



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK



Christoph Jansen, Georg Schollmeyer and Thomas Augustin

A Probabilistic Evaluation Framework for Preference Aggregation Reflecting Group Homogeneity

Technical Report Number 193, 2018
Department of Statistics
University of Munich

<http://www.statistik.uni-muenchen.de>



A Probabilistic Evaluation Framework for Preference Aggregation Reflecting Group Homogeneity

C. Jansen^{a,*}, G. Schollmeyer^a, T. Augustin^a

^a*Department of Statistics, LMU Munich, Ludwigsstr. 33, 80539 Munich, Germany*

Abstract

Groups differ in the homogeneity of their members' preferences. Reflecting this, we propose a probabilistic criterion for evaluating and comparing the adequateness of preference aggregation procedures that takes into account information on the considered group's homogeneity structure. Further, we discuss two approaches for approximating our criterion if information is only imperfectly given and show how to estimate these approximations from data. As a preparation, we elaborate some general minimal requirements for measuring homogeneity and discuss a specific proposal for a homogeneity measure. Finally, we investigate our framework by comparing aggregation rules in a simulation study.

Keywords: Aggregation procedure, preference profile, voting theory, imprecise probabilities, maximum entropy, homogeneity measure, group decision making. JEL classification: C1, C6

1. Introduction

One of the fundamental tasks in social choice theory is to define adequately justified rules for aggregating the preferences of a group of individuals into one global consensus order. Due to the generality of this problem, it is hardly surprising that many different rules have been proposed since the pioneering works by de Borda (1781); de Condorcet (1785); Hare (1857) (see Brams and Fishburn (2002) for a survey). More generally, the question of aggregating collections of binary relations in a meaningful way does not exclusively concern social choice theory, but also appears in classification problems in statistics (see, e.g., Maniqueta and Mongin (2016)), benchmarking of algorithms in the computer sciences (see, e.g., Mersmann et al. (2015)) or problems of judgment aggregation in philosophy (see, e.g., Hartmann and Sprenger (2012)) to name only a few examples.

Given the diversity of aggregation rules, criteria for evaluating and comparing their quality need to be established. Many different criteria have been proposed, and comparisons of aggregation rules with respect to them have been studied intensively (see, e.g., Grofman and Feld (2004)). However, almost all these criteria are non-group-specific: They are intended to be valid independently of the group whose members' preferences are to be aggregated. But what is a perfectly adequate aggregation procedure for one group may not be as appropriate for another one. The adequateness of an aggregation procedure may, beyond compatibility with non-group-specific criteria, additionally

*Corresponding author

Email address: christoph.jansen@stat.uni-muenchen.de (C. Jansen)

depend on certain characteristics of the specific group under consideration. One such characteristic is the homogeneity of the group members (see Section 2.2 for a discussion of the literature on homogeneity). In this paper we propose a group-specific quality criterion for aggregation rules that takes into account information on the homogeneity of group members' preference structure. Moreover, we show different ways to approximate our criterion under partial probabilistic information and discuss how to estimate these approximations in the presence of data or expert knowledge.

More precisely, the paper is structured as follows: In Section 2, we discuss measures for quantifying the homogeneity of a group that is represented by a fixed profile (R_1, \dots, R_n) of strict weak orders. Specifically, in Section 2.3, we elaborate a list of three minimal requirements that every reasonable measure should satisfy. In Section 2.4, we then propose a concrete measure, the maximum consensus homogeneity, and discuss why it is reasonable beyond its mere compatibility with these minimal requirements. Section 3, after reviewing some basics on Bayesian theory in Section 3.1, introduces a framework for evaluating and comparing aggregation procedures in the presence of probabilistic information on the considered group. This involves three steps: In Section 3.2, we introduce an optimality criterion that requires perfect knowledge of the probabilities with respect to which the group constitutes different profiles (R_1, \dots, R_n) . Section 3.3 discusses approaches for approximating this criterion if the probabilistic information on the group is partial in the sense that only the probability distribution of some homogeneity measure is given. Finally, Section 3.4 discusses several statistical approaches for estimating this distribution in the presence of data, expert knowledge, or both. Section 4 starts by briefly reviewing some common aggregation procedures relevant to our context (Section 4.1). Afterwards, Section 4.2 summarizes an aggregation procedure recently proposed in Schollmeyer (2017).¹ In Section 5, we investigate the aggregation procedures reviewed, in respect to our criterion in a simulation study. Section 6 is reserved for concluding remarks as well as an outlook on future research questions.

2. Measuring Homogeneity of Preference Profiles

We begin the section by introducing our notation and terminology (Section 2.1) and surveying some related work on the topic (Section 2.2). Subsequently, we establish and discuss a weak set of conditions (Section 2.3) as well as a concrete proposal (Section 2.4) for measuring the homogeneity of a fixed collection (R_1, \dots, R_n) of strict weak orders each of which representing the opinion of a member of a group of size n .

2.1. Notation and Terminology

Throughout the paper, C denotes a finite set of at least two consequences. The elements of C have to be ranked by the members of a specific group G_n of fixed size $n \geq 2$, where certain requirements of rationality regarding the individual orders involved are imposed. Specifically, we work with the following spaces of binary relations on C :

$$\mathcal{R} := \{R \subset C^2 : R \text{ asymmetric, negatively transitive}\} \quad (1)$$

$$\mathcal{Q} := \{Q \subset C^2 : Q \text{ asymmetric}\} \quad (2)$$

In the sequel, every $R \in \mathcal{R}$ is termed a strict weak order on C . For every $R \in \mathcal{R}$, define the usual equivalence relation \sim_R on C by setting $a \sim_R b$ if and only if $(a, b) \notin R \wedge (b, a) \notin R$. Given this,

¹For an explanation of the procedure and a discussion see Section 4.2.

interpret $(a, b) \in R$ as a is strictly preferred to b and $(a, b) \in \sim_R$ as indifference between a and b . The elements of \mathcal{R} are associated with the individual orders of the group members. Hence, the group members are assumed to have asymmetric and negatively transitive preferences. Importantly, note that our model of the individual preferences excludes incomparability of consequences: For alternatives $a, b \in C$ chosen arbitrarily, every group member is thus assumed to be able to decide if she strictly prefers a to b , or b to a , or if she ranks them equally desirable. Thus, we explicitly assume that incomparability with respect to $R \in \mathcal{R}$ is interpreted as indifference (see, e.g., Kreps (1988, Chapter 2) for a discussion of this convention).² For $n \geq 2$, an element $\underline{R} := (R_1, \dots, R_n) \in \mathcal{R}^n$ is called a preference profile on C and each component of \underline{R} is interpreted as the opinion of a member of G_n about how the consequences in C should be ranked.

Contrarily, every element $Q \in \mathcal{Q}$ is called a consensus order (or group preference). Except for asymmetry, we do not impose any further restrictions on the consensus order. This allows for also investigating aggregation procedures for which the group preference is not always as well-behaved as the individual orders (this includes, e.g., Condorcet's method, see Section 4.1, which might yield intransitive consensus orders). In this context, every mapping $S : \mathcal{R}^n \rightarrow \mathcal{Q}$ is called a preference aggregation function. Particularly, for every preference profile $\underline{R} \in \mathcal{R}^n$, the image $S(\underline{R}) \in \mathcal{Q}$ is the consensus order of the group represented by \underline{R} with respect to the aggregation procedure described by S .

2.2. Preference Homogeneity in Related Work

In literature on social choice theory at least two different lines of how to establish a notion of homogeneity of groups can be identified. One line (see, e.g., Niemi (1969); Jamison and Luce (1972); Berg (1985); Gehrlein and Lepelley (2010); Lepelley and Valognes (2003)), which could be called “model-based”, builds up stochastic models that govern the constitution of profiles and have specific parameters implicitly regulating the group's homogeneity. One prominent example is the multivariate Pólya-Eggenberger urn model (see, e.g., Johnson and Kotz (1977)), which has been used for instance in Berg (1985); Gehrlein and Lepelley (2010); Lepelley and Valognes (2003) in order to analyze the relationship between group homogeneity and the probability of the voting paradox or the manipulability of different aggregation functions. The Pólya-Eggenberger model contains two other well-established models as special cases: impartial culture and impartial anonymous culture, which are also often presumed in studies of the voting paradox and the manipulability of aggregation procedures (see, e.g., Aleskerov et al. (2012); Diss et al. (2012); Pritchard and Slinko (2006)). Other model-based approaches, in which the orders in the profile are assumed to be randomly drawn with replacement, measure the homogeneity of the generating process by the probabilities p_i ($i = 1, \dots, |C|!$) with respect to which the order R_i is drawn: Natural measures of homogeneity are then the variance of the p_i s used for instance in Abrams (1976) or the Herfindahl index $\sum_{i=1}^{|C|!} p_i^2$ used, for instance, in Gehrlein (1981). Measures that only rely on the values of the p_i 's and not on the concrete associated orders R_i are called non-profile specific measures (see Gehrlein (1981)). Since they are related to the probabilities p_i , they are also called population specific homogeneity measures in Gehrlein and Lepelley (2010, p. 191).

²An alternative approach would be to directly model the individual preferences by weak orders, i.e. complete and transitive binary relations $P \subset C^2$. To every such relation we then can associate its strict part $R_P \subset C^2$ by setting $(a, b) \in R_P$ if and only if $(a, b) \in P \wedge (b, a) \notin P$ for all $a, b \in C$. The relation R_P is then asymmetric and negatively transitive. Our model thus explicitly assumes that the individual orders $R \in \mathcal{R}$ arise as strict parts of a weak order.

A second line of establishing a notion of homogeneity, which can be called “data-based”, relates homogeneity not to a probabilistic model but to the actually observed data in a profile. For example, in the above approaches, one can replace the probability p_i of observing the order R_i in a profile with the relative frequency of the associated order in the actually observed profile. Then one arrives at a notion of homogeneity that is no related to a generating process, but instead related to the observed profile. Such measures are called situation specific homogeneity measures in Gehrlein and Lepelley (2010, p. 192). A further type of such data-based measures are distance-based measures, which additionally utilize the information in the orders of the profile. These measures, arising not only in social choice theory but also in statistics and computer sciences (see, e.g., Fligner and Verducci (1986); Dwork et al. (2001)), rather rely on a geometric understanding and first introduce a distance between pairs of orders. Based on this distance, one defines a measure of heterogeneity by computing the average distance of all pairs of orders in the profile. Homogeneity of the profile is then measured by comparing the maximal distance to this average distance. This type of measures is local in the sense that not the whole group is examined simultaneously, as only pairs are considered. Another data-based measure of homogeneity, especially used in social choice theory (see, e.g., Fishburn (1973)), is the W coefficient introduced in Kendall and Smith (1939). This measure intends to analyze the whole population simultaneously by looking at the variance of the vector of the summarized ranks of each consequence. However, note that also this measure, as shown by Kendall and Smith (1939), could be alternatively represented as the average Spearman correlation coefficient of pairs of rank-vectors and is thus also local in the above sense.

Beyond concrete proposals for data-based homogeneity measures, axiomatic approaches have also been studied (see, e.g., Bosch (2006); Alcalde-Unzu and Vorsatz (2013)). Here, the terms consensus and cohesiveness are used instead of homogeneity. For the concept of polarization, a concept very similar, but not identical to the concept of heterogeneity,³ an axiomatic characterization of a measure of polarization of profiles is given in Can et al. (2015).

2.3. Minimal Requirements for Measuring Homogeneity

Before introducing a concrete non-local and profile-specific homogeneity measure in the next section, we first set out to agree on some minimal requirements that, in our eyes, every reasonable candidate for such a measure should necessarily satisfy. We list these requirements in the following definition. Afterwards, a discussion of each is given.

Definition 1. *A preference homogeneity measure (for a group of size n) is a map $A_n : \mathcal{R}^n \rightarrow [0, 1]$ satisfying the following three properties:*

- (S1) **Consensus sensitivity:** $A_n(\underline{R}) = 1$ if and only if $\underline{R} = (R^*, \dots, R^*)$ for some $R^* \in \mathcal{R}$.
- (S2) **Anonymity:** Let $\phi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ be a bijective map. Then $A_n(R_1, \dots, R_n) = A_n(R_{\phi(1)}, \dots, R_{\phi(n)})$ for all $(R_1, \dots, R_n) \in \mathcal{R}^n$.
- (S3) **Majority strengthening:** Let $\underline{R} \in \mathcal{R}^n$. Define $k(j) := \{i : R_i = R_j\}$. If there exists $j_0 \in \{1, \dots, n\}$ such that $n > |k(j_0)| \geq \lfloor \frac{n}{2} \rfloor$, choose $j_1 \in \{1, \dots, n\} \setminus k(j_0)$ and define

³While the notion of heterogeneity refers here to the diversity of the orders in the profile, polarization means that the orders in the profile are clustered in two or more “opposite” subgroups. A clear cut rigorous disambiguation between polarization and heterogeneity for the case of preference profiles is, as far as the authors are aware, not yet established. For a more elaborate disambiguation between polarization and heterogeneity/inequality in the context of, for instance, poverty measurement, see, e.g., Esteban and Ray (1994); Duclos et al. (2004).

$\phi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ by $\phi(j) = j_0$, if $j \in k(j_0) \cup \{j_1\}$ and $\phi(j) = j$ else. Then we have $A_n(R_1, \dots, R_n) \leq A_n(R_{\phi(1)}, \dots, R_{\phi(n)})$.

So, why is it reasonable to require (S1), (S2), and (S3) as a minimal basis for measuring homogeneity of preference profiles? Consensus sensitivity states that every measure should be capable of identifying perfect consensus by attaining its maximal value 1 if and only if all group members share identical preferences. This is certainly reasonable since it reflects the fact that we know exactly that the identical profiles are ideal and superior to the non-identical ones with respect to homogeneity. This knowledge should not get lost by the construction of the measure. Anonymity ensures that the homogeneity value of a profile does not depend on the order in which the individuals state their preferences, as no individual has greater influence. Finally, majority strengthening can be interpreted as a weak demand for monotonicity: If a subgroup consisting of at least $\lfloor \frac{n}{2} \rfloor$ group members shares identical preferences and one member from outside this subgroup changes her mind towards this subgroup, then the homogeneity value of the modified profile should not decrease.

Clearly, all three conditions rely solely on the categorical and not the ordinal scale of measurement of the orders in the profile, i.e. one conceptually only distinguishes between equal and non-equal orders and does not make use of for example a notion of how similar different orders are by e.g. counting edges that two different orders have in common. Of course, one could also establish a notion of (S3) that uses the ordinal structure by stating for instance that if one order R in the profile is changed towards another order R' that is more similar to the order of the majority, then the homogeneity should not decrease. However, this would require a notion of what the terms “majority” and “more similar order” then exactly mean. Note further that adequately axiomatizing more subtle aspects like the difference between heterogeneity and polarization seems to be not possible if one only relies on the categorical scale of measurement of the orders in the profile. In this sense, the conditions (S1), (S2), and (S3) should indeed be understood as minimal requirements for a notion of homogeneity that leaves much space for content matter considerations in the final choice of the measure.⁴

2.4. The Maximum Consensus Homogeneity

We now introduce a specific homogeneity measure, the so-called maximum consensus homogeneity, show that this measure satisfies the minimal requirements given in Definition 1, and discuss why it is a reasonable choice for our purposes beyond its mere compatibility with the minimal requirements. The basic idea of the measure is to compare, for each pair (a, b) separately, the maximal number of coinciding opinions about that pair in the profile to the maximal possible number n .⁵

Some additional notation is needed: Let $n \geq 2$ and let $R_0 \in \mathcal{R}$ with $\sim_{R_0} = \{(c, c) : c \in C\}$ be fixed, such that R_0 always contains exactly one of the pairs (a, b) or (b, a) for all distinct $a, b, \in C$. For a fixed preference profile $\underline{R} \in \mathcal{R}^n$ and a fixed pair of distinct consequences $(a, b) \in C^2$, we define the expressions $c_{\underline{R}}(a, b) := |\{i : (a, b) \in R_i\}|$ and $e_{\underline{R}}(a, b) := |\{i : (a, b) \in \sim_{R_i}\}|$ to be, respectively, the number of individuals in \underline{R} that prefer a to b and the number of individuals that are indifferent between these options.

⁴An (in parts) similar axiomatization, however stronger, is given in Alcalde-Unzu and Vorsatz (2013) in the context of measuring cohesiveness of preferences profiles.

⁵A similar measure is introduced in Can et al. (2015): There, the authors first list a set of axioms for measures of *polarization* that uniquely characterize a measure that is closely related to the one used in this work.

Definition 2. The mapping $\delta_n : \mathcal{R}^n \rightarrow [0, 1]$ defined by

$$\delta_n(\underline{R}) := \frac{\sum_{(a,b) \in R_0} \max\{c_{\underline{R}}(a,b), c_{\underline{R}}(b,a), e_{\underline{R}}(a,b)\}}{n \cdot \binom{|C|}{2}} \quad (3)$$

for all $\underline{R} \in \mathcal{R}^n$ is called *maximum consensus homogeneity*.

Importantly, note that the definition of δ_n does not depend on the choice of $R_0 \in \mathcal{R}$ with the desired properties (see Appendix A1). As a first step in our discussion of the proposed maximum consensus measure δ_n , we show that it does indeed satisfy the conditions (S1), (S2), and (S3). Therefore, we consider it compatible with the minimal requirements a measure of homogeneity should satisfy. This is the assertion made in the following proposition. The proof consists in straightforwardly verifying (S1), (S2), and (S3) from Definition 1 and is given in Appendix A1.

Proposition 1. *The maximum consensus homogeneity δ_n satisfies (S1), (S2), and (S3).*

So, why is the maximum consensus measure δ_n a reasonable candidate for measuring homogeneity and what makes it preferable to the measures discussed in Section 2.2 for our purposes? First, δ_n utilizes the information encoded in the orders collected in the inserted profiles and does not solely rely on the shares p_i of identical orders. Accordingly, δ_n is profile-specific (in contrast to, e.g., Herfidahl's index). This certainly is a desirable property, since any measure of homogeneity should be capable of distinguishing between profiles of very similar yet not identical orders and profiles of completely opposed orders. Second, δ_n is not a local measure in the sense of being only based on pairwise comparisons of individual orders (see the discussion in Section 2.2): For computing the value $\delta_n(\underline{R})$ the whole profile needs to be examined simultaneously. This is a very desirable property conceptually, since group homogeneity should depend on the group as a whole rather than on comparisons of pairs of individuals only. Note that, in addition to this argument, distance-based homogeneity measures also satisfy the minimal conditions (S1) to (S3) from Definition 1.⁶ Finally, note that the classical measure W of Kendall and Smith mentioned in Section 2.2 does not satisfy majority strengthening: For a counterexample, take $C = \{a, b, c, d, e\}$ and consider the profiles $\underline{R} = (R_1, R_1, R_1, R_2, R_3)$ and $\underline{R}' = (R_1, R_1, R_1, R_1, R_3)$ where relation R_1 ranks $a b c d e$, relation R_2 ranks $a b c e d$ and relation R_3 ranks $e b c a d$. Clearly, the majority strengthening condition (S3) requires assigning higher homogeneity to the profile \underline{R}' , but simple calculations yield $W(\underline{R}') = 0.584 < 0.592 = W(\underline{R})$.

3. A Probabilistic Evaluation Framework for Preference Aggregation Functions

Section 3.1 recalls required concepts from Bayesian statistics (see, e.g., Berger (1980); Gelman et al. (2004) for monographs). Subsequently, based on the concept of preference homogeneity measures from the previous section, we propose a probabilistic criterion for evaluating the adequateness of a preference aggregation function S for a fixed group G_n of size n . Specifically, this will involve three steps: Firstly, in Section 3.2, we introduce a theoretical criterion $m_{G_n}^u(S)$ that measures the expected similarity that S yields given the true probability measure \mathbb{P}_{G_n} with respect to which the group constitutes different profiles $\underline{R} \in \mathcal{R}^n$. Secondly, reflecting the fact that in reality the

⁶A formal justification of the non-locality and a discussion of distance-based homogeneity measures in the light of Definition 1 are given in Appendix 3.

measure \mathbb{P}_{G_n} will typically be unknown, Section 3.3 shows how to construct approximations for it if only the distribution of some homogeneity measure is available instead. Finally, in Section 3.4, we discuss different methods for estimating the distribution of a homogeneity measure.

3.1. Required Concepts of Bayesian Statistics

Roughly stated, Bayesian theory addresses two fundamental questions: (Q1) *How to model an agent's beliefs in the light of uncertainty* and (Q2) *How to update the model once new information is gained*. In classical Bayesian theory as pioneered by de Finetti's concept of subjective probability (see, in particular, de Finetti (1974)), Question (Q1) is addressed by the assumption that an agent's beliefs/information about any uncertain phenomenon (independent of whether that phenomenon is random or not) can be perfectly characterized by a unique subjective probability measure π on the space of potential outcomes of the phenomenon. All reasoning should then be based on this unique probability measure π .

However, apart from classic Bayesian theory, this uniqueness is often strongly doubted for being too demanding in regard to the consistency of the agent's beliefs. Instead, beliefs are assumed to be only partial in that they specify a whole set \mathcal{M} of probability measures compatible with them. Then, two main approaches are followed: The first one establishes criteria for choosing one particular distribution from the set \mathcal{M} and, subsequently, bases all further analyses on the chosen representative. The most common choice for such a criterion is Jayne's maximum entropy principle (see Jaynes (1957)): Among all measures compatible with the beliefs, choose the one that is least informative and thus best captures the complete ignorance among the compatible measures (see Rosenkrantz (1977, Section 3.5)). Particularly, the informativeness of a distribution used in the definition of the principle is measured by means of Shannon's entropy (see Shannon (1949)). For a more recent justification of the maximum entropy principle see, e.g., Landes and Williamson (2013). The second approach treats the set \mathcal{M} of all compatible distributions, also called credal set⁷ in this context, as an entity of its own: The agent's beliefs are represented by all members of \mathcal{M} , not just by one single representative. Clearly, an argument supporting this approach is that it avoids any selection: Even a well-justified criterion might select a rather bad representative in certain situations and therefore could yield misleading reasoning. Contrarily, reasoning based solely on the credal set obviously produces less informative results. For a detailed discussion of the advantages and disadvantages of the two directions and a decision-theoretical justification of maximum entropy see Walley (1991, Section 5.12). From a practical point of view, it often makes sense to consider both approaches simultaneously: Use a well-established selection criterion (such as maximum entropy) and analyze the credibility and robustness of the derived inferences by additionally considering the set of inferences drawn by the credal set.

Let us turn now to Question (Q2): Suppose (new) information x about the uncertain phenomenon is gained (e.g. in the form of data). The agent then updates the unique⁸ measure π describing her beliefs, also called prior distribution in this context, to a posterior measure $\pi|x$ according to Bayes' rule. This posterior measure $\pi|x$ is then assumed to appropriately express the updated beliefs about the uncertain phenomenon given the data x . In the specific context

⁷The name credal set is attributed to Isaac Levi (see Levi (1980)). For the general framework of imprecise probabilities, working with sets of probabilities or interval-valued assignments, see Walley (1991) and Weichselberger (2001), or, for a recent introduction, see Augustin et al. (2014).

⁸For Question (Q2), for the sake of brevity, we restrict presentation to classic Bayesian theory. For discussions on how to adequately update credal sets in light of new information see, e.g., the discussions in Walley (1991).

of Bayesian statistics as used here, this translates as follows: Suppose some random variable $X : \Omega \rightarrow \mathcal{X}$, mapping from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to a measurable space $(\mathcal{X}, \sigma(\mathcal{X}))$, for which we can specify its distribution up to a parameter θ from a parameter space Θ , i.e. we know $X \sim \mathbb{P}_\theta$ given $\theta \in \Theta$ is the true parameter. Following the ideas of Bayesian theory described above, we can then describe the uncertainty about the true parameter θ by a random variable $V : \Omega \rightarrow \Theta$ taking values in the measurable space $(\Theta, \sigma(\Theta))$ with prior distribution π , i.e. $V \sim \pi$, and we know that $\mathbb{P}_\theta(A) = \mathbb{P}(X \in A | V = \theta)$ for all $\theta \in \Theta$ and $A \in \sigma(\mathcal{X})$. After having observed a sample x of X , one then computes the posterior distribution $\pi|x$ by setting $\pi|x(B) := \mathbb{P}(V \in B | X = x)$ for all $B \in \sigma(\Theta)$ and utilizing Bayes' rule.

If an estimate for the parameter rather than the posterior measure itself is of interest, popular choices are to use the expectation, the median or the mode of the posterior distribution. For sake of computational convenience, so-called conjugate families of distributions are often used: A family of distributions $\mathcal{D}_1(\Xi)$ with parameter space Ξ is called conjugate to another family $\mathcal{D}_2(\Theta)$ with parameter space Θ if, whenever $\pi \in \mathcal{D}_1(\Xi)$ and $X \sim \mathbb{P}_\theta \in \mathcal{D}_2(\Theta)$, it holds that $\pi|x \in \mathcal{D}_1(\Xi)$, where x is an observation of X . Hence, such models guarantee that the posterior belongs to the same distribution family as the prior and, therefore, that the posterior as well as its moments can basically be computed by updating only the parameter of the prior distribution. An example for such a conjugate model is the Dirichlet-categorical model, which will be used in Section 3.4.

3.2. A Probabilistic Criterion for Evaluating Preference Aggregation Functions

We now turn to the first step of the construction of our criterion. Therefore, for the moment, we assume the probabilities according to which the members of the group G_n constitute the different profiles contained in \mathcal{R}^n are known. More formally, we consider the measurable space $(\mathcal{R}^n, 2^{\mathcal{R}^n})$ together with a known group-specific probability measure \mathbb{P}_{G_n} , and we interpret the value $\mathbb{P}_{G_n}(\{\underline{R}\})$ as the probability that the members of G_n constitute the preference profile \underline{R} . In order to utilize the probabilistic information (given by \mathbb{P}_{G_n}) in the construction of our criterion, we want to compute the similarity of the individual orders collected in \underline{R} and the consensus order $S(\underline{R})$ that an aggregation rule S yields in expectation. However, before such a criterion can be defined, we need to be more precise about what we mean by the similarity of a consensus relation to a profile of relations. Specifically, we will consider similarity measures of the following kind:

Definition 3. *Let $u : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be a monotone increasing function and let $S : \mathcal{R}^n \rightarrow \mathcal{Q}$ be a preference aggregation function. The mapping*

$$Y_S^u : \mathcal{R}^n \rightarrow \mathbb{R} \quad , \quad \underline{R} \mapsto \sum_{i=1}^n u(|R_i \cap S(\underline{R})|) \quad (4)$$

is called the similarity measure for S with respect to u .

The basic idea underlying a similarity measure in the above sense is to quantify similarity of pairs of relations R_1 and R_2 by computing the cardinality of their intersection $|R_1 \cap R_2|$ or, in other words, by counting the edges shared by both relations. An axiomatic justification is given in Kemeny and Snell (1962), where the authors show that the distance measure $d(R_1, R_2) := |R_1 \Delta R_2| = |(R_1 \cup R_2) \setminus (R_1 \cap R_2)| = |R_1| + |R_2| - 2|R_1 \cap R_2|$ is unique in satisfying four desirable conditions (including the properties of a metric). Together with the assumption that similarity of R_1 and R_2 should be high whenever their uniquely determined distance $d(R_1, R_2)$ is low and vice versa, using $|R_1 \cap R_2| = \frac{1}{2}(|R_1| + |R_2| - d(R_1, R_2)) = \frac{1}{2}(|C|(|C| + 1) - d(R_1, R_2))$ is a natural

choice. For a fixed profile $\underline{R} = (R_1, \dots, R_n)$, the idea of a similarity measure Y_S^u is then to compute the pairwise similarity $|R_i \cap S(\underline{R})|$ between every individual order R_i and the group order $S(\underline{R})$ separately and, afterwards, sum up monotone transformations $u(|R_i \cap S(\underline{R})|)$ of these values. The role of u is to control the influence of high similarity values $|R_{i_0} \cap S(\underline{R})|$ for certain orders R_{i_0} on the global similarity $Y_S^u(\underline{R})$: If u is chosen to be a convex function, high pairwise similarity values will have a strong influence on the global similarity, whereas if u is concave, then increasing the similarity of an inadequately represented group member contributes more to global similarity than doing the same for a group member that is already appropriately represented.

Once having decided which specific similarity measure Y_S^u to use, one can go on to construct a quality criterion for the aggregation function S : We evaluate S by computing the expectation of the chosen similarity function with respect to the group specific probability \mathbb{P}_{G_n} .

Definition 4. Let G_n be a group consisting of n members and let \mathbb{P}_{G_n} denote its group specific probability measure on $(\mathcal{R}^n, 2^{\mathcal{R}^n})$. For a preference aggregation function S and a monotone increasing function $u : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with associated similarity measure Y_S^u , we define the value

$$m_{G_n}^u(S) := \mathbb{E}_{\mathbb{P}_{G_n}}(Y_S^u) = \sum_{\underline{R} \in \mathcal{R}^n} Y_S^u(\underline{R}) \cdot \mathbb{P}_{G_n}(\{\underline{R}\}) \quad (5)$$

Then $m_{G_n}^u(S)$ is called the expected similarity of the aggregation function S with respect to Y_S^u .

The criterion $m_{G_n}^u(\cdot)$ is intended to be applied as follows: Given two aggregation functions S_1 and S_2 and a group G_n that agrees to measure similarity by Y_S^u , the group should prefer aggregation rule S_1 whenever $m_{G_n}^u(S_1) \geq m_{G_n}^u(S_2)$, i.e. if S_1 yields higher expected similarity than S_2 .

In practice, this criterion will often not be directly applicable, since \mathbb{P}_{G_n} cannot be fully specified. However, in many applications there will be at least some information about the homogeneity structure of the preferences of the group under investigation. In the following Section 3.3, we demonstrate how to construct approximations for the true group-specific measure \mathbb{P}_{G_n} if this information is given in the form of the probability distribution of some homogeneity measure A_n and how these approximations can be utilized for estimating expected similarity.

3.3. Constructing Approximations for Expected Similarity

This leads us to the second step of our construction: Let $A_n : \mathcal{R}^n \rightarrow [0, 1]$ denote a fixed preference homogeneity measure attaining exactly the values $k_1 < k_2 < \dots < k_\xi \in [0, 1]$. We assume that the available information on the homogeneity of G_n can be specified as the probability distribution of this homogeneity measure A_n . More formally, we (for the moment) assume to know $\alpha := (\alpha_1, \dots, \alpha_\xi) \in \Delta^{\xi-1} := \{x \in [0, 1]^\xi : \sum_{i=1}^\xi x_i = 1\}$ such that $\mathbb{P}_{G_n}(A_n = k_j) = \alpha_j$ for all $j = 1 \dots, \xi$. Substantially, this relates to the assumption that, even if the full group-specific measure \mathbb{P}_{G_n} is unknown, we still know the probabilities α that the group G_n constitutes a certain degree of homogeneity, which is characterized by the chosen preference homogeneity measure A_n . Given this, our goal is to approximate the true underlying group-specific probability measure \mathbb{P}_{G_n} such that the available knowledge on the distribution of A_n is utilized in the best possible way.

To reach this goal, first note that our assumption naturally characterizes a set of probability measures on $(\mathcal{R}^n, 2^{\mathcal{R}^n})$, namely the credal set \mathcal{M}_α of all probability measures that are compatible with the available information on the distribution of A_n . Formally, we have

$$\mathcal{M}_\alpha := \{\pi \in \mathcal{P}(\mathcal{R}^n) : \pi(A_n^{-1}(k_j)) = \alpha_j \text{ for all } j = 1, \dots, \xi\} \quad (6)$$

where $\mathcal{P}(\mathcal{R}^n)$ denotes the set of all probability measures on the space of profiles $(\mathcal{R}^n, 2^{\mathcal{R}^n})$. Consequently, any element of \mathcal{M}_α is a candidate for the true group specific measure and, therefore, a plausible candidate for approximating it. As discussed in Section 3.1, (at least) two different approaches exist for dealing with the ambiguity between the compatible measures in \mathcal{M}_α in such situations: Applying the maximum entropy principle to \mathcal{M}_α in order to specify the least informative measure or directly working with the set \mathcal{M}_α as an entity of its own. We contrast both approaches and the approximations for expected similarity obtained by them in the sequel:

Maximum entropy approach: Given the distribution of A_n , we know that the probability of the homogeneity class $A_n^{-1}(k_j) := \{\underline{R} \in \mathcal{R}^n : A_n(\underline{R}) = k_j\}$ equals α_j . Contrarily, there is complete ignorance between all measures fixing the probabilities of these classes. Applying the maximum entropy principle, we choose the representative among the compatible measures in \mathcal{M}_α that maximizes Shannon's entropy and that therefore can be viewed as the least informative one. The measure satisfying the desired property is induced by the assignment

$$\mathbb{P}_\alpha^*(\{\underline{R}\}) := \frac{\alpha_{\phi(A_n(\underline{R}))}}{|A_n^{-1}(A_n(\underline{R}))|} \quad (7)$$

for all $\underline{R} \in \mathcal{R}^n$, where $\phi(k_j) := j$ for $j = 1, \dots, \xi$. Among all measures fixing the probability values of the homogeneity classes $A_n^{-1}(k_j)$ to α_j , the resulting probability \mathbb{P}_α^* is exactly the one giving equal probability mass to all profiles belonging to the same class. Therefore, beyond maximizing entropy, the measure \mathbb{P}_α^* is also intuitively appealing: Why should two profiles with coinciding homogeneity value be assumed to have different probability?

Credal set approach: Directly approximating \mathbb{P}_{G_n} with the credal set \mathcal{M}_α protects against possibly misleading inferences based on an unlucky selection of a representative. Obviously, the set \mathcal{M}_α contains exactly these probability measures that are compatible with the probabilities α_j of the homogeneity classes $A_n^{-1}(k_j)$. By construction, we therefore have $\mathbb{P}_{G_n} \in \mathcal{M}_\alpha$ and $\mathbb{P}_\alpha^* \in \mathcal{M}_\alpha$, i.e. the true measure and the maximum entropy measure are contained in the credal set.

By using the two approaches just described, the expectation in (5) can now straightforwardly be approximated by replacing the true measure \mathbb{P}_{G_n} in the expression with the corresponding approximation \mathbb{P}_α^* or \mathcal{M}_α . Note that for the credal set approach this will lead to a set-valued approximation, each element of the set representing the expected similarity with respect to a different distribution from \mathcal{M}_α . Formally, this leads to the following two approaches for approximation.

Maximum entropy approximation: Compute the expected similarity with respect to the maximum entropy measure \mathbb{P}_α^* . We then arrive at the following real-valued approximation:

$$m_{G_n}^{u*}(S) := \mathbb{E}_{\mathbb{P}_\alpha^*}(Y_S^u) = \sum_{\underline{R} \in \mathcal{R}^n} Y_S^u(\underline{R}) \cdot \mathbb{P}_\alpha^*(\{\underline{R}\}) \quad (8)$$

The maximum entropy is represented by a single real number and, therefore, allows for easy comparisons of different aggregation functions S_1 and S_2 . However, the maximum entropy approximation \mathbb{P}_α^* might differ from the true underlying measure \mathbb{P}_{G_n} in a way yielding $m_{G_n}^{u*}(S_1) > m_{G_n}^{u*}(S_2)$ but $m_{G_n}^u(S_2) > m_{G_n}^u(S_1)$ and, thus, might produce misleading comparisons.

Credal approximation: Compute the expectation with respect to the set \mathcal{M}_α , i.e. the interval ranging from the lowest to the highest expected similarity value compatible with a measure from \mathcal{M}_α . We arrive at the following interval-valued approximation:

$$M_{G_n}^u(S) := [\underline{M}_{G_n}^u(S), \overline{M}_{G_n}^u(S)] := \left[\inf_{\pi \in \mathcal{M}_\alpha} \mathbb{E}_\pi(Y_S^u), \sup_{\pi \in \mathcal{M}_\alpha} \mathbb{E}_\pi(Y_S^u) \right] \quad (9)$$

Again, by construction, it holds that $m_{G_n}^u(S) \in M_{G_n}^u(S)$ and $m_{G_n}^{u*}(S) \in M_{G_n}^u(S)$ and thus both the true expected similarity value and its maximum entropy approximation are contained in the interval given by the credal approximation. The smaller the width of the credal interval, the less ambiguity underlies the situation. Consequently, analyses based on the maximum entropy approximation are then more reliable.

For the computation of the approximations (8) and (9), we give a proposition showing that, once the preimages of the homogeneity values are computed, one only has to compute the scalar products of the weight vector α with corresponding fixed vectors associated to the previously computed preimages. This will prove very valuable also to our study in Section 5. Checking the validity of the proposition is straightforward and therefore omitted.

Proposition 2. *For the maximum entropy approximation and the credal approximation defined in (8) and (9), the following equations hold, respectively:*

$$\begin{aligned} i) \quad m_{G_n}^{u*}(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \frac{1}{|A_n^{-1}(k_j)|} \sum_{\underline{R} \in A_n^{-1}(k_j)} Y_S^u(\underline{R}) \right) \\ ii) \quad \underline{M}_{G_n}^u(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \min_{\underline{R} \in A_n^{-1}(k_j)} Y_S^u(\underline{R}) \right) \\ iii) \quad \overline{M}_{G_n}^u(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \max_{\underline{R} \in A_n^{-1}(k_j)} Y_S^u(\underline{R}) \right) \end{aligned}$$

3.4. Estimation of Homogeneity Class Probabilities

Finally, we turn to the last step of the construction described at the beginning of Section 3: In real-world applications, not only the group-specific probability \mathbb{P}_{G_n} , but also the precise homogeneity class probabilities α_j will typically be unknown. Accordingly, an estimate $\hat{\alpha} := (\hat{\alpha}_1, \dots, \hat{\alpha}_\xi)$ for these probabilities has to be obtained. In principle, different ways of addressing this estimation task are conceivable. Firstly, one can draw on expert knowledge, i.e. ask experts from the investigated field for their probability estimates of the homogeneity classes. If more than one expert is involved, an estimate could be received by using either an weighted average of the experts' estimates or by directly working with the credal set containing all of them.

Secondly, one can collect data. For this purpose, a questionnaire can be designed consisting of d items (covering a relevant topic), each of which demanding the participant to order $q := |C|$ alternatives by preference. Each of the n group members participates in the survey such that, after combining the questionnaires, each item produces a preference profile of the relevant group and, therefore, a collection of d preference profiles $\underline{R}_1, \dots, \underline{R}_d$ is received. For each of these profiles we compute the homogeneity measure and receive data $\underline{x} := (x_1, \dots, x_d)$, where $x_s := A_n(\underline{R}_s)$ for $s = 1, \dots, d$. We then estimate α_j by computing relative frequencies

$$\hat{\alpha}_j := \frac{1}{d} \cdot \sum_{s=1}^d \mathbb{1}_{\{k_j\}}(x_s) \quad (10)$$

Finally, expert knowledge and available data can be combined by following a Bayesian approach: A preference homogeneity measure $A_n : \mathcal{R}^n \rightarrow [0, 1]$ defines a categorically distributed random variable⁹ taking values in $\{k_1, \dots, k_\xi\}$. Specifically, since $\alpha_j = \mathbb{P}(A_n = k_j)$, we have $A_n \sim \text{Cat}(\alpha)$.

⁹A random variable X with possible values $\{x_1, \dots, x_k\}$ is called categorically distributed with parameter vector $\lambda = (\lambda_1, \dots, \lambda_k) \in \Delta_{k-1}$, formally $X \sim \text{Cat}(\lambda)$, if the probability that X attains value x_i equals λ_i for all $i = 1, \dots, k$. The categorical distribution is that special case of the multinomial distribution (see, e.g., Berger (1980, p. 562)) where the sample size $n = 1$.

If we, as described in Section 3.1, interpret the parameter $\alpha = (\alpha_1, \dots, \alpha_\xi)$ as a random quantity with a Dirichlet distribution with (hyper-)parameter $\gamma := (\gamma_1, \dots, \gamma_\xi) \in \mathbb{R}_+^\xi$ as a prior¹⁰, we can use the data (x_1, \dots, x_d) from above and compute the posterior distribution of α given \underline{x} . For specifying the parameter γ of the prior distribution drawing on expert knowledge seems to be reasonable. As the family of Dirichlet distributions is conjugate to the family of categorical distributions (see Section 3.1), the posterior is again a Dirichlet distribution with updated posterior-parameter $\gamma|\underline{x} := (\gamma_1|\underline{x}, \dots, \gamma_\xi|\underline{x})$, where $\gamma_j|\underline{x} := \gamma_j + \sum_{s=1}^q \mathbb{1}_{\{k_j\}}(x_s)$ for $j = 1, \dots, \xi$.¹¹ The common choice for estimating α is then the posterior expectation given by

$$\hat{\alpha}_j := \frac{\gamma_j|\underline{x}}{\sum_{l=1}^{\xi} \gamma_l|\underline{x}} \quad (11)$$

Which approach to follow also depends on the situation: If q is large and the homogeneity measure can attain many different values, taking the relative frequencies will often fail, since doing so requires too many data points to be stable. Particularly, in such cases the Bayesian approach has certain advantages. However, this approach needs to specify a hyper-parameter γ .¹² Note that, when it comes to eliciting experts, the advantages of the proposed framework become perfectly clear: Instead of directly asking experts for their probability estimates on the space of profiles \mathcal{R}^n , which contains $(q!)^n$ different elements, one could let them specify a distribution α on the much smaller space $\{k_1, \dots, k_\xi\}$. Due to its very intuitive interpretation as a relatively small homogeneity scale, the distribution α is much easier to enquire about: How homogeneous do you think the considered group is in probability?

4. Aggregation Rules investigated in the Study

In Section 4.1, we briefly survey some common preference aggregation procedures and demonstrate how they straightforwardly extend to our definition of preference aggregation functions. Importantly, it should be noted that all preference aggregation procedures listed in the following section 4.1 are adaptations of the classic rules from literature to the framework that is used in the present paper. In Section 4.2, we shortly describe a new aggregation method, recently proposed in Schollmeyer (2017), which is based on a generalized concept of quantiles on complete lattices.

4.1. Adaptations of some common Aggregation Procedures

Mean rank (Borda count): For $R \in \mathcal{R}$ and $a \in C$, let $\text{rank}_R(a)$ denote the rank of alternative a with respect to R .¹³ The mean rank aggregation function $\text{MR} : \mathcal{R}^n \rightarrow \mathcal{Q}$ is defined by $(a, b) \in$

¹⁰The Dirichlet distribution with parameter vector $\mu = (\mu_1, \dots, \mu_k) \in \mathbb{R}_+^k$ is a probability distribution on the unit simplex Δ_{k-1} . It can therefore be used as a prior distribution for the parameter of a categorical distribution with k categories. For details, see, e.g., Berger (1980, p. 561).

¹¹For further details concerning the Dirichlet-Categorical Model see, e.g., Gelman et al. (2004).

¹²If no expert knowledge for specifying the hyper-parameter is available, a so-called near-vacuous prior model, such as the imprecise Dirichlet model (IDM), can be chosen (see Walley (1996) for the original work or Bernard (2005) for further interesting properties).

¹³Formally, we have $\text{rank}_R(a) := |\{b \in C : (a, b) \in R\}| + \frac{1}{2}|\{b \in C : (a, b) \in \sim_R \wedge a \neq b\}| + 1$. Note that this definition of the rank, as common in statistics, assigns the mean value of all possible ranks to the members of the equivalence classes of \sim_R (see for instance, Yule and Kendall (1924)). Other, less common, choices are to assign to these consequences the minimum or the maximum rank.

$\text{MR}(\underline{R})$ if and only if $\sum_{i=1}^n (\text{rank}_{R_i}(a) - \text{rank}_{R_i}(b)) > 0$, where $\underline{R} \in \mathcal{R}^n$. Specifically, the group assigns each alternative its average rank and prefers alternative a to alternative b iff the latter achieves a strictly lower average rank. a and b are equivalent with respect to $\sim_{\text{MR}(\underline{R})}$ if and only if they have equal average rank in the profile \underline{R} .

Condorcet's method: The Condorcet aggregation function $\text{CO} : \mathcal{R}^n \rightarrow \mathcal{Q}$ is defined by $(a, b) \in \text{CO}(\underline{R})$ if and only if $(c_{\underline{R}}(a, b) > c_{\underline{R}}(b, a) \wedge c_{\underline{R}}(a, b) > e_{\underline{R}}(a, b))$, where $\underline{R} \in \mathcal{R}^n$. Hence, for each pair (a, b) we decide if the majority of the group prefers a to b or vice versa or if the majority of the group is indifferent between a and b . The consequences a and b are considered equivalent with respect to $\sim_{\text{CO}(\underline{R})}$ if and only if either at least half of the group is indifferent between them or an equal number of persons prefer a to b and vice versa. Importantly, note that this is one adaptation of Condorcet's method to our framework: In principle there may be other plausible ways of defining the group's indifference relation.

Instant runoff (Hare's method): Instant runoff is a sequential aggregation procedure: In the first step, all alternatives with the fewest number of first-place votes are excluded from \mathcal{C} . These form the alternatives that are least preferred by the group, and between them the group is indifferent. Afterwards, we exclude the alternatives with the fewest first-place votes in the profile on the reduced space of alternatives and receive a set of alternatives that the group prefers second least. Again, between these alternatives the group is indifferent, but each of them is preferred to every alternative excluded in the first step. Successively repeating this procedure, we end up with a set of optimal options with the same number of first-place votes. This describes a preference aggregation function $\text{IR} : \mathcal{R}^n \rightarrow \mathcal{Q}$ defined by $(a, b) \in \text{IR}(\underline{R})$ if and only if a is excluded at a later stage than b , where $\underline{R} \in \mathcal{R}^n$. The consequences a and b are equivalent with respect to $\sim_{\text{IR}(\underline{R})}$ if and only if they are excluded at the same stage of the procedure.

*Coombs' rule:*¹⁴ The basic idea of Coombs' rule is very similar to that of instant runoff voting, as it is also based on sequential exclusion of alternatives. However, in contrast to instant runoff voting, we exclude the alternatives with the maximal number of last-place votes in every step. The corresponding aggregation function $\text{CM} : \mathcal{R}^n \rightarrow \mathcal{Q}$ is defined by $(a, b) \in \text{CM}(\underline{R})$ if and only if a is excluded at a later stage than b , where $\underline{R} \in \mathcal{R}^n$. Moreover, a and b are equivalent with respect to $\sim_{\text{CM}(\underline{R})}$ if and only if they are excluded at the same stage.

*Kemeny's rule:*¹⁵ Given a profile $\underline{R} := (R_1, \dots, R_n) \in \mathcal{R}^n$, the idea of Kemeny's rule is to choose that consensus order $Q \in \mathcal{R}$ that minimizes the sum of the distances $d(R_i, Q)$ to the individual orders (see Section 3.2). Formally, we say Q^* is a Kemeny consensus ranking for \mathcal{R} whenever it holds that $Q^* \in \text{argmin}_{Q \in \mathcal{R}} \sum_{i=1}^n d(R_i, Q) =: \text{KE}(\underline{R})$. Since such Q^* will generally not be unique, in order to define an aggregation function from this rule, we need to choose a choice function $f : 2^{\mathcal{R}} \setminus \{\emptyset\} \rightarrow \mathcal{R}$ satisfying $f(\mathcal{C}) \in \mathcal{C}$ for all $\mathcal{C} \in 2^{\mathcal{R}} \setminus \{\emptyset\}$. The Kemeny aggregation function $\text{KE}_f : \mathcal{R}^n \rightarrow \mathcal{R}$ with respect to f is then given by $\text{KE}_f(\underline{R}) := f(\text{KE}(\underline{R}))$ for all $\underline{R} \in \mathcal{R}^n$.¹⁶ In

¹⁴Cf. Coombs and Cohen (1984) for a discussion or Grofman and Feld (2004) for comparisons with Hare's method.

¹⁵Originally proposed in Kemeny (1959); some nice properties of the method are shown in Young and Levenglick (1978). Generally, note that determining a Kemeny consensus ranking for \underline{R} is NP-hard (see Bartholdi et al. (1989). Also, compare Ali and Meila (2012) for a comparison of algorithms alleviating the NP-hardness.

¹⁶Note that, since $\sum_{i=1}^n d(R_i, Q)$ is minimal iff $\sum_{i=1}^n |R_i \cap Q|$ is maximal (see Section 3.2), this implies $Y_{\text{KE}_f}^{id}(\underline{R}) \geq Y_S^{id}(\underline{R})$ for all $\underline{R} \in \mathcal{R}^n$ for every other preference aggregation function S , where $id(x) := x$ for all $x \in \mathbb{R}$. Consequently, this implies $\mathbb{E}_{\mathbb{P}}(Y_{\text{KE}_f}^{id}) \geq \mathbb{E}_{\mathbb{P}}(Y_S^{id})$ for any probability measure \mathbb{P} and, thus, Kemeny's rule is superior to every other method if similarity is measured based on the function $u = id$.

the study in Section 5, we enumerated the orders and defined $f(\mathcal{C}) = R_{j_0}$, where j_0 is the smallest index of an order belonging to \mathcal{C} .

Dictatorship: For $i_0 \in \{1, \dots, n\}$, the dictatorship aggregation function $\text{DI}_{i_0} : \mathcal{R}^n \rightarrow \mathcal{Q}$ is defined by $(a, b) \in \text{DI}_{i_0}(\underline{R})$ iff $(a, b) \in R_{i_0}$, where $\underline{R} \in \mathcal{R}^n$. Hence, the group prefers a to b whenever individual i_0 does. Even if this does not seem like the fairest way of aggregating preferences, it might be worth investigating how the dictatorship function performs in comparison to the others.

4.2. An aggregation rule based on quantiles on complete lattices: commonality sharing

We now briefly sketch the idea of the commonality sharing aggregation rule that was recently proposed in Schollmeyer (2017) and that initially arose in a different context, namely through attempts to generalize concepts of centrality and outlyingness of observations to partially ordered data. Opposed to the other methods investigated here, this aggregation procedure does not locally look at different alternatives or pairs of alternatives, but takes into account the full order of all individuals and embeds these orders into the complete lattice of binary relations on C equipped with the set intersection and set union as meets and joins, respectively. Then, a notion of outlyingness in this space, described in Schollmeyer (2017), is used to select one or more orders of individuals who are most centered in the population. Concretely, the following procedure can be applied:

For a given minimum size k , one looks at every possible sub-population M_k^i consisting of at least k individuals. Then, one considers the set Q_k of all individuals q_k^j who share with every sub-population M_k^i all commonalities of this sub-population (i.e., all edges (a, b) that the population M_k^i has in common should also be edges of every order q_k^j in Q_k). The set Q_k of individuals who share with every sub-population of size $\geq k$ its commonalities is to some extent representative for every such sub-population. If k is too small, then Q_k is empty. In contrast, for $k = n$ the set Q_k is the whole population. Now, for a given order q , the smaller the smallest k such that Q_k still contains q , the more central is the order q , since then q is a representative for a bigger collection of sub-populations including smaller sub-populations with bigger and thus more specific commonalities. Finally, to select a consensus order, choose k as small as possible under the restriction that $Q_k \neq \emptyset$ and choose the arising Q_k as the set of candidates for the consensus order. If Q_k has more than one element then for a unique consensus order choose arbitrarily from the set Q_k or apply some further aggregation rule to the orders in Q_k . In the study of Section 5 we apply the first approach and, like in the procedure for Kemeny's rule, choose that most central order with the smallest index.

Note that the commonality sharing consensus rule is in fact a non-local rule in the sense that if for example two individuals in a profile both prefer all alternatives in the set $\{a, b, c\}$ over the alternatives in the set $\{d, e, f\}$, but with different orders within these sets, then the consensus order could possibly change if the individuals swap their orders within the set $\{a, b, c\}$ with each other while maintaining their orders within the set $\{d, e, f\}$. Thus, the orders do in fact play a role as a whole. This is the main difference from the aggregation rules of Section 4.1 (except dictatorship): There, for example, it does not matter which individual an alternative gets its score from in the mean rank aggregation or where edges for pairs of alternatives in Condorcet's method are counted without differentiating between edges belonging to the same individual and edges belonging to different individuals. Note that commonality sharing can be computed in $\mathcal{O}(n \cdot |C|^2)$ time, much simpler than one would expect from the conceptual description.

5. A Simulation Study for the Case $n = 8$ and $|C| = 4$

In this section, we apply the evaluation framework developed in Sections 2 and 3 to the aggregation functions discussed in Section 4. Specifically, we design a study allowing for the comparison of the appropriateness of these aggregation functions for groups of varying degrees of homogeneity in Section 5.1 and then discuss the results in Section 5.2.

5.1. Setup of the Study

Throughout this study, we consider groups G_8 consisting of $n = 8$ group members, each of which has ranked $|C| = 4$ consequences. We further assume that within the individual preferences there is no indifference, i.e., we restrict analysis to the set \mathcal{H} of all strict weak orders $R \in \mathcal{R}$ satisfying $\sim_R = \{(c, c) : c \in C\}$. In order to analyze the appropriateness of different aggregation functions S for groups of varying degrees of homogeneity, we compute and compare the maximum entropy approximation from (8) and the credal set approximation from (9) for different choices of the homogeneity class probabilities α . Therefore, we need to specify three things: A preference homogeneity measure A_n in the sense of Definition 1, a similarity measure Y_S^u in the sense of Definition 3, and a sequence of probability vectors $\alpha^{[i]}$, each determining a distribution of A_n while representing a different degree of group homogeneity.

Firstly, for A_n , we choose the maximum consensus homogeneity δ_n from Definition 2 restricted to the domain \mathcal{H}^8 , which in this case reduces to $\delta_8(\underline{R}) = \frac{1}{48} \sum_{(a,b) \in R_0} \max\{c_{\underline{R}}(a, b), c_{\underline{R}}(b, a)\}$ for all $\underline{R} \in \mathcal{H}^8$. Secondly, for measuring similarity between a profile and its consensus order with respect to an aggregation function S , we consider three different choices for the function u in the similarity measure Y_S^u , taking into account the discussion after Definition 3: The linear function $u_1(x) = x$ giving equal influence to all group members, the convex function $u_2(x) = x^2$ giving higher influence to appropriately represented group members, and the concave function $u_3(x) = 36 - (6 - x)^2$ strengthening the influence of inadequately represented members. Note that u_1 and u_3 were also proposed in Kemeny (1959).

Finally, in order to model varying degrees of group homogeneity of the considered groups G_8 , first note that δ_8 attains the values $k_1 = \frac{24}{48} < k_2 = \frac{25}{48} < \dots < k_{25} = \frac{48}{48}$, where $\frac{1}{2}$ indicates minimal homogeneity and 1 indicates perfect homogeneity of the inserted profile. Accordingly, assuming δ_8 to be categorically distributed with parameter vector $\alpha_{min} = (1, 0, \dots, 0)$ obviously represents a group G_8 with a lower degree of homogeneity than assuming the same for parameter vector $\alpha_{max} = (0, \dots, 0, 1)$. Generalizing from this idea, we can construct a sequence $\alpha^{[0]}, \dots, \alpha^{[50]} \in \Delta^{24}$ of parameter vectors representing groups G_8 of increasing degrees of homogeneity by setting

$$\alpha_j^{[i]} := \text{Bin}(24, \frac{i}{50})(\{j\}) = \binom{24}{j} \cdot (\frac{i}{50})^j \cdot (1 - \frac{i}{50})^{24-j} \quad (12)$$

with $i = 0, \dots, 50$ and $j = 0, \dots, 24$, where $\alpha_j^{[i]}$ denotes the j th component of the i th parameter vector and $\text{Bin}(n, p)$ denotes the binomial distribution with parameters n and p . Using the constructed sequence then allows for analyzing the performance of different aggregation function for varying degrees $\alpha^{[i]}$ of homogeneity. Note that, due to Proposition 2, computation of the approximations is possible without computing the whole assessment for every single $\alpha^{[i]}$.

5.2. Discussion of the Results

The results of the study described in Section 5.1 for the similarity measure $Y_S^{u_2}$ are visualized in Figures 1 and 2, while the results for the similarity measures $Y_S^{u_1}$ and $Y_S^{u_3}$ can be found in Figures 3 and 4 in Appendix A4. We will refer to these in our discussion at several points.

Figure 1 shows the credal approximation $M_{G_n}^{u_2}(S)$ (gray shaded region) and the maximum entropy approximation $m_{G_n}^{u_2^*}(S)$ (black line) for different choices of the aggregation function S along an increasing degree of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \dots, 50\}$. Specifically, an abscissa value of i indicates that the information on the group G_8 is given by $\delta_8 \sim \text{Cat}(\alpha^{[i]})$. The ordinate ranges from 0 to 288, where 0 is the minimal and 288 is the maximal attainable expected similarity value.

Clearly, for all investigated aggregation functions the width of the intervals $M_{G_n}^{u_2}(S)$ does depend on the homogeneity of the group, where very homogeneous and very inhomogeneous groups tend to produce more narrow intervals than groups of medium homogeneity. Moreover, for all functions except instant-runoff, both the upper expected similarity $\overline{M}_{G_n}^u(S)$ and the maximum entropy approximation $m_{G_n}^{u_2^*}(S)$ strictly increase along increasing homogeneity. Contrarily, the lower expected similarity $\underline{M}_{G_n}^u(S)$ does not strictly increase along homogeneity for mean rank, dictatorship, Kemeny's rule, and commonality sharing, while it does for the other functions (the same is true if analysis is based on $Y_S^{u_1}$ or $Y_S^{u_3}$). Finally, the results show that Kemeny's rule, commonality sharing, and Condorcet's method produce rather narrow credal approximations compared to the other functions, whereas dictatorship, Coombs' rule, and instant-runoff lead to rather wide intervals. Note that the comparison of the width of the credal intervals provides highly relevant information as it indicates how sensitively the evaluation of an aggregation procedure reacts to choosing one (possibly wrong) approximating measure from the credal set: If we mistakenly evaluate an aggregation function by using maximum entropy, the average error we make will be higher for aggregation functions whose evaluation reacts very sensitively to the choice of a representative.

Figure 2 consists of two different graphs: The upper graph compares the lower expected similarity $\underline{M}_{G_n}^{u_2}(S)$ along an increasing degree of homogeneity $\alpha^{[i]}$ for all aggregation procedures. With the exception of very inhomogeneous groups ($i \leq 2$) where dictatorship exceeds Kemeny's rule, it shows that Kemeny's rule and commonality sharing outperform all other methods independent of the underlying $\alpha^{[i]}$. Comparing commonality sharing with Kemeny's rule, we see that the latter exceeds the former for groups with $i \geq 6$, while the opposite is the case for inhomogeneous groups with $i \leq 5$. Further, it turns out that for rather inhomogeneous groups ($i \leq 18$), choosing a dictatorship performs better than all other methods except commonality sharing and Kemeny's rule. For medium to perfect homogeneity ($i > 18$), however, Condorcet's rule shows the third best performance, for high homogeneity ($i \geq 35$), very closely followed by mean rank and Coombs' rule. Moreover, it is interesting to note that instant-runoff is outperformed by all other aggregation functions independent of the underlying $\alpha^{[i]}$. For $Y_S^{u_1}$ and $Y_S^{u_3}$ (see Appendix A4), the main difference is that the lower expected similarity values $\underline{M}_{G_n}^{u_1}(S)$ and $\underline{M}_{G_n}^{u_3}(S)$ of Kemeny's rule exceed those of the other functions, independent of $\alpha^{[i]}$.

Comparing the maximum entropy approximations $m_{G_n}^{u_2^*}(S)$ in the lower graph gives a similar picture. Again, commonality sharing and Kemeny's rule, whose (numerically) coinciding approximations cannot be distinguished, are superior to all the other aggregation methods. However, mean rank aggregation now outperforms Condorcet's method independent of homogeneity and is already superior to a dictatorship for groups with a rather low homogeneity value ($i \geq 10$). For homogeneous groups ($i \geq 30$), all methods except dictatorship and instant-runoff show a very similar performance. Again, instant-runoff is outperformed by all other functions.

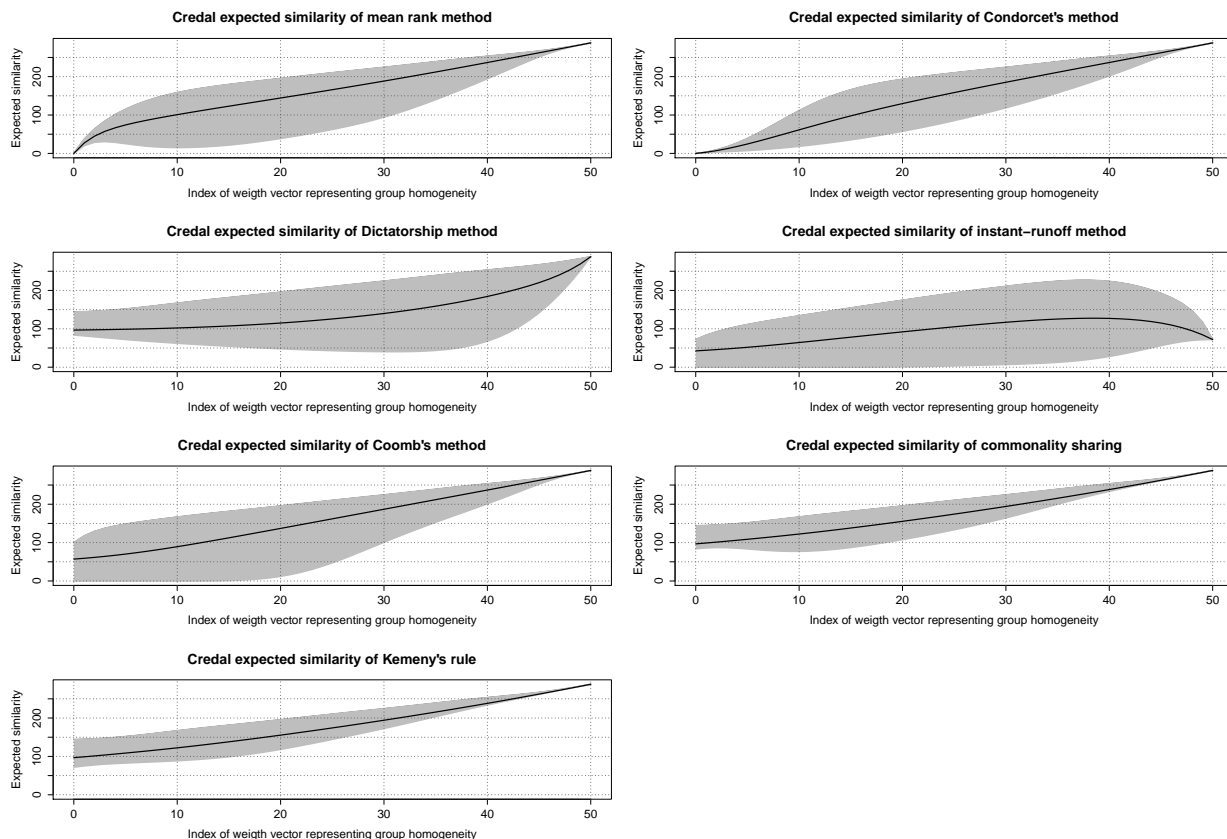


Figure 1: The figures shows the credal approximation $M_{G_n}^{u_2}(S)$ (gray shaded region) and the maximum entropy approximation $m_{G_n}^{u_2^*}(S)$ (black line) for different choices of the aggregation function S along an increasing degree of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \dots, 50\}$ and $u_2(x) = x^2$.

6. Summary, concluding remarks, and discussion

In this paper, we introduced the expected similarity $m_{G_n}^u(S)$ as a theoretical criterion for evaluating and comparing the performance of different preference aggregation functions S if perfect probabilistic information on the homogeneity structure of the group members' preferences is available. Approaches for approximating the true value of $m_{G_n}^u(S)$ under imperfect probabilistic information are fundamentally based on the concept of preference homogeneity measures, for which we gave both a set of minimal requirements and a concrete proposal. Specifically, we studied two different ways to approximate $m_{G_n}^u(S)$: the maximum entropy approximation and the credal set approximation. Finally, by comparing these approximations, we investigated the performance of six common aggregation procedures as well as the recently proposed commonality sharing rule by means of a simulation study for groups along varying degrees of homogeneity. Specifically, we were able to show that the adequateness of a preference aggregation function for a fixed group indeed depends on the group's homogeneity structure.

In future research this framework needs to be applied to real-world data. Particularly, we plan to apply the proposed estimation procedures and evaluation framework to survey data on political opinions and investigate whether groups of significantly differing degrees of homogeneity can be

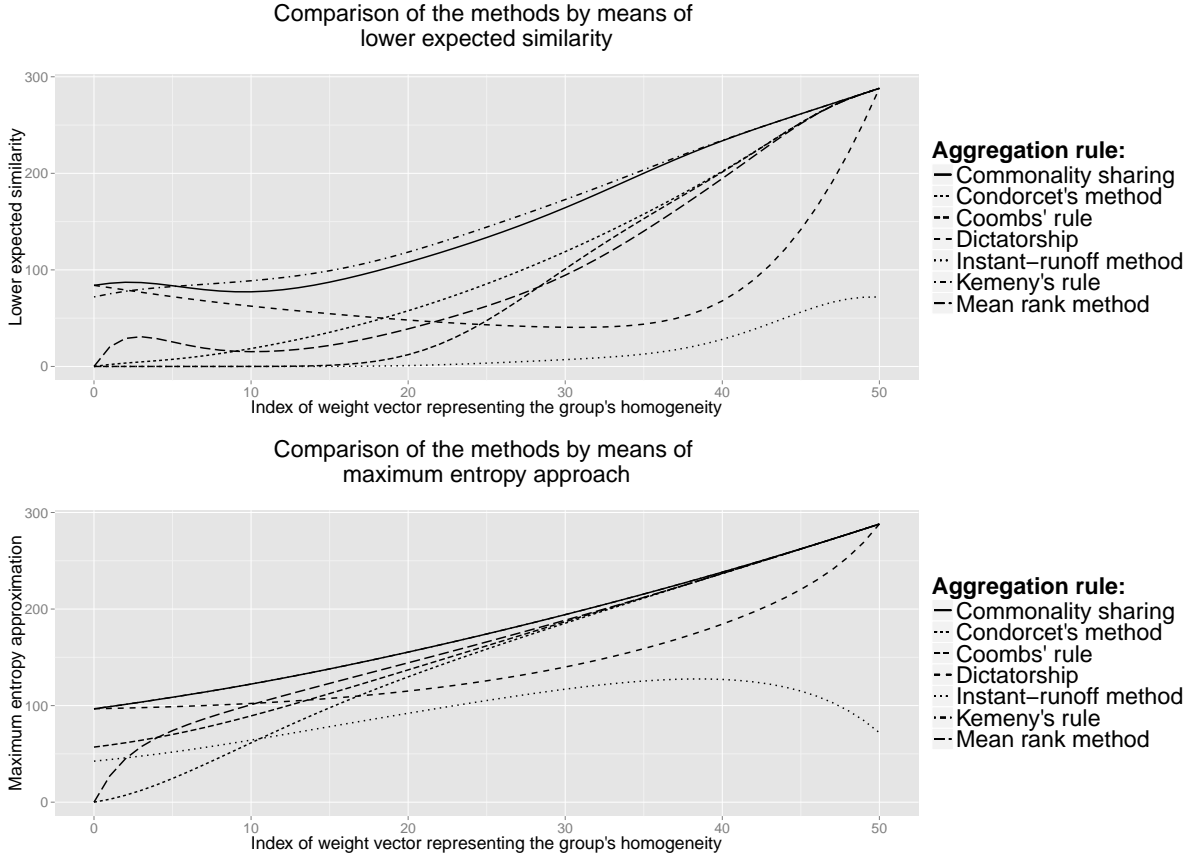


Figure 2: The upper graph shows the lower expected similarity $\underline{M}_{G_n}^{u_2}(S)$ for different choices of the aggregation function S along increasing degrees of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \dots, 50\}$ and $u_2(x) = x^2$. The lower graph shows the same for the maximum entropy approximation $m_{G_n}^{u_2^*}(S)$.

identified in empirical studies. Beyond this, three further extensions seem particularly promising:

Partial individual preferences: The preferences of the group members are modeled by asymmetric and negatively transitive relations $R \in \mathcal{R}$. This explicitly excludes the case of group members judging certain consequences in C incomparable (since incomparability with respect to R is treated as indifference, see Section 2.1 and Footnote 2 in particular). Allowing also for incomparability of consequences could lead to a more realistic model in certain situations. Of course, this requires appropriate adaptations of the aggregation rules from Section 4.

Axiomatic foundations: The conditions that have been listed in Definition 1 are to be understood as minimal requirements for measures of preference homogeneity. However, they are rather weak, as they only look at the profile on a categorical scale. Going beyond the categorical scale in the spirit of Bosch (2006) and Alcalde-Unzu and Vorsatz (2013) could give a more detailed picture of what is actually meant by homogeneity on an axiomatic level.

Efficient algorithms for simulation: In Section 5, we presented a study for a group of $n = 8$ members ranking $|C| = 4$ alternatives. For this setting, the approximations for expected similarity were able to be computed analytically. However, for larger settings this becomes computer-intensive, and computation using simulations has to be applied instead. A proposal for simulation designs

is given in Appendix A2. More efficient designs, comparable to the MCMC-driven approaches already used in the statistical analysis of networks (see, e.g., Hunter et al. (2012)), are planned to be investigated in future research.

Acknowledgement

The authors would like to thank the reviewers and the associate editor for very supportive and helpful comments which improved the paper considerably.

References

- Abrams, R. (1976). The voter’s paradox and the homogeneity of individual preference orders. *Public Choice*, 26(1):19–27.
- Alcalde-Unzu, J. and Vorsatz, M. (2013). Measuring the cohesiveness of preferences: an axiomatic analysis. *Social Choice and Welfare*, 41(4):965–988.
- Aleskerov, F., Karabekyan, D., Sanver, M. R., and Vyacheslav, Y. (2012). On the manipulability of voting rules: the case of 4 and 5 alternatives. *Mathematical Social Sciences*, 64(1):67–73.
- Ali, A. and Meila, M. (2012). Experiments with Kemeny ranking: what works when? *Mathematical Social Sciences*, 64(1):28–40.
- Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors (2014). *Introduction to Imprecise Probabilities*. Wiley.
- Bartholdi, J., Tovey, C., and Trick, M. (1989). Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6(2):157–165.
- Berg, S. (1985). Paradox of voting under an urn model: the effect of homogeneity. *Public Choice*, 47(2):377–387.
- Berger, J. O. (1980). *Statistical Decision Theory and Bayesian Analysis*. Springer.
- Bernard, J. (2005). An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2–3):123–150.
- Bosch, R. (2006). *Characterizations on Voting Rules and Consensus Measures (Ph.D. dissertation)*. Tilburg University, The Netherlands.
- Brams, S. and Fishburn, P. (2002). Voting procedures. In Arrow, K., Sen, A., and Suzumura, K., editors, *Handbook of Social Choice and Welfare, Vol. 1*, pages 173–236. North-Holland.
- Can, B., Ozkes, A. I., and Storcken, T. (2015). Measuring polarization in preferences. *Mathematical Social Sciences*, 78:76–79.
- Coombs, C. H. and Cohen, J. L. (1984). Social choice observed: five presidential elections of the American Psychological Association. *The Journal of Politics*, 46(2):479–502.
- de Borda, J. C. (1781). *Memoire sur les elections au scrutin*. Historie de l’Academie Royale des Sciences.
- de Condorcet, N. (1785). *Essai sur l’application de l’analyse a la probabillite des decisions rendues a la pluralite des voix*. Paris.
- de Finetti, B. (1974). *Theory of Probability*. Wiley.
- Diss, M., Louichi, A., Merlin, V., and Smaoui, H. (2012). An example of probability computations under the IAC assumption: The stability of scoring rules. *Mathematical Social Sciences*, 64(1):57–66.
- Duclos, J.-Y., Esteban, J., and Ray, D. (2004). Polarization: concepts, measurement, estimation. *Econometrica*, 72(6):1737–1772.
- Dwork, C., Kumar, R., Naor, M., and Sivakumar, D. (2001). Rank aggregation methods for the web. In Shen, V., Saito, N., Lyu, M. R., and Zurko, M. E., editors, *Proceedings of the 10th International Conference on World Wide Web*, pages 613–622. ACM.
- Esteban, J.-M. and Ray, D. (1994). On the measurement of polarization. *Econometrica*, 62(4):819–851.
- Fishburn, P. C. (1973). Voter concordance, simple majorities, and group decision methods. *Behavioral Science*, 18(5):364–376.
- Fligner, M. A. and Verducci, J. S. (1986). Distance-based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):359–369.
- Gehrlein, W. V. (1981). Non-profile specific measures of social homogeneity. *Proceedings of the Southeast American Institute for Decision Sciences*, pages 167–169.

- Gehrlein, W. V. and Lepelley, D. (2010). *Voting Paradoxes and Group Coherence: The Condorcet Efficiency of Voting Rules*. Springer.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman and Hall.
- Grofman, B. and Feld, S. (2004). If you like the alternative vote (a.k.a. the instant runoff), then you ought to know about the Coombs rule. *Electoral Studies*, 23:641–659.
- Hare, T. (1857). *The Machinery of Representation*. Maxwell.
- Hartmann, S. and Sprenger, J. (2012). Judgment aggregation and the problem of tracking the truth. *Synthese*, 187(1):209–221.
- Hunter, D. R., Krivitsky, P. N., and Schweinberger, M. (2012). Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics*, 21(4):856–882.
- Jamison, D. and Luce, E. (1972). Social homogeneity and the probability of intransitive majority rule. *Journal of Economic Theory*, 5(1):79–87.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.
- Johnson, N. L. and Kotz, S. (1977). *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*. Wiley.
- Kemeny, J. and Snell, J. (1962). *Mathematical Models in the Social Sciences*. Ginn and Company.
- Kemeny, J. G. (1959). Mathematics without numbers. *Quantity and Quality*, 88(4):577–591.
- Kendall, M. G. and Smith, B. B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, 10(3):275–287.
- Kreps, D. M. (1988). *Notes on the theory of choice (Underground Classics in Economics)*. Westview Press Incorporated.
- Landes, J. and Williamson, J. (2013). Objective Bayesianism and the maximum entropy principle. *Entropy*, 13:3528–3591.
- Lepelley, D. and Valognes, F. (2003). Voting rules, manipulability and social homogeneity. *Public Choice*, 116(1-2):165–184.
- Levi, I. (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT press.
- Maniqueta, F. and Mongin, P. (2016). A theorem on aggregating classifications. *Mathematical Social Sciences*, 79:6–10.
- Mersmann, O., Preuss, M., Trautmann, H., Bischl, B., and Weihs, C. (2015). Analyzing the BBOB results by means of benchmarking concepts. *Evolutionary Computation Journal*, 23(1):161–185.
- Niemi, R. G. (1969). Majority decision-making with partial unidimensionality. *American Political Science Review*, 63(2):488–497.
- Pritchard, G. and Slinko, A. (2006). On the average minimum size of a manipulating coalition. *Social Choice and Welfare*, 27(2):263–277.
- Rosenkrantz, R. D. (1977). *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*. Springer.
- Schollmeyer, G. (2017). Application of lower quantiles for complete lattices to ranking data: analyzing outlyingness of preference orderings. *Technical Report 208, Department of Statistics, LMU Munich (2017)*. Available under: <https://epub.ub.uni-muenchen.de/40452/>.
- Shannon, C. E. (1949). A mathematical theory of communication data. *The Bell System Technical Journal*, 27:379–423.
- Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.
- Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):3–57.
- Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeinen Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept*. Physica.
- Young, H. P. and Levenglick, A. (1978). A consistent extension of Condorcet’s election principle. *SIAM Journal of Applied Mathematics*, 35:285–300.
- Yule, G. U. and Kendall, M. G. (1924). *An introduction to the theory of statistics*. C. Griffin.

Appendix

A1: Proof of Proposition 1

First, note that the definition of δ_n does not depend on the choice of $R_0 \in \mathcal{R}$ with $\sim_{R_0} = \text{diag}(C^2)$, since every such relation contains exactly one of the pairs (a, b) and (b, a) for all $a, b \in C$ with $a \neq b$ and summation is commutative. Moreover, one easily verifies that $\text{Im}(\delta_n) \subset [0, 1]$. Hence, δ_n is well-defined.

(S1): We have to show that 1 is attained iff the profile consists of identical orders. Obviously, δ_n equals 1 for identical profiles by construction. In contrast, if $\underline{R} := (R_1, \dots, R_n) \in \mathcal{R}^n$ is a non-identical profile, a pair $(a, b) \in R_0$ exists such that $\max\{c_{\underline{R}}(a, b), c_{\underline{R}}(b, a), e_{\underline{R}}(a, b)\} < n$. This gives $\delta_n(\underline{R}) < 1$.

(S2): Let $\phi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ be a bijective map and $\underline{R} := (R_1, \dots, R_n) \in \mathcal{R}^n$. By definition, we have $c_{(R_1, \dots, R_n)}(a, b) = c_{(R_{\phi(1)}, \dots, R_{\phi(n)})}(a, b)$ and $e_{(R_1, \dots, R_n)}(a, b) = e_{(R_{\phi(1)}, \dots, R_{\phi(n)})}(a, b)$. This implies $\delta_n(R_1, \dots, R_n) = \delta_n(R_{\phi(1)}, \dots, R_{\phi(n)})$.

(S3): Let $\underline{R} := (R_1, \dots, R_n) \in \mathcal{R}^n$ be a preference profile such that exactly $k \in \{\lfloor \frac{n}{2} \rfloor, \dots, n-1\}$ group members share identical preferences. Without loss of generality, we assume it holds that $R_1 = \dots = R_k =: R^*$ (otherwise we can rearrange the profile in this way due to (S2)). For all distinct pairs $(a, b) \in C^2$ it then holds that $f_{\underline{R}}(a, b) := \max\{c_{\underline{R}}(a, b), c_{\underline{R}}(b, a), e_{\underline{R}}(a, b)\} \geq k$, since each pair (a, b) is identically ranked within the orders R_1, \dots, R_k . Choose an arbitrary index $j_0 \in \{k+1, \dots, n\}$ and define $\underline{Q} := (Q_1, \dots, Q_n)$ to be the profile that arises from \underline{R} by exchanging order R_{j_0} with order R^* . We show that $\delta_n(\underline{R}) \leq \delta_n(\underline{Q})$. Therefore, let $a_0, b_0 \in C$, $a_0 \neq b_0$ be arbitrary but fixed. We distinguish two cases:

Case 1: $f_{\underline{R}}(a_0, b_0) = k$. Clearly, this implies $f_{\underline{Q}}(a_0, b_0) = k+1 > k = f_{\underline{R}}(a_0, b_0)$, since (a_0, b_0) is then identically ranked by exactly Q_1, \dots, Q_k and Q_{j_0} .

Case 2: $f_{\underline{R}}(a_0, b_0) > k$. For arbitrary but fixed $R \in \mathcal{R}$ and $a, b \in C$, $a \neq b$, define the expression $R_{\{a,b\}} := \{(x, y) : x, y \in \{a, b\} \wedge (x, y) \in R\}$. We then distinguish two sub-cases:

Sub-case 1: $\forall j \in \{k+1, \dots, n\} : R_{\{a_0, b_0\}}^* \neq (R_j)_{\{a_0, b_0\}}$.

This implies that $(R_{j_1})_{\{a_0, b_0\}} = (R_{j_2})_{\{a_0, b_0\}}$ for all $j_1, j_2 \in \{k+1, \dots, n\}$ (and that $k = \lfloor \frac{n}{2} \rfloor$ and n is odd), since otherwise $f_{\underline{R}}(a_0, b_0) > k$ would not be possible. Hence, the pair (a_0, b_0) is ranked identically by $\lfloor \frac{n}{2} \rfloor + 1$ members and, therefore, we have $f_{\underline{R}}(a_0, b_0) = \lfloor \frac{n}{2} \rfloor + 1$. However, it also holds that $f_{\underline{Q}}(a_0, b_0) = \lfloor \frac{n}{2} \rfloor + 1$, since (a_0, b_0) is identically ranked by exactly Q_1, \dots, Q_k and Q_{j_0} .

Sub-case 2: $\exists j \in \{k+1, \dots, n\} : R_{\{a_0, b_0\}}^* = (R_j)_{\{a_0, b_0\}}$.

Then, if $R_{\{a_0, b_0\}}^* = (R_{j_0})_{\{a_0, b_0\}}$ we have $f_{\underline{R}}(a_0, b_0) = f_{\underline{Q}}(a_0, b_0)$, and if $R_{\{a_0, b_0\}}^* \neq (R_{j_0})_{\{a_0, b_0\}}$ we have $f_{\underline{R}}(a_0, b_0) < f_{\underline{Q}}(a_0, b_0)$. In either case, we have $f_{\underline{R}}(a_0, b_0) \leq f_{\underline{Q}}(a_0, b_0)$.

Thus, we have shown that, in every case, it holds that $f_{\underline{R}}(a_0, b_0) \leq f_{\underline{Q}}(a_0, b_0)$. Since the pair (a_0, b_0) was chosen arbitrarily, this implies $\delta_n(\underline{R}) \leq \delta_n(\underline{Q})$, completing the proof. \square

A2: A possible simulation design for larger settings

The simulation can be done in the following way: Draw N random samples $\underline{R}_1, \dots, \underline{R}_N$ from the space of profiles \mathcal{R}^n . For all $j = 1, \dots, \xi$, define the set $N_j := \{\underline{R}_i : A_n(\underline{R}_i) = k_j\}$ of all samples mapped to homogeneity class k_j . For given weights $\alpha := (\alpha_1, \dots, \alpha_\xi)$ and aggregation function S , we use the characterization of the assessments given in Proposition 2 and receive

$$\begin{aligned}
m_{G_n}^{u*}(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \frac{1}{|A_n^{-1}(k_j)|} \sum_{\underline{R} \in A_n^{-1}(k_j)} Y_S^u(\underline{R}) \right) \approx \sum_{j=1}^{\xi} \left(\alpha_j \cdot \frac{1}{|N_j|} \sum_{\underline{R} \in N_j} Y_S^u(\underline{R}) \right) \\
\underline{M}_{G_n}^u(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \min_{\underline{R} \in A_n^{-1}(k_j)} Y_S^u(\underline{R}) \right) \approx \sum_{j=1}^{\xi} \left(\alpha_j \cdot \min_{\underline{R} \in N_j} Y_S^u(\underline{R}) \right) \\
\underline{M}_{G_n}^u(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \max_{\underline{R} \in A_n^{-1}(k_j)} Y_S^u(\underline{R}) \right) \approx \sum_{j=1}^{\xi} \left(\alpha_j \cdot \max_{\underline{R} \in N_j} Y_S^u(\underline{R}) \right)
\end{aligned}$$

However, note that this simulation design requires a sample satisfying the condition $N_j \neq \emptyset$ for all $j = 1, \dots, \xi$, i.e., the sample needs to be rich enough that every homogeneity class has been met at least once. Consequently, such a design becomes computationally intensive as n and $|C|$ increase.

A simulation design producing fewer computational costs can be realized by taking advantage of the fact that the maps A_n and Y_S are invariant under permutations of the inserted profile. Let Φ denote the set of all bijective maps $\phi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. For $\underline{R} := (R_1, \dots, R_n) \in \mathcal{R}^n$ and $\phi \in \Phi$, we set $\underline{R}_\phi := (R_{\phi(1)}, \dots, R_{\phi(n)})$ and define an equivalence relation \sim_Φ on \mathcal{R}^n by setting

$$\underline{R} \sim_\Phi \underline{Q} \quad :\Leftrightarrow \quad \exists \phi \in \Phi : \underline{R} = \underline{Q}_\phi$$

Moreover, let $\mathcal{R}_{\sim_\Phi}^n$ denote the quotient space produced by \sim_Φ and let $f : \mathcal{R}_{\sim_\Phi}^n \rightarrow \mathcal{R}^n$ be any choice function satisfying $f(\mathcal{C}) \in \mathcal{C}$ for all $\mathcal{C} \in \mathcal{R}_{\sim_\Phi}^n$. Further, for every possible homogeneity value k_j , where $j = 1, \dots, \xi$, we define the set $\mathcal{L}_j := \{\mathcal{C} \in \mathcal{R}_{\sim_\Phi}^n : A_n(f(\mathcal{C})) = k_j\}$ of all equivalence classes with members that are mapped to k_j . Due to Proposition 2 and the fact that both A_n and Y_S^u are constant on every $\mathcal{C} \in \mathcal{R}_{\sim_\Phi}^n$ (as they are invariant under permutations of the inserted profile), one easily verifies the following identities:

$$\begin{aligned}
m_{G_n}^{u*}(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \frac{\sum_{\mathcal{C} \in \mathcal{L}_j} Y_S^u(f(\mathcal{C})) \cdot |\mathcal{C}|}{\sum_{\mathcal{C} \in \mathcal{L}_j} |\mathcal{C}|} \right) \\
\underline{M}_{G_n}^u(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \min_{\mathcal{C} \in \mathcal{L}_j} Y_S^u(f(\mathcal{C})) \right) \\
\underline{M}_{G_n}^u(S) &= \sum_{j=1}^{\xi} \left(\alpha_j \cdot \max_{\mathcal{C} \in \mathcal{L}_j} Y_S^u(f(\mathcal{C})) \right)
\end{aligned}$$

Using the above identities allows the application of a similar simulation design as proposed above, however, instead of drawing samples from the space \mathcal{R}^n , we can now sample from the smaller space $\mathcal{R}_{\sim_\Phi}^n$. In our context, this means we can sample from the space of all n -combinations of \mathcal{R} instead of sampling from the space of n -permutations of \mathcal{R} .

A3: *Non-locality of the maximum consensus homogeneity δ_n*

The measure δ_n is not local as it cannot be represented as an average similarity of pairs of orders: For a counterexample, consider the profile $\underline{R} = (R_1, R_1, R_1, R_2, R_2)$ on $C = \{a, b, c\}$, where R_1 ranks $a b c$ and R_2 ranks $c b a$. An arbitrary homogeneity measure h_n based on average pairwise similarities would satisfy $h_5(\underline{R}) = \frac{1}{10}[3 \cdot h_2(R_1, R_1) + 6 \cdot h_2(R_1, R_2) + h_2(R_2, R_2)]$. However, the maximum homogeneity measure δ_n does not satisfy this identity, since we have $\delta_5(\underline{R}) = 0.6$, but at the same time $\frac{1}{10}[3 \cdot \delta_2(R_1, R_1) + 6 \cdot \delta_2(R_1, R_2) + \delta_2(R_2, R_2)] = 0.7$.

Note that, despite their locality, homogeneity measures based on pairwise distances of relations satisfy the minimal requirements listed in Definition 1. Conditions (S1) and (S2) are trivially true. For (S3), the triangle inequality can be used: If an order R changes from a non-majority order to the majority order R^* then the distances $d(R, R^*)$ change to $d(R^*, R^*) = 0$ and the distances $d(R, R_j)$ from R to non-majority orders R_j change to $d(R^*, R_j)$. With $d(R^*, R_j) \leq d(R^*, R) + d(R, R_j)$ we get $d(R, R^*) \geq d(R^*, R_j) - d(R, R_j)$. Since in addition to the order R that changes to R^* there are as least as many majority orders as non-majority orders, we can match every increase in distance associated with a pair (R, R_j) to a decrease associated to $d(R, R^*)$ that is greater or equal, so the overall change in the sum of all distances can only be a decrease or zero.

A4: Results for the linear function $u_1(x) = x$ and the concave function $u_3(x) = 36 - (6 - x)^2$

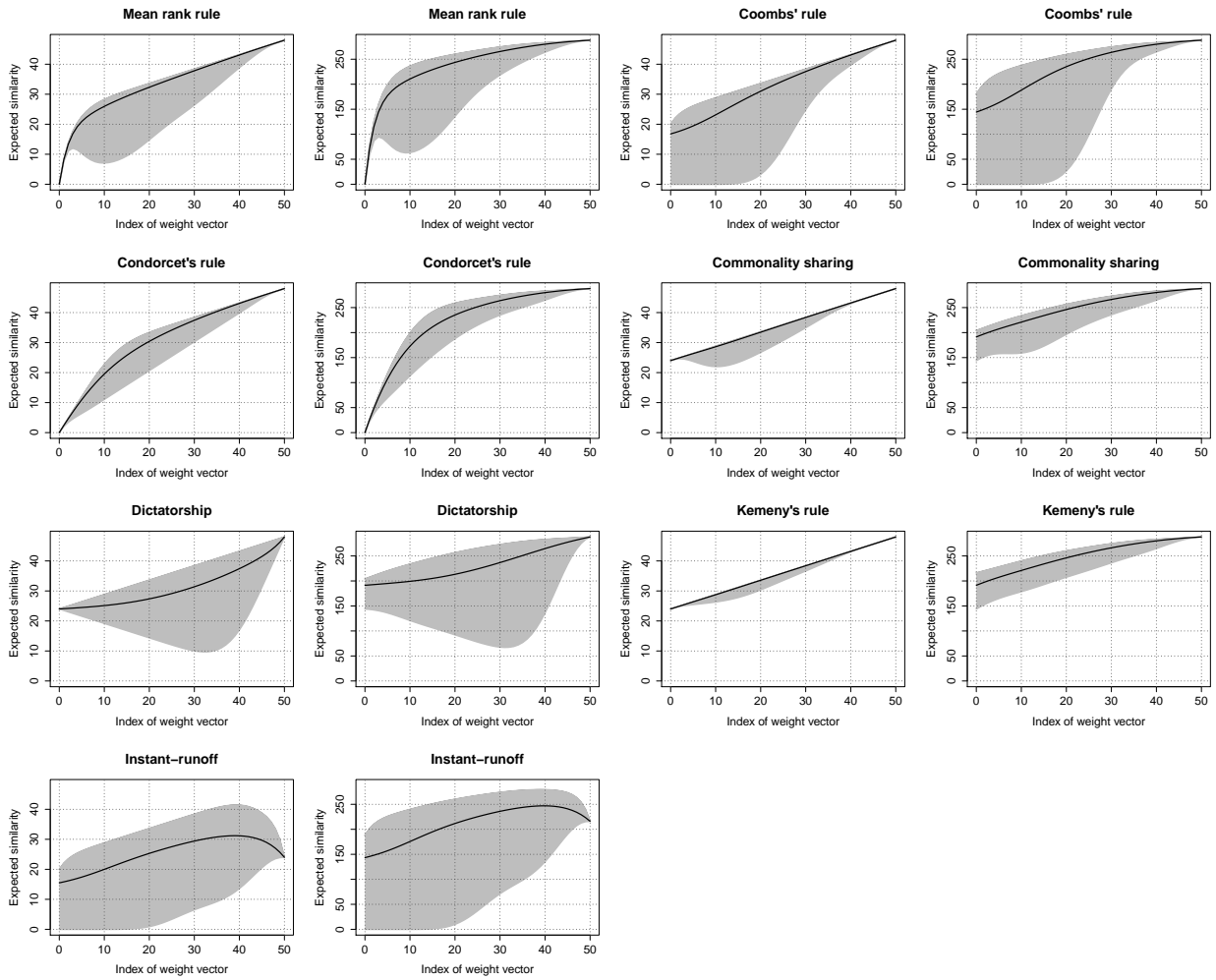


Figure 3: The graphs show the credal approximation $M_{G_n}^u(S)$ (gray shaded region) and the maximum entropy approximation $m_{G_n}^{u*}(S)$ (black line) for different choices of the aggregation function S along an increasing degree of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \dots, 50\}$. For the same S , the left graph corresponds to the choice $u(x) = u_1(x) = x$, whereas the right graph corresponds to the choice $u(x) = u_3(x) = 36 - (6 - x)^2$.

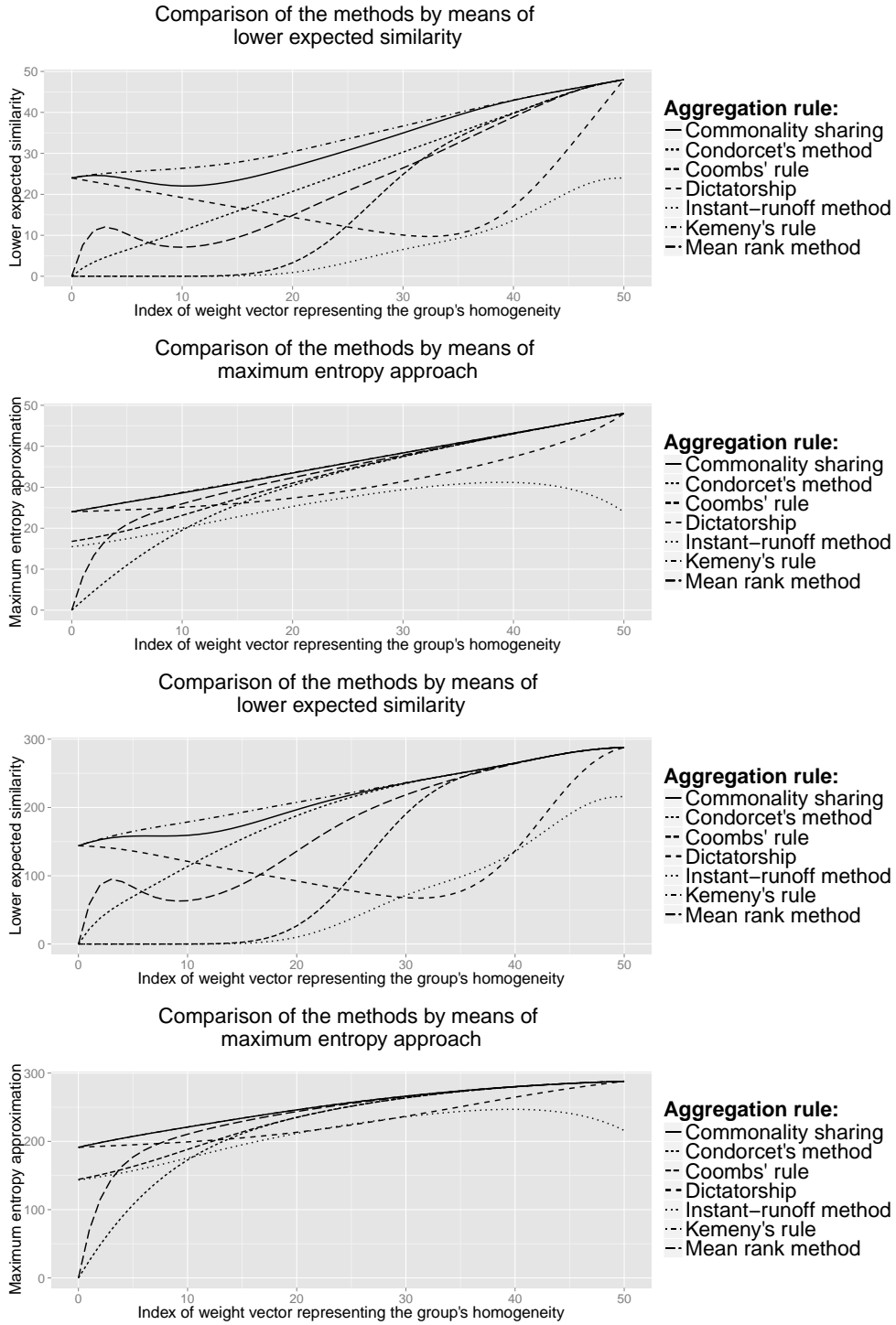


Figure 4: Graphs 1 and 3 show the lower expected similarity $M_{G_n}^u(S)$ for different choices of the aggregation function S along an increasing degree of group homogeneity $\alpha^{[i]}$, where $i \in \{1, \dots, 50\}$ for the choices $u(x) = u_1(x) = x$ and $u(x) = u_3(x) = 36 - (6 - x)^2$. Graphs 2 and 4 show the same for the maximum entropy approximation $m_{G_n}^{u,*}(S)$.