcal{E}(f)$ with its associated partial input rankings and election matrices. If there are just two alternatives, then ϕ must be simple majority rule and hence equal to K (Young [40], footnote 18 in Young [41]). Note that partial input rankings do not create any problems for this result to go through (see page 51 in Young [40]). When there are more than two alternatives, LIIA together with the previous majority rule for all profiles of two alternatives imply that the rule must satisfy the Condorcet property of Young and Levenglick [43] (see footnote 18 in Young [41]). As a result, Lemma 1 in Young and Levenglick [43] implies that for every $L \in \mathcal{L}$, $\phi(L)$ depends only on the election matrix Y^L. Again, partial input rankings do not affect the proof of this result. Hence, the characterization of K may be established on the domain $Y^{\mathcal{L}^Q}$ (instead of $\mathcal{L}^Q$), which, by Lemma 1 above, is equal to $\mathcal{Y}^Q$. The latter is the same domain on which Young and Levenglick [43] prove their characterization of Kemeny's rule (see their Lemma 1 on page 292 and the discussion immediately preceding and following it). Hence, $\phi = K$ follows from Young and Levenglick [43] and the argument delineated in footnote 18 of Young [41].

■

Theorem 2. Recall that $L = (\boldsymbol{X}_\mathcal{A}, f, u)$. Suppose graph G^L has M strongly connected components denoted by $G_1 = (\mathcal{A}_1, E_1), G_2 = (\mathcal{A}_j, E_2), , ..., G_M = (\mathcal{A}_M, E_M)$.

For each $m = 1, 2, ..., M$ define $g_m \equiv \frac{in(\mathcal{A}_m, \mathcal{A} \backslash \mathcal{A}_m)}{|\mathcal{A}_m|}$. Now for all pairs (G_i, G_j) either all vertices in G_i are pointing to all vertices in G_j, or vice versa. For, if this were not true, it would be possible to form a cycle that included vertices in both G_i and G_j contradicting the fact that they are separate strongly connected components. If all vertices in G_i are pointing to all vertices in G_j, then they are also pointing to all of the vertices G_j is pointing to (if not there would exist a cycle), implying $g_i < g_j$. Conversely, if all vertices in G_j are pointing to all vertices in G_i then $g_i > g_j$. Thus, we may deduce that $g_i < g_j$ if and only if all vertices in G_i are pointing to all vertices in G_j.

The above implies that it is possible to order the g_i's in strictly ascending order and we may assume, without loss of generality, that $g_1 < g_2 < .... < g_M$. With this ordering in mind, define the sets $\mathcal{A}_m^- = \mathcal{A}_1 \cup \mathcal{A}_2 \cup ... \cup \mathcal{A}_m$ and $\mathcal{A}_m^+ = \mathcal{A}_m \cup \mathcal{A}_{m+1} \cup ... \cup \mathcal{A}_M$, for all $m = 1, 2, ..., M$.

We proceed by proving that K^L must be such that, for all $m = 1, 2, ...M - 1$, all alternatives in $\mathcal{A}_m^-$ are ranked before all alternatives in $\mathcal{A}_{m+1}^+$. To wit, consider the alternatives in $\mathcal{A}_m^-$. Since $g_i < g_j$ for all (i, j) such that $i \leq m$ and $j > m$, we know that all alternatives in $\mathcal{A}_m^-$ are pointing to all alternatives in $\mathcal{A}_{m+1}^+$. Hence $Y_{ij}^L > 0$ for all pairs $(a_i, a_j) \in \mathcal{A}_m^- \times \mathcal{A}_{m+1}^+$. Thus, since K satisfies the extended-Condorcet property, any Kemeny-optimal ranking will have to rank all alternatives in $\mathcal{A}_m^-$ before all alternatives in $\mathcal{A}_{m+1}^+$.

Now, note that by the definition of G^L we have $|\mathcal{A}_m^-| = \frac{in(\mathcal{A}_{m+1}, \mathcal{A}\backslash\mathcal{A}_{m+1})}{|\mathcal{A}_{m+1}|}$. Thus, we obtain the following relation for all $R \in K^L$:

$$R(a) \in \left\{ \frac{in(\mathcal{A}_m, \mathcal{A}\backslash\mathcal{A}_m)}{|\mathcal{A}_m|} + 1, \ \frac{in(\mathcal{A}_m, \mathcal{A}\backslash\mathcal{A}_m)}{|\mathcal{A}_m|} + |\mathcal{A}_m| \right\}, \ \forall a \in \mathcal{A}_m, \ \ m = 1, 2, ..., M. \quad (25)$$

To determine the precise rank of the alternatives in each $\mathcal{A}_m$ within the ranges given by Eq.(25), it is sufficient, by LIIA, to solve for $K^{L_{\mathcal{A}_m}}$, and update accordingly. ■

Theorem 3. Let us first concentrate on the denominator of (18). We perform the following two operations on the elements of W^ϵ: (a) we translate them by $-(1-\epsilon)\bar{\boldsymbol{w}}$, and then (b) multiply them by $1/\epsilon$. The resulting polytope is Δ^{I-1}, the standard $(I-1)$-simplex. The volume of the standard simplex Δ^{I-1} (which has a side length of $\sqrt{2}$) is given by

$$Vol\left(\Delta^{I-1}\right) = \frac{\sqrt{2}^{I-1}\sqrt{I}}{(I-1)!\sqrt{2^{I-1}}} = \frac{\sqrt{I}}{(I-1)!} \quad (26)$$

Basic linear algebra (see Lang [23]) implies that

$$Vol(W^\epsilon) = \sqrt{\epsilon^{2(I-1)}} Vol\left(\Delta^{I-1}\right) = \epsilon^{I-1}\frac{\sqrt{I}}{(I-1)!}. \quad (27)$$

Now let us focus on the numerator. Recall the difference vector $\boldsymbol{d}^\gamma$ and the constant $D^{\epsilon,\gamma} = -\frac{1-\epsilon}{\epsilon}(\boldsymbol{d}^\gamma)'\bar{\boldsymbol{w}}$, defined in Eq. (20). Performing the same affine transformation as before, namely $\boldsymbol{w} \leftarrow \frac{\boldsymbol{w}-(1-\epsilon)\bar{\boldsymbol{w}}}{\epsilon}$, the polytope $W_{a_i a_j}^{\epsilon,\gamma}$ is transformed into

$$\widehat{W}_{a_i a_j}^{\epsilon,\gamma} = \left\{\widehat{\boldsymbol{w}} \in \Delta^{I-1} : \ \ (\boldsymbol{d}^\gamma)'\widehat{\boldsymbol{w}} \geq D^{\epsilon,\gamma}\right\}, \quad (28)$$

which in turn implies

$$Vol\left(W_{a_i a_j}^{\epsilon,\gamma}\right) = \epsilon^{I-1} Vol\left(\widehat{W}_{a_i a_j}^{\epsilon,\gamma}\right). \quad (29)$$

Putting Eqs. (27) and (29) together, we obtain

$$V_{ij}^{\epsilon,\gamma} = \frac{Vol\left(W_{a_i a_j}^{\epsilon,\gamma}\right)}{Vol\left(W^\epsilon\right)} = \frac{(I-1)!}{\sqrt{I}} Vol\left(\widehat{W}_{a_i a_j}^{\epsilon,\gamma}\right). \quad (30)$$

Eqs (20) and (28) imply that the volume of $\widehat{W}^{\epsilon,\gamma}_{a_i a_j}$ is increasing in ϵ if $(\boldsymbol{d}^\gamma)'\bar{\boldsymbol{w}} < 0$, decreasing if $(\boldsymbol{d}^\gamma)'\bar{\boldsymbol{w}} = 0$, and constant if $(\boldsymbol{d}^\gamma)'\bar{\boldsymbol{w}} = 0$. The result now follows from Eq. (30). ■

Theorem 4. Assume without loss of generality that $i^* = I$. Consider the polytope $W^{\epsilon,\gamma,*}_{a_i a_j}$ of Eq. (21), obtained by using the equality constraint $\sum_{i=1}^{I} w_1 = 1$ to eliminate variable I from polytope $\widehat{W}^{\epsilon,\gamma}$:

$$W^{\epsilon,\gamma,*}_{a_i a_j} = \left\{ \boldsymbol{w^*} \in \Re^{I-1} : \ \boldsymbol{w^*} \geq 0, \ \sum_{i \in \mathcal{I}^*} w^*_i \leq 1, \ (\boldsymbol{d^*} - d^\gamma_{i^*}\boldsymbol{e})'\boldsymbol{w^*} + d^\gamma_{i^*} \geq D^{\epsilon,\gamma} \right\}.$$

The affine transformation h which maps polytope $W^{\epsilon,\gamma,*}_{a_i a_j}$ to $\widehat{W}^{\epsilon,\gamma}_{a_i a_j}$ is given by $h : \Re^{I-1} \to \Re^I$, satisfying $h(\boldsymbol{w^*}) = T \cdot \boldsymbol{w^*} + [0, 0, ..., 0, 1]'$, where T is an $I \times (I-1)$ matrix equal to:

$$T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & & 1 \\ -1 & -1 & -1 & \cdots & -1 \end{bmatrix} \Rightarrow T' \cdot T = \begin{bmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ \vdots & & \ddots & & \\ 1 & 1 & \cdots & 2 & 1 \\ 1 & 1 & 1 & \cdots & 2 \end{bmatrix}$$

Thus we have $\det[T' \cdot T] = I$. Once again, basic linear algebra [23] implies that

$$Vol\left(\widehat{W}^{\epsilon,\gamma}_{a_i a_j}\right) = \sqrt{\det[T' \cdot T]} \cdot Vol\left(W^{\epsilon,\gamma,*}_{a_i a_j}\right) = \sqrt{I} \cdot Vol\left(W^{\epsilon,\gamma,*}_{a_i a_j}\right). \tag{31}$$

Eqs. (30)-(31) together imply

$$V^{\epsilon,\gamma}_{a_i a_j} = (I-1)! \, Vol\left(W^{\epsilon,\gamma,*}_{a_i a_j}\right).$$

■

Proposition 2. We first identify the vertices of polytope $W^{\epsilon,\gamma*}_{a_i a_j}$. In doing so, we divide the set of indicators $\mathcal{I}^* = \{1, 2, ..., I-1\}$ into $\mathcal{I}^*_1$ and $\mathcal{I}^*_2$, such that

$$\mathcal{I}^*_1 = \{i \in \mathcal{I}^* : \ \boldsymbol{d}^{\gamma*}_i > D^{\epsilon,\gamma}\}, \ \mathcal{I}^*_2 = \{i \in \mathcal{I}^* : \ \boldsymbol{d}^{\gamma*}_i < D^{\epsilon,\gamma}\}.$$

Assumption 1 ensures that $\{\mathcal{I}^*_1, \mathcal{I}^*_2\}$ is a partition of $\mathcal{I}^*$, and we let $I^*_1 = |\mathcal{I}^*_1|$ and $I^*_2 = |\mathcal{I}^*_2|$. It will be useful to express polytope $W^{\epsilon,\gamma*}_{a_i a_j}$ in the following way:

$$W^{\epsilon,\gamma,*}_{a_i a_j} = \{w^* \in \Re^{I-1} : \ \boldsymbol{y}'_{\boldsymbol{k}}\boldsymbol{w^*} \leq \boldsymbol{b}\}, \tag{32}$$

where the $(I-1)$-dimensional vectors $\boldsymbol{y_k}$, $k = 1, 2, ..., I+1$, and $\boldsymbol{b}$ satisfy (a) $\boldsymbol{y_k} = -\boldsymbol{e_k}$[15] and $\boldsymbol{b}_k = 0$ for $k = 1, ..., I-1$, (b) $\boldsymbol{y_I} = [1, 1, 1, ..., 1]'$ and $\boldsymbol{b}_I = 1$, and (c) $\boldsymbol{y_{I+1}} = -\boldsymbol{d}^{\gamma*} + d^\gamma_{i^*}\boldsymbol{e}$ and $\boldsymbol{b}_{I+1} = -D^{\epsilon,\gamma} + d^\gamma_{i^*}$.

[15] Here $\boldsymbol{e_k}$ denotes the corresponding standard basis vector in $\Re^{I-1}$.

With representation (32) in mind, a vector $\boldsymbol{v}$ is a vertex of $W^{\epsilon,\gamma*}_{a_ia_j}$ if it satisfies $I-1$ linearly independent inequality constraints with equality [6]. The structure of vectors $\boldsymbol{y_k}$ for $k=1,2,...,I+1$ and $\boldsymbol{b}$ imply that a vertex of $W^{\epsilon,\gamma*}_{a_ia_j}$ can have at most 2 nonzero entries. Furthermore, Assumption 1 ensures primal nondegeneracy so that every vertex $\boldsymbol{v}$ will correspond to a unique basis matrix B_v, i.e., a unique set of linearly independent constraints satisfied with equality.

We may distinguish between four kinds of vertices and their corresponding bases:

(V1) $\boldsymbol{v_0}=\boldsymbol{0}$. $B_0=\{\boldsymbol{y_k'}:k=1,2,..,I-1\}$

(V2) $\boldsymbol{v_i}=\boldsymbol{e_i}$ for all $i\in\mathcal{I}_1^*$. Here $B_i=\{\boldsymbol{y_k'}:k=1,...,i-1,i+1,..,I-1,I\}$, for all $i\in\mathcal{I}_1^*$.

(V3) $\boldsymbol{v_j}=\pi_j\boldsymbol{e}_j$, where $\pi_j=\frac{d^{\gamma}_{i^*}-D^{\epsilon,\gamma}}{d^{\gamma}_{i^*}-\boldsymbol{d}^{\gamma*}_j}$, for all $j\in\mathcal{I}_2^*$. Here $B_j=\{\boldsymbol{y_k'}:k=1,...,j-1,j+1,..,I-1,I+1\}$, for all $j\in\mathcal{I}_2^*$.

(V4) $\boldsymbol{v_{ij}}=\pi_{ij}\boldsymbol{e}_i+(1-\pi_{ij})\boldsymbol{e}_j$, where $\pi_{ij}=\frac{D^{\epsilon,\gamma}-\boldsymbol{d}^{\gamma*}_j}{\boldsymbol{d}^{\gamma*}_i-\boldsymbol{d}^{\gamma*}_j}$, for all $i\in\mathcal{I}_1^*$ and $j\in\mathcal{I}_2^*$. Here $B_{ij}=\{\boldsymbol{y_k'}:k=1,...,i-1,i+1,..,j-1,j+1,...,I,I+1\}$, for all $i\in\mathcal{I}_1^*$ and $j\in\mathcal{I}_2^*$.

Two vertices are connected by an edge if they share $I-2$ common linearly independent active constraints [6]. An examination of the preceding expressions for the vertices of $W^{\epsilon,\gamma,*}_{a_ia_j}$ and their bases, implies that we may identify the following eight kinds of undirected edges:

(E1) $(\boldsymbol{v_0},\boldsymbol{v_i})$, for all $i\in\mathcal{I}_1^*$.

(E2) $(\boldsymbol{v_0},\boldsymbol{v_j})$, for all $j\in\mathcal{I}_2^*$.

(E3) $(\boldsymbol{v_i},\boldsymbol{v_k})$ for all pairs (i,k) where $i,k\in\mathcal{I}_1^*$.

(E4) $(\boldsymbol{v_j},\boldsymbol{v_k})$ for all pairs (j,k) where $j,k\in\mathcal{I}_2^*$.

(E5) $(\boldsymbol{v_i},\boldsymbol{v_{ij}})$ for all pairs (i,j) where $i\in\mathcal{I}_1^*$ and $j\in\mathcal{I}_2^*$.

(E6) $(\boldsymbol{v_j},\boldsymbol{v_{ij}})$ for all pairs (i,j) where $i\in\mathcal{I}_1^*$ and $j\in\mathcal{I}_2^*$.

(E7) $(\boldsymbol{v_{ij}},\boldsymbol{v_{ik}})$ for all triplets (i,j,k) where $i\in\mathcal{I}_1^*$ and $j,k\in\mathcal{I}_2^*$.

(E8) $(\boldsymbol{v_{ij}},\boldsymbol{v_{kj}})$ for all triplets (i,j,k) where $i,k\in\mathcal{I}_1^*$ and $j\in\mathcal{I}_2^*$.

Recall that we wish to exhibit a vector $\boldsymbol{c}\in\Re^{I-1}$ such that the function $h(\boldsymbol{w^*})=\boldsymbol{c}'\boldsymbol{w^*}$ is non-constant on each edge of $W^{\epsilon,\gamma*}_{a_ia_j}$. To this end, recall the vertices of $W^{\epsilon,\gamma*}_{a_ia_j}$ enumerated above as (V1)-(V4) and the values of π_j for $j \in \mathcal{I}_2^*$ and π_{ij} for all $i \in \mathcal{I}_1^*$ and $j \in \mathcal{I}_2^*$. Define the following two quantities

$$\delta_i = \min_{i\in\mathcal{I}_1^*,\ j,k\in\mathcal{I}_2^*} \{|\pi_{ij} - \pi_{ik}| : \pi_{ij} \neq \pi_{ik}\}. \text{ If undefined, set } \delta_i = 1.$$

$$\delta_j = \min_{i,k\in\mathcal{I}_1^*,\ j\in\mathcal{I}_2^*} \{|\pi_{ij} - \pi_{kj}| : \pi_{ij} \neq \pi_{kj}\}. \text{ If undefined, set } \delta_j = 1.$$

Consequently let $\delta = \min\{\delta_i, \delta_j\}$ and define $C = \frac{2}{\delta}$.

Finally, relabel the elements of set $\mathcal{I}_2^* = \{j_1, j_2, ..., j_{I_2^*}\}$ so that $\pi_{j_1} \leq \pi_{j_2} \leq ... \leq \pi_{j_{I_2^*}}$. Now, define the vector $\boldsymbol{c}$ satisfying

$$c_l \quad = \quad \begin{cases} C + \frac{k}{(I_1^*+1)}, & \text{if } \ l = i_k \in \mathcal{I}_1^*, \\ \frac{k}{I_2^*(I_1^*+1)}, & \text{if } \ l = j_k \in \mathcal{I}_2^*. \end{cases} \tag{33}$$

With this choice of $\boldsymbol{c}$ we can check all eight kinds of edges (E1)-(E8) and verify that $\boldsymbol{c}'\boldsymbol{v} \neq \boldsymbol{c}'\boldsymbol{u}$ for all pairs of adjacent vertices $(\boldsymbol{v}, \boldsymbol{u})$. Thus the function $f(\boldsymbol{w}^*) = \boldsymbol{c}'\boldsymbol{w}^*$ is non-constant on each edge of $W^{\epsilon,\gamma*}_{a_i a_j}$. Hence, in conjunction with Assumption 1, we may apply Theorem 1 in Lawrence [24] to conclude

$$Vol\left(W^{\epsilon,\gamma,*}_{a_i a_j}\right) = \sum_{\substack{\text{vertices } \boldsymbol{v} \\ \text{of } W^{\epsilon,\gamma*}_{a_i a_j}}} \frac{(\boldsymbol{c}'\boldsymbol{v})^{I-1}}{(I-1)!\,|\det(B_{\boldsymbol{v}})| \prod_{i=1}^{I-1} \left[(B'_{\boldsymbol{v}})^{-1}\boldsymbol{c}\right]_i}. \tag{34}$$

■

	Initial Score		K^0	K^1/6	K^1/3	K^1/2
Harvard University	100.00		1	1	1	1
Stanford University	72.58		2	2	2	3
University of California, Berkeley	71.27		3	3	4	4
Massachusetts Institute of Technology (MIT)	71.12		4	4	3	2
University of Cambridge	69.57		5	5	5	5
California Institute of Technology	62.87		6	6	6	6
Princeton University	61.92		7	7	7	7
Columbia University	59.76		8	8	8	8
University of Chicago	57.07		9	9	9	9
University of Oxford	55.92		10	10	10	10
Yale University	55.36		11	11	11	11
University of California, Los Angeles	52.89		12	12	12	12
Cornell University	50.02		13	13	13	13
University of California, San Diego	49.92		14	14	15	15
University of Pennsylvania	49.66		15	15	14	14
University of Washington	48.32		16	16	16	16
The Johns Hopkins University	46.88		17	17	17	17
University of California, San Francisco	46.14		18	18	18	18
University of Wisconsin - Madison	44.89		19	19	19	19
Swiss Federal Institute of Technology Zurich	43.47		20	20	20	20
University College London	43.02		21	21	22	22
The University of Tokyo	42.95		22	22	21	21
University of Michigan - Ann Arbor	42.62		23	23	23	23
The Imperial College of Science, Technology and Medicine	41.65		24	24	24	24
University of Illinois at Urbana-Champaign	41.07		25	25	25	25
Kyoto University	40.85		26	26	26	26
New York University	40.46		27	27	27	27
University of Toronto	40.34		28	28	28	28
University of Minnesota, Twin Cities	39.67		29	29	29	29
Northwestern University	38.84		30	30	30	30
Duke University	38.08		31	31	31	31
Washington University in St. Louis	37.53		32	32	32	33
University of Colorado at Boulder	37.32		33	33	34	34
Rockefeller University	37.12		34	34	33	32
University of California, Santa Barbara	35.88		35	35	35	35
The University of Texas at Austin	35.46		36	37	37	37
Pierre and Marie Curie University - Paris 6	35.28		37	36	36	36
University of Maryland, College Park	34.72		38	38	38	39
University of Paris Sud (Paris 11)	34.54		39	39	39	38
University of British Columbia	34.21		40	40	41	41
The University of Manchester	34.01		41	42	42	42
University of Copenhagen	33.78		42	41	40	40
University of North Carolina at Chapel Hill	33.69		43	43	43	44
Karolinska Institute	32.68		44	44	44	43
University of California, Irvine	32.38		45	45	45	45
The University of Texas Southwestern Medical Center at Dallas	31.41		46	46	46	48
University of California, Davis	31.34		47	47	49	51
University of Southern California	31.33		48	48	50	60
Vanderbilt University	30.99		49	50	51	55
Technical University Munich	30.59		50	49	47	46
The University of Edinburgh	30.51		51	52	52	50
Carnegie Mellon University	30.42		52	51	48	47
Utrecht University	30.33		53	53	53	54
Pennsylvania State University - University Park	30.18		54	58	60	59
University of Heidelberg	30.17		55	55	58	56
University of Melbourne	30.17		56	54	57	58
Purdue University - West Lafayette	30.09		57	59	61	61
McGill University	29.85		58	56	55	52
The Hebrew University of Jerusalem	29.78		59	57	56	53
University of Zurich	29.70		60	61	63	63
Rutgers, The State University of New Jersey - New Brunswick	29.51		61	64	64	65
University of Pittsburgh	29.50		62	63	62	62
University of Munich	29.48		63	60	59	57
University of Bristol	29.20		64	65	65	68
The Ohio State University - Columbus	28.97		65	69	69	70
The Australian National University	28.87		66	66	68	67
Brown University	28.78		67	67	66	66
King's College London	28.78		68	71	71	71
University of Oslo	28.72		69	70	70	69
University of Geneva	28.71		70	68	67	64
Ecole Normale Superieure - Paris	28.51		71	62	54	49
University of Florida	28.49		72	72	72	73
Uppsala University	28.03		73	73	74	74
Leiden University	27.78		74	74	73	72
Boston University	27.42		75	75	76	77
University of Helsinki	27.22		76	76	75	76
Technion-Israel Institute of Technology	26.57		77	77	78	78
University of Arizona	26.52		78	79	79	79
Moscow State University	26.12		79	78	77	75
Arizona State University - Tempe	26.05		80	82	86	89
Aarhus University	25.94		81	80	82	82
Stockholm University	25.81		82	81	80	81
University of Basel	25.59		83	83	81	80
University of Nottingham	25.57		84	87	87	88
Osaka University	25.57		85	86	85	87
Indiana University Bloomington	25.54		86	88	88	90
The University of Queensland	25.50		87	85	84	85
University of Utah	25.47		88	90	95	99
Ghent University	25.45		89	89	90	86
University of Rochester	25.40		90	84	83	83
The University of Western Australia	24.95		91	92	91	91
University of Groningen	24.93		92	95	98	97
Weizmann Institute of Science	24.90		93	93	94	94
McMaster University	24.89		94	94	97	96
Michigan State University	24.89		95	99	100	100
Rice University	24.86		96	91	89	84
University of Strasbourg	24.75		97	97	93	93
University of Sydney	24.74		98	98	96	95
Case Western Reserve University	24.58		99	96	92	92
University of Freiburg	24.32		100	100	99	98

Figure 2: 2013 ARWU Top-100: Initial scores and ranking K^0, and Kemeny-optimal rankings K^ϵ for $\epsilon \in \{1/6, 1/3, 1/2\}$.

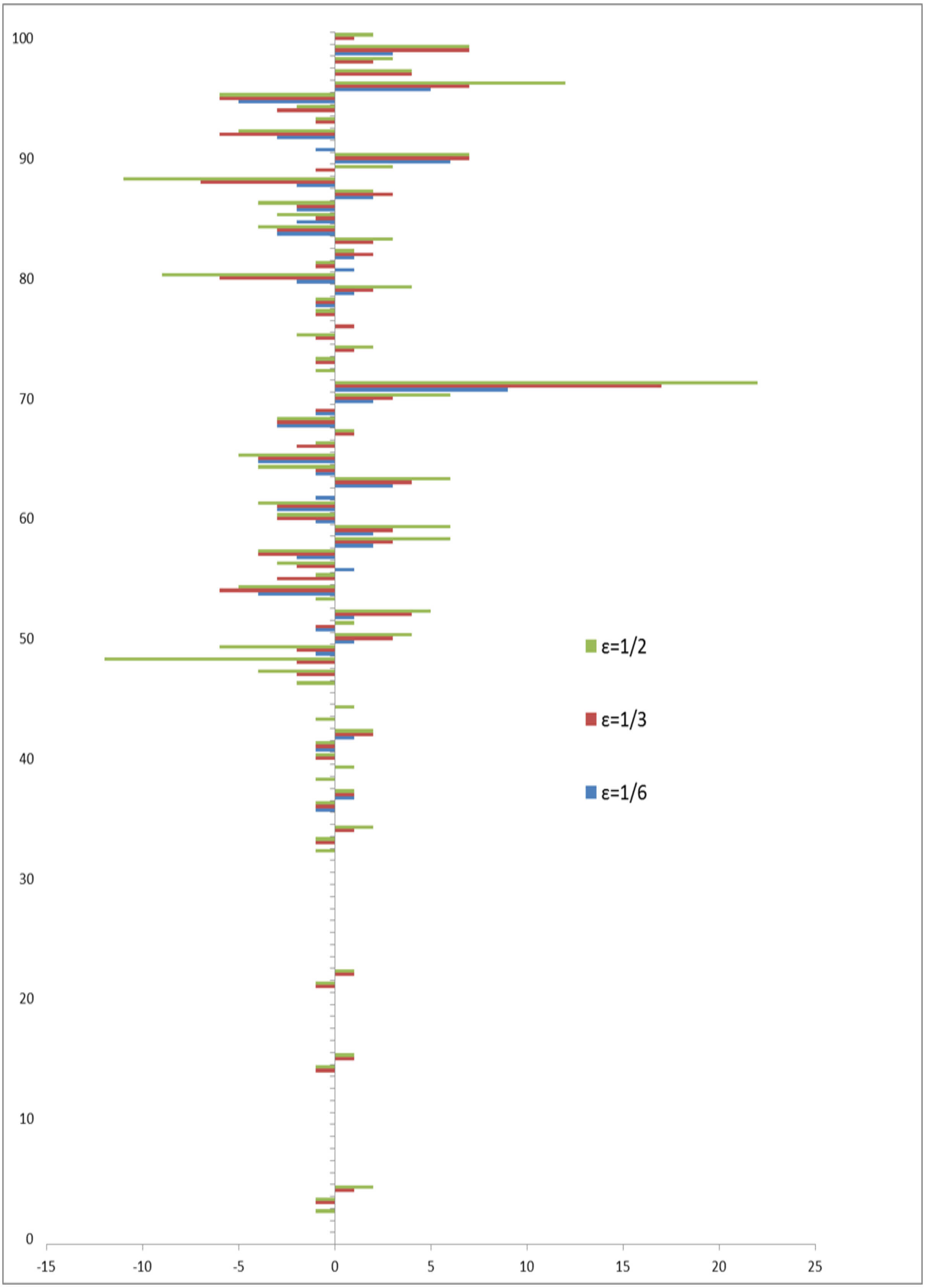

Figure 3: 2013 ARWU Top-100: $K^0 - K^\epsilon$.

References

[1] N. Ailon (2007), "Aggregation of Partial Rankings, p-Ratings and Top-m Lists," *Proc 18th Annual Sympos. Discrete Algorithms (SODA '07)*, 415-424.

[2] S. Alkire and J.E. Foster (2011), "Counting and Multidimensional Poverty Measurement," *Journal of Public Economics*, 95, 476–487.

[3] S. Anand and A.K. Sen (1997). "Concepts of Human Development and Poverty: a Multidimensional Perspective," UNDP. New York.

[4] G. Anderson, I. Crawford, and A. Leicester (2011), "Welfare Rankings from Multivariate Data, a Nonparametric Approach," *Journal of Public Economics*, 95, 247–252.

[5] A. Atkinson (2003), "Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches," *Journal of Economic Inequality*, 1, 51-65.

[6] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA, USA, 1997.

[7] J. Bartholdi III, C. A. Tovey, and M. A. Trick (1989), "Voting schemes for which it can be difficult to tell who won the election", *Social Choice and Welfare*, 6, 157165.

[8] J. Berger and L. Berliner (1986), "Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors," *Annals of Statistics*, 14, 461–486.

[9] B. Bueler, A. Enge, and K. Fukuda (2000), "Exact volume computation for convex polytopes: A practical study," In G. Kalai and G. M. Ziegler, editors, *Polytopes - Combinatorics and Computation*, DMV Seminar Volume 29, 131-154.

[10] F. Bourguignon and S.R. Chakravarty (2003), "The Measurement of Multidimensional Poverty," *Journal of Economic Inequality*, 1, 25-49.

[11] V. Conitzer (2012), "Should Social Network Structure be Taken Into Account in Elections?" *Mathematical Social Sciences*, 64, 100-102.

[12] K. Decancq and M.A. Lugo (2013), "Weights in Multidimensional Indices of wellbeing: An Overview," *Econometric Reviews*, 32, 7–34.

[13] J.-Y. Duclos, D. E. Sahn, and S.D. Younger (2006), "Robust Multidimensional Poverty Comparisons," *Economic Journal*, 116, 943-968.

[14] J.-Y. Duclos, D. E. Sahn, and S.D. Younger (2011), "Partial Multidimensional Inequality Orderings," *Journal of Public Economics*, 95, 225-238.

[15] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar (2001), "Rank Aggregation Methods for the Web," *Proceedings of 10th International World Wide Web Conference*, 613-622.

[16] J.E. Foster, M. McGillivray, and S. Seth (2013), "Composite Indices: Rank Robustness, Statistical Association, and Redundancy," *Econometric Reviews*, 32, 35-56.

[17] J.E. Foster and A.K. Sen. *On Economic Inequality. After a Quarter Century.* Annexe to the Expanded Edition of *On Economic Inequality* by A.K. Sen., Clarendon Press, Oxford, 1997.

[18] Gilboa, I. and M. Marinacci (2013), "Ambiguity and the Bayesian Paradigm," in *Advances in Economics and Econometrics, Tenth World Congress of the Econometric Society: Economic Theory*, Volume I, Eds. D. Acemoglu, M. Arellano, E. Dekel, Cambridge University Press.

[19] David F. Gleich (2009). http://www.mathworks.it/matlabcentral/fileexchange/24134-gaimc-graph-algorithms-in-matlab-code/content/gaimc/scomponents.m . Copyright, Stanford University, 2008-2009.

[20] J.G. Kemeny (1959), "Mathematics without Numbers," *Daedalus*, 88, 575–591.

[21] P. Klibanoff, M. Marinacci, and S. Mukerji (2005), "A Smooth Model of Decision Making Under Ambiguity," *Econometrica*, 73, 1849–1892.

[22] I. Kopylov (2009), "Choice deferral and ambiguity aversion," *Theoretical Economics*, 4, 199–225.

[23] S. Lang, *Linear Algebra*, Springer 3rd Edition, New Haven, CT, USA, 1987.

[24] J. Lawrence (1991), "Polytope Volume Computation," *Mathematics of Computation*, 57, 259–271.

[25] H. Moulin, *Axioms of Cooperative Decision Making*, Econometric Society Monographs, Cambridge University Press, Cambridge, 1988.

[26] K.G. Nishimura and H. Ozaki (2004), "Search and Knightian uncertainty," *Journal of Economic Theory*, 119, 299–333.

[27] K.G. Nishimura and H. Ozaki (2006) "An axiomatic approach to ϵ-contamination," *Economic Theory*, 27, 333–340.

[28] K.G. Nishimura and H. Ozaki (2007), "Irreversible investment and Knightian uncertainty," *Journal of Economic Theory*, 136, 668–694.

[29] OECD and European Commission JRC, *Handbook on Constructing Composite Indicators*, OECD Publications, Paris, France, 2008.

[30] M. Pinar, T. Stengos, and N. Topaloglou (2013), "Measuring Human Development: A Stochastic Dominance Approach," *Journal of Economic Growth*, 18, 69–108.

[31] M. Ravallion (2012), "Troubling Tradeoffs in the Human Development Index," *Journal of Development Economics*, 99, 201-209.

[32] M. Ravallion (2012), "Mashup Indices of Development," *World Bank Research Observer,* 27, 1-32.

[33] M. Saisana, A. Saltelli, and S. Tarantola (2005), "Uncertainty and Sensitivity Analysis as Tools for the Quality of Composite Indicators," *Journal of the Royal Statistical Society A*, 168, 1–17.

[34] M. Saisana, B. D'Hombres, and A. Saltelli (2011), "Rickety Numbers: Volatility of University Rankings and Policy Implications," *Research Policy*, 40, 165–177.

[35] A.K. Sen, *Commodities and Capabilities.* Elsevier, Amsterdam; New York. North-Holland press, 1985.

[36] A.K. Sen (1993), "Capability and Well-Being," in: Nussbaum, M., Sen, A.K. (Eds). Quality of Life, Clarendon Press, Oxford, 30-53.

[37] R. Tarjan (1972), "Depth-first Search and Linear Graph Algorithms," *SIAM Journal on Computing*, 1, 146–160.

[38] K. Tsui (1995), "Multidimensional Generalizations of the Relative and Absolute Inequality Indices: The Atkinson-Kolm-Sen Approach," *Journal of Economic Theory*, 67, 251-266.

[39] United Nations Development Programme, *2013 Human Development Report. The Rise of the South: Human Progress in a Diverse World*, New York, NY, 2013.

[40] H. P. Young (1974), "An Axiomatization of Borda's Rule," *Journal of Economic Theory*, 9, 43-52.

[41] H. P. Young (1988), "Condorcet's Theory of Voting," *American Political Science Review*, 82, 1231–1244

[42] H. P. Young (1995), "Optimal Voting Rules," *Journal of Economic Perspectives*, 9, 51-64.

[43] H. P. Young and A. Levenglick (1978), "A Consistent Extension of Condorcet's election principle," *SIAM Journal on Applied Mathematics*, 35, 285–300.

[44] A. van Zuylen and D. Williamson (2009), "Deterministic Pivoting Algorithms for Constrained Ranking and Clustering Problems," *Mathematics of Operations Research*, 34, 594–620.