LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Christoph Jansen, Georg Schollmeyer, Thomas Augustin

Probabilistic Evaluation of Preference Aggregation Functions: A Statistical Approach in Social Choice Theory

Technical Report Number 193, 2016
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Probabilistic Evaluation of Preference Aggregation Functions: A Statistical Approach in Social Choice Theory

C. Jansen[a,*], G. Schollmeyer[a], T. Augustin[a]

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

Abstract

A statistical criterion for evaluating the appropriateness of preference aggregation functions for a fixed group of persons is introduced. Specifically, we propose a method comparing aggregation procedures by relying on probabilistic information on the homogeneity structure of the group members' preferences. For utilizing the available information, we give a minimal axiomatization as well as a proposal for measuring homogeneity and discuss related work. Based on our measure, the group specific probability governing the constitution of preference profiles is approximated, either relying on maximum entropy or imprecise probabilities. Finally, we investigate our framework by comparing aggregation rules in a small study.

Keywords: Preference aggregation, preference profile homogeneity, voting theory, imprecise probabilities, maximum entropy. JEL classification: C1, C6

1. Introduction

One of the main issues in *Social Choice Theory* is the question how to combine the individual preferences of the members of a group into one *fair* social preference. Pioneers in facing these kind of problems from a rather mathematical point of view were the French mathematicians Nicolas de Condorcet and Jean-Charles de Borda, who are best known for proposing and investigating two different preference aggregation procedures: *Condorcet's method* and the *Borda count* (see de Condorcet (1785) and de Borda (1781), respectively). However, both methods are known to cause certain disadvantages and, consequently, are not unanimously seen as a completely satisfying solution for the problems of Social Choice. More than 150 years after Condorcet's and Borda's proposals, the economist Kenneth Arrow proposed an axiomatic approach for addressing the problem. Specifically, he formulated three very appealing minimal requirements that every preference aggregation procedure (or *social welfare function* in Arrow's terminology) should satisfy: (1) *unanimity*, (2) *independence of irrelevant alternatives* and (3) *no dictatorship*.[1] Unfortunately, Arrow's impossibility result from 1951 startlingly shows the mutual inconsistency of these three axioms. More precisely, it states: *Every social welfare function that satisfies unanimity and independence*

*Corresponding author

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

[1] Originally, Arrow formulated five axioms. However, if one demands social welfare functions to map n-tuples of arbitrary *weak orders* to *weak orders* again, the other axioms are implicitly satisfied. The three axioms recalled in the text, in some sense, form the core of the Arrovian framework.

of irrelevant alternatives violates no dictatorship (see Arrow (1950)) and, therefore, rules out the possibility of the existence of fair procedures for preference aggregation at least at first sight.

Ever since its publication, much effort has been put on getting around Arrow's impossibility theorem. Very roughly, one can distinguish between two main classes of approach here: (1) Restricting the domain of the aggregation rule on (in some sense) *feasible* preferences (e.g. on *single-peaked* or *single-caved preferences*, see Inada (1964)) or (2) weakening/modifying the axioms by taking into account some notion of *strength* of the individual preferences (e.g. by using value difference functions (see (French, 1988, pp. 280-323)) or preferences in accordance with an expected utility model (see Bacharach (1975))). However, while the first class dismisses certain preferences as being irrational in advance and therefore restricts the free will of the individuals, the second one assumes more than an *ordinal* information on the preferences of the group members involved.[2] Accordingly, both ways appear to be too restrictive for the description of certain situations.

For this reason, we intend to follow a different approach: Taking the (potentially) ordinal structure of the group members' preferences and their free will seriously, we avoid making restricting assumptions concerning the domain of the aggregation rule or some cardinal scale underlying the individual preferences. Instead, we propose a *statistical*[3] performance criterion that allows for comparing the appropriateness of different aggregation procedures for a *fixed group* of individuals under consideration. Specifically, we propose a method for assessing a *group specific* probabilistic model on the space of all preference profiles that is driven by the degree of *homogeneity* of the preferences of the members of the considered group. Applying this model then makes it possible to evaluate and compare different procedures for preference aggregation by utilizing the available information on the *group specific homogeneity structure* of preferences.

The paper is structured as follows: In section 2, we first briefly recall some basic mathematical preliminaries and introduce the notations used in this paper. Afterwards, we successively develop a *statistical* framework for evaluating the appropriateness of a certain preference aggregation procedure for a fixed group under considerations if information on the *homogeneity* structure of the group members' preferences is available. Here, the evaluation framework is statistical in the sense that the performance criterion is based on a *probabilistic* model, which is assessed by utilizing the homogeneity information for the considered group. Section 3 is divided in two parts: In the first part, we briefly recall some well-investigated preference aggregation procedures, namely the *Borda count*, *Condorcet's method*, *Instant-runoff voting*, *Coomb's rule* and *Dictatorship*, and characterize them in terms of preference aggregation functions. In the second part, we summarize a recently proposed aggregation procedure, which addresses the problem by a purely order theoretical approach. In section 4, we apply the introduced evaluation framework in a small group setting. Particularly, we investigate the performance of the aggregation functions recalled in section 3 with respect to the proposed criterion of optimality for groups of varying degree of homogeneity and, subsequently, discuss the results of the study. Section 5 is preserved for concluding remarks as well as an outlook on future research questions.

[2]The demand for ordinality of the individual preferences in the Arrovian framework is mainly coded in the independence axiom. Consequently, most attempts of modification have this axiom as a target.

[3]In contrast, some of the common criteria known from Social Choice Theory (e.g. *avoidance of Condorcet losers*) have a rather axiomatic character and do not allow for using group specific probabilistic information or, in concrete applications, corresponding statistical estimates.

2. A statistical framework for evaluating preference aggregation functions

Within this paper, our goal is to establish a statistical framework for comparing different procedures of *preference aggregation* for *one specific* group of individuals of interest. Specifically, the comparison of the procedures will be based on *expected aggregation quality*, where the probability measure with respect to which the expectation is taken is approximated by using the available information on the homogeneity of preferences of the group under consideration and in concrete application is estimated from data or expert knowledge. Consequently, this makes it necessary to firstly introduce some notion of homogeneity for collections of individual preference orders. Before we can start, some terminology has to be fixed.

2.1. Mathematical preliminaries

Throughout the whole 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 of fixed size $n \geq 2$, where certain requirements of rationality regarding the involved individual orderings are imposed. Precisely, we will work on the following spaces:

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

$$\mathcal{Q} := \{Q \subset C^2 : Q \text{ asymmetric}\} \tag{2}$$

We call every $R \in \mathcal{R}$ a *preference order* on C. Moreover, for every $R \in \mathcal{R}$, we define an equivalence relation $\sim_R$ on C by setting $a \sim_R b$ if and only if $(a,b) \notin R \wedge (b,a) \notin R$. Having fixed this, we interpret $(a,b) \in R$ as *a is strictly preferred to b*, whereas $(a,b) \in \sim_R$ is interpreted 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. For $n \geq 2$, we call every element $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$ a *preference profile* on C, where each component is interpreted as the opinion about the consequences of a member of a group of size n.

In contrast, every element $Q \in \mathcal{Q}$ is called a *consensus order* (or *group preference*). Except of asymmetry, we do not impose any further restricting assumptions on the consensus order.[4] This allows for investigating also aggregation procedures for which the group preference is not always as well-behaving as the individual orders (this includes e.g. *Condorcet's method*, see section 3.1). Given this setting, we call every mapping

$$S : \mathcal{R}^n \to \mathcal{Q} \quad , \quad \underline{R} \mapsto S(\underline{R}) \tag{3}$$

a *preference aggregation function*. Specifically, for every preference profile $\underline{R} \in \mathcal{R}^n$, the image $S(\underline{R}) \in \mathcal{Q}$ gives the consensus order of the group with respect to the aggregation procedure described by S.

2.2. Preference homogeneity in related work

As already foreshadowed in the introduction, our main argument is that the quality of an aggregation procedure S for a fixed group depends on how *homogeneous* the group members tend to be in opinion. In the literature on Social Choice Theory there are mainly two different lines of establishing a notion of homogeneity of groups. The first direction (cf., e.g., Niemi (1969); Jamison

[4] This is an important difference to the Arrovian framework, where the group order has to be transitive.

and Luce (1972); Berg (1985); Gehrlein and Lepelley (2011); Lepelley and Valognes (2003)) consists of establishing a stochastic model that governs the constitution of profiles and has a specific parameter that implicitly regulates the group's homogeneity. One prominent model is the multivariate Pólya-Eggenberger urn model (cf., e.g., Johnson and Kotz (1977)) used in Berg (1985); Gehrlein and Lepelley (2011); Lepelley and Valognes (2003) to analyze the relationship between group homogeneity and the probability of the voting paradox or the manipulability of different aggregation functions. Here, all different possible orderings are coded as differently colored balls in an urn. Then, n balls are drawn, but not simply with or without replacement but instead with the replacement of an amount of σ additional balls of the same color as the color of the last drawn ball. This modified replacement in a way models the interaction between the persons' opinion, and higher values for σ lead to stronger interactions and thus to a higher expected homogeneity of the drawn profile. The Pólya-Eggenberger model contains as special cases also the so called *Impartial Culture* and the *Impartial Anonymous Culture* that are often presumed in studies of the voting paradox and the manipulability of aggregation procedures (cf. Aleskerov et al. (2012); Diss et al. (2012); Pritchard and Slinko (2006)). Note that in the Pólya-Eggenberger model, usually one assumes that the urn is initially composed with exactly one ball for every possible order and thus there is perfect symmetry with respect to all orders. This seems to be a very unrealistic assumption and thus is often criticized, especially in the field of Behavioral Social Choice (see, e.g., Tsetlin et al. (2003) and Regenwetter et al. (2009)).

The second direction of establishing a notion of group homogeneity is to measure directly the homogeneity of a profile by some homogeneity measure. One sort of such measures are the so called non-profile specific measures of homogeneity (cf., Gehrlein (1981)) that use only the categorical scale of the different orders of a profile and are thus based only on the proportions p_i of the different orderings in the profile but not on the orderings themselves. Common measures are quadratic forms in the p_i's, for example Herfindahl's index $H = \sum p_i^2$. The other sort of measures does use the information in the orderings of the profile. Not only in Social Choice Theory, but also e.g. in the field of statistics and computer science (cf., e.g., Fligner and Verducci (1986); Dwork et al. (2001)), there are many concepts based on a geometric understanding that introduces a notion of distance between two orderings. Based on this, it is easy to define a measure of heterogeneity of a group via the average distance of all pairs of orders in the profile. This idea is very similar to the construction of the classical variance of a real-valued vector that can be represented both as the mean of the squared deviations from the mean as well as the average squared distance of all possible pairs of the vector. This type of measures of heterogeneity could be called *local* in the sense that not the whole population is examined simultaneously, but only pairs are considered and afterwards the mean over all pairs is taken. Another measure of homogeneity that was used especially in Social Choice Theory (cf., e.g., Fishburn (1973)) is the measure W of Kendall and Smith (see Kendall and Smith (1939)). This measure intends to analyze the whole population simultaneously by looking at the variance of the vector of the sums of the ranks that the group altogether assigns to the different alternatives. 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 that are assigned by pairs of persons and thus could also be called local. Beyond proposals of concrete homogeneity measures, axiomatic approaches for measures of homogeneity are given in e.g. Bosch (2006); Alcalde-Unzu and Vorsatz (2013), where the terms consensus and cohesiveness are used instead of the term homogeneity. For the concept of polarization, a concept

very similar, but not identical to the concept of heterogeneity,[5] an axiomatic characterization of a measure of polarization of profiles is given in Can et al. (2015).

2.3. A minimal axiomatization for homogeneity of preferences

In section 2.4, we propose a new measure of preference homogeneity, where we intend to utilize the information inherent in the order structure of the profile. Moreover, we aim at analyzing the population as a whole, while the measures mentioned above rather rely on the analysis of pairs of persons or on a solely categorical viewpoint. Against this background, we first formulate some minimal requirements ("axioms") our measure should satisfy:[6]

Definition 1. *A preference homogeneity measure for a group of size n is a map $A_n : \mathcal{R}^n \to [0,1]$, such that the following properties are satisfied:*

(S1) ***Consensus sensitivity****: $A_n(\underline{R}) = 1$ if and only if $\underline{R} = (R^*, \ldots, R^*)$ for some $R^* \in \mathcal{R}$.*

(S2) ***Fair weighing****: Let $\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a bijective map. Then $A_n(R_1, \ldots, R_n) = A_n(R_{\phi(1)}, \ldots, R_{\phi(n)})$ for all $(R_1, \ldots, R_n) \in \mathcal{R}^n$.*

(S3) ***Majority strengthening****: Let $\underline{R} \in \mathcal{R}^n$. Define $k(j) := \{i : R_i = R_j\}$. If there exists $j_0 \in \{1, \ldots, n\}$ such that $n > |k(j_0)| \geq \lfloor \frac{n}{2} \rfloor$, choose $j_1 \in \{1, \ldots, n\} \setminus k(j_0)$ and define $\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ by $\phi(j) = j_0$, if $j \in k(j_0) \cup \{j_1\}$ and $\phi(j) = j$ else. Then we have $A_n(R_1, \ldots, R_n) \leq A_n(R_{\phi(1)}, \ldots, R_{\phi(n)})$.*

Axiom (S1) states that A_n should be sensible for identifying perfect consensus, i.e. its maximal value 1 is attained iff all group members share identical preferences. Axiom (S2) ensures that the homogeneity value of a profile does not depend on the order in which the individuals state their preferences and, therefore, that no individual should have a bigger influence on the homogeneity value than any of the others. Finally, (S3) 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 his mind towards this subgroup, then the homogeneity value of the modified profile should not decrease. Clearly, all three axioms solely rely on the categorical and not on the ordinal structure of the orderings in the profile. Of course, one could also establish a notion of (S3) that uses the ordinal structure by stating for instance that if one ordering R in the profile is changed towards another ordering R' that is more similar to the ordering of the majority, then the homogeneity should not decrease. This is possible but would require a notion of what the terms "majority" and "more similar ordering" then exactly mean. (In the categorical case "majority" and "more similar ordering" straightforwardly translate to "more than the half of the population" and "identical ordering".) Accordingly, the axioms (S1), (S2) and (S3) are to be understood as a minimal characterization of homogeneity leaving definitorical freedom for substance matter considerations.[7]

[5] While the notion of heterogeneity refers here to the diversity of the orders in the profile, polarization means that the orderings 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 knowledge of the authors goes, not yet established for the more elaborated disambiguation between polarization and heterogeneity/inequality in the context of e.g. poverty measurement, see, e.g., Esteban and Ray (1994); Duclos et al. (2004).

[6] Note that the otherwise also compelling measure of Kendall and Smith does not satisfy these axioms, see below.

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

2.4. A proposal for a preference homogeneity measure

Next we propose a concrete measure satisfying the axioms given in Definition 1. The basic idea 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.[8] Let $n > 2$ and let $R_0 \in \mathcal{R}$ with $\sim_{R_0} = \operatorname{diag}(C^2)$ be fixed, such that R_0 always contains *exactly one* of the pairs (a, b) or (b, a) for all $a, b, \in C$ with $a \neq b$. For a fixed preference profile $\underline{R} \in \mathcal{R}^n$ and a fixed pair of consequences $(a, b) \in C^2$, $a \neq b$, we define the expressions

- $c_{\underline{R}}(a, b) := |\{i : (a, b) \in R_i\}|$
- $e_{\underline{R}}(a, b) := |\{i : (a, b) \in \sim_{R_i}\}|$

to be the number of individuals in $\underline{R}$ that prefer a to b and the number of individuals that are indifferent between these options, respectively. Using this, we define $\delta_n : \mathcal{R}^n \to [0, 1]$ by setting

$$\delta_n(\underline{R}) := \frac{\sum_{(a,b)\in R_0} \max\Big\{c_{\underline{R}}(a, b), c_{\underline{R}}(b, a), e_{\underline{R}}(a, b)\Big\}}{n \cdot \binom{|C|}{2}} \tag{4}$$

Note that if all orderings in the profile $\underline{R}$ have no equivalences, then δ_n is always at least 0.5. Additionally, if we define a similarity measure $s(f, g) := \sum_{(a,b)\in C^2} f((a, b)) \cdot g((a, b))$ for valuation functions $f, g : C^2 \to \mathbb{R}$, then δ_n can be written as

$$\delta_n(\underline{R}) = s(f_m, g_c) = \frac{1}{q^2} \sum_{(a,b)\in C^2} f_m((a, b)) \cdot g_c((a, b)) \tag{5}$$

where $f_m((a, b)) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{R_i}((a, b))$ corresponds to the "mean" and $g_c((a, b)) = \mathbb{1}_{\mathrm{CO}(\underline{R})}((a, b))$ corresponds to the Condorcet consensus ordering of the profile, respectively (with $\mathrm{CO}(\underline{R})$ as defined in equation (15)). Additionally, since δ_n sums over all pairs of alternatives it could be called local w.r.t. alternatives, but it is not local w.r.t. persons in the sense that it cannot be represented as an average similarity of pairs of orderings: For a counterexample, consider a profile $\underline{R}$ consisting of three orderings equal to R_1 and two orderings equal to R_2, where $C = \{a, b, c\}$ and aR_1bR_1c and cR_2bR_2a, a homogeneity measure h_n based on average pairwise similarities would automatically satisfy

$$h_5(\underline{R}) = \frac{3 \cdot h_2((R_1, R_1)) + 6 \cdot h_2((R_1, R_2)) + h_2((R_2, R_2))}{10}$$

However, the measure δ_n does not satisfy this identity[9], since we have $\delta_4(\underline{R}) = 0.6$, but

$$\frac{3 \cdot \delta_2((R_1, R_1)) + 6 \cdot \delta_2((R_1, R_2)) + \delta_2((R_2, R_2))}{10} = 0.7$$

Next we explicitly state that our construction is successful; the proposed measure indeed defines a preference homogeneity measure:

[8] Note that the proposed measure is an (indirect) generalization of the *probabilistic distance metric* proposed in Haddaway and Ha (2003) to the case $n > 2$. However, in our context we are interested in homogeneity rather than distance. Another very similar measure is introduced in Can et al. (2015): Here the authors first list up a set of axioms for measures of *polarization* that uniquely characterizes a measure that is closely related to the one used in this work.

[9] Also if one used $2 \cdot (\delta_n - 0.5)$ instead of δ_n this would not change the situation.

Proposition 1. *The mapping δ_n from in equation (5) defines a preference homogeneity measure.*

Proof. The proof of the proposition consists in straightforwardly verifying the axioms (S1), (S2) and (S3) from Definition 1. A detailed proof is given in appendix A1. □

In contrast to δ_n, the homogeneity measure W of Kendall and Smith does not satisfy the *majority strengthening* property (S3): Take for example $C = \{a, b, c, d, e\}$ and $\underline{R} = (R_1, R_1, R_1, R_2, R_3)$ as well as $\underline{R'} = (R_1, R_1, R_1, R_1, R_3)$ with $aR_1bR_1cR_1dR_1e$; $aR_2bR_2cR_2eR_2d$ and $eR_3bR_3cR_3aR_3d$. Then, *majority strengthening* would require $W(\underline{R'}) \geq W(\underline{R})$, but we actually have $W(\underline{R'}) = 0.584 < 0.592 = W(\underline{R})$. Note also that *majority strengthening* is not a too strong requirement since, for example, heterogeneity measures based on average pairwise distances satisfy an analogous property, where a non-decreasing homogeneity translates to a non-increasing heterogeneity.[10]

2.5. A criterion for evaluating preference aggregation functions

In this section, we want to demonstrate how the notion of a preference homogeneity measure can be used in order to determine approximations of the *group specific* probability measure on the space of profiles. Consider the measurable space $(\mathcal{R}^n, 2^{\mathcal{R}^n})$. For a fixed group G_n consisting of n members, we assume the existence of a *true* probability measure $\mathbb{P}_{G_n}$ on this space that is interpreted as follows: For a profile $\underline{R} \in \mathcal{R}^n$, the value $\mathbb{P}_{G_n}(\{\underline{R}\})$ is the probability that the group members of G_n constitute the preference profile $\underline{R}$ at some *randomly chosen* time-point t. The main idea behind this assumption is the following: As time changes, the preferences of the group members may change. However, we assume that there exists a fixed, however latent, *degree of group homogeneity* that remains *stable over time*. Accordingly, the more heterogeneous the group G_n is in principle, the more likely it is that its members will constitute a profile which represents a high degree of heterogeneity. Similarly, very homogeneous groups will tend to constitute profiles representing a high degree of homogeneity.

Since we are interested in the influence of group homogeneity on the appropriateness of a certain aggregation procedure, we first need to introduce a measure for the *quality* of a given preference aggregation function. To express this formally, let S denote an arbitrary, however fixed preference aggregation function. In order to evaluate the aggregation quality of S for a specific profile $\underline{R} \in \mathcal{R}^n$, we define the following, intuitively appealing, mapping:

$$Y_S : \mathcal{R}^n \to \mathbb{R} \quad , \quad \underline{R} \mapsto \sum_{i=1}^{n} |R_i \cap S(\underline{R})| \tag{6}$$

Specifically, the value $\frac{1}{n} \cdot Y_S(\underline{R})$ describes the average number of pairs that is shared by each individual order R_i with the consensus order $S(\underline{R})$ and, accordingly, can be interpreted as a measure for the performance of aggregation rule S for the specific profile $\underline{R}$. Formally, for every individual relation R_i we compute the cardinality of the intersection with the relation $S(\underline{R})$ and, afterwards,

[10] The reason is the triangle inequality: If an ordering R changes from a non-majority ordering to the majority ordering 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 orderings 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 beside the order R that changes to R^* there are as least as many majority orderings as non-majority orderings 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 decreasing or zero.

sum up over all these values.[11]

For the moment, suppose the true $\mathbb{P}_{G_n}$ of group G_n is known. In this case, a straightforward performance criterion for evaluating the performance of an aggregation function S *for the group* G_n under consideration is to take the expectation of the random variable Y_S with respect to the measure $\mathbb{P}_{G_n}$. Clearly, the higher this value is, the higher is the *expected similarity* of the individual preferences contained in $\underline{R}$ with the group preference $S(\underline{R})$. Formally, we arrive at the performance criterion:

$$m_{G_n}(S) := \mathbb{E}_{\mathbb{P}_{G_n}}(Y_S) \tag{7}$$

From a practical point of view, criterion (7) has of course a disadvantage: The true group specific measure will, in general, be unknown, since there simply is no *perfect* probabilistic information available. However, in many applications there will be at least *some* information on the homogeneity structure of the preferences of the group under investigation. In the following section 2.6, we demonstrate how to construct approximations for the true group specific measure $\mathbb{P}_{G_n}$ if the distribution of the chosen homogeneity measure A_n is assumed to be known. Subsequently, we show how these approximations can be used to find assessments for expected similarity $m_{G_n}(S)$.[12] Finally, in section 2.7, we turn to the elicitation of these quantities from expert knowledge and data.

2.6. Constructing assessments for expected similarity

Let A_n denote some fixed preference homogeneity measure on the space of profiles $\mathcal{R}^n$. Moreover, suppose the mapping A_n can *exactly* attain the values $k_1 < k_2 < \cdots < k_\xi \in [0,1]$. For a fixed group G_n consisting of n members, our goal then is to approximate the group specific probability measure $\mathbb{P}_{G_n}$ such that the available knowledge on the distribution of A_n, that is the information concerning the homogeneity structure of the group's preferences, is best possibly utilized. For the moment, suppose we know the vector $\alpha := (a_1, \ldots, \alpha_\xi) \in \Delta^{\xi-1} := \{x \in [0,1]^\xi : \sum_{i=1}^{\xi} x_i = 1\}$ containing the probabilities of the homogeneity values $k_1, \ldots, k_\xi$, i.e. we have $\mathbb{P}_{G_n}(A_n = k_j) = \alpha_j$ for all $j = 1 \ldots, \xi$. 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 operationalized by the chosen preference homogeneity measure A_n.

[11] The idea underlying Y_S could easily be used in order to define a distance function on the space $\mathcal{R}^2$. An axiomatic framework for such distance functions together with a strong representation result is discussed in Kemeny (1959).

[12] At this point, it might be worth taking a little detour: Formally, one can define a preference aggregation function S^* by demanding $S^*(\underline{R}) \in \operatorname{argmax}_{Q \in \mathcal{Q}} \sum_{i=1}^{n} |R_i \cap Q|$ for all $\underline{R} \in \mathcal{R}^n$. By definition, this implies $Y_{S^*}(\underline{R}) \geq Y_S(\underline{R})$ for all $\underline{R} \in \mathcal{R}^n$ for *every* other preference aggregation function S. Consequently, we then have that $\mathbb{E}_{\mathbb{P}}(Y_{S^*}) \geq \mathbb{E}_{\mathbb{P}}(Y_S)$ for no matter which probability measure $\mathbb{P}$. Thus, S^* is optimal w.r.t. the proposed criterion for every group, independently of its specific homogeneity structure. Although this approach, which basically corresponds to *Kemeny's rule* (Kemeny (1959)), has some nice properties and characteristics (see, e.g., Young and Levenglick (1978)), there are, however, strong arguments why S^* might not be satisfactory in certain situations: Far beyond computational aspects referring to the NP-hardness of finding $S^*(\underline{R})$ for a concrete profile $\underline{R}$ (see Bartholdi et al. (1989) for the original result; compare also Ali and Meila (2012) for a comparison of algorithms alleviating the NP-hardness), there is a fundamental problem: choosing S^* very strongly relates to *Goodhart's law*, which roughly states: *When an indicator becomes a target, it ceases to be a good indicator* (see Goodhart (1975) for details). Specifically, the value $Y_S(\underline{R})$ is intended to be an indicator for the *latent* degree of similarity of the orders collected in $\underline{R}$ to the group order $S(\underline{R})$. However, choosing S^* puts pressure on this indicator by optimizing it. If we had chosen another appropriate indicator for similarity in advance, S^* might perform pretty poorly. An axiomatization for similarity as listed as a future research goal in section 5 would allow to judge how serious the impact of the indicator choice would be indeed.

Note that the above assumption naturally gives rise *not* to a single probability on $(\mathcal{R}^n, 2^{\mathcal{R}^n})$, but to a *whole set* of such probabilities, namely the set of all probability measures that are compatible with the available information on the homogeneity structure of the group, which defines a probability over a partition of $\mathcal{R}^n$ as sample space and not $\mathcal{R}^n$ itself. If we take the information seriously, *any* element of this set is a candidate for the true group specific measure and, therefore, an equally plausible choice for approximating it. Principally, there are (at least) two different general approaches for dealing with the *ambiguity* between the compatible measures. The first approach consists in defining criteria for choosing *one* representative from the set of compatible measures and, subsequently, bases all further analyses on the chosen representative. The most prominent example for such a criterion, the *maximum entropy principle* (see Jaynes (1957) or, e.g., Wu (1996)), has its origins in information theory and will be discussed in the following. The second approach treats the set of all compatible measures, also called *credal set*[13] in this context, as an entity on its own: The uncertainty underlying the situation is then described by a whole set of probability measures, not just by one single representative. The argument for doing so is that choosing a representative, for no matter which criterion is applied, necessarily involves a certain degree of arbitrariness. The chosen measure might be rather different from the true one and, accordingly, all inferences based on this representative might be completely misleading. On the other hand, analyses based on the credal set might produce less informative results, however, the results are more credible, since they do not depend on a hardly justifiable reduction of the underlying ambiguity.

We contrast both approaches and the assessments based on them in the sequel:

Maximum entropy approach: According to our assumption, within each homogeneity class $A_n^{-1}(k_j)$, $j = 1, \ldots, \xi$, we are completely vacuous between the probabilities of the contained profiles, whereas the class itself is of known probability α_j. As already foreshadowed above, we then choose that representative in the set of compatible measures that maximizes *Shannon's entropy* (Shannon (1949) for details). The measure that does this job is induced by the assignment

$$\mathbb{P}^*_\alpha(\{\underline{R}\}) := \frac{\alpha_{\phi(A_n(\underline{R}))}}{|A_n^{-1}(A_n(\underline{R}))|} \tag{8}$$

for all $\underline{R} \in \mathcal{R}^n$, where $\phi(k_j) := j$ for $j = 1, \ldots, \xi$. Note that this assignment corresponds exactly to fixing the probability values of our partition given by the equivalence classes $A_n^{-1}(k_j)$, where $j = 1, \ldots, \xi$, and assuming uniform distribution within each $A_n^{-1}(k_j)$. Historically, this quite strong assumption was also often justified by means of the *principle of insufficient reason*, which states that, in absence of evidence, there is no reason for assuming anything different from uniformity.

Credal set approach: As argued above, if we intend to avoid making not fully justifiable assumptions and a possibly misleading approximation of the true probability, we should take into account the set of *all* probability measures on $(\mathcal{R}^n, 2^{\mathcal{R}^n})$ that are compatible with the uncertainty underlying. Consequently, a cautious and therefore more credible approximation is given by the *credal set*

$$\mathcal{M}_\alpha := \Big\{ \pi \in G(\mathcal{R}^n, 2^{\mathcal{R}^n}) : \ \pi(A_n^{-1}(k_j)) = \alpha_j \text{ for all } j = 1, \ldots, \xi \Big\} \tag{9}$$

[13]The name *credal set* is contributed to I. 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, Augustin et al. (2014). For our purposes here, it suffices to note that uncertainty is no longer measured by one single probability, but by the set of all probabilities compatible with the information available.

where $G(\mathcal{R}^n, 2^{\mathcal{R}^n})$ denotes the set of all probability measures on $(\mathcal{R}^n, 2^{\mathcal{R}^n})$. Obviously, the set $\mathcal{M}_\alpha$ contains exactly the probability measures compatible with the class probabilities. In particular, we have $\mathbb{P}_{G_n} \in \mathcal{M}_\alpha$ and $\mathbb{P}^*_\alpha \in \mathcal{M}_\alpha$, that is, both the true measure and its maximum entropy approximation are contained in the credal set.

By using the two approaches just described, the construction of an assessment for the expected similarity value $m_{G_n}(S)$ for a fixed preference aggregation function S can be done pretty straightforwardly: Instead of taking the expectation with respect to the true group specific measure, we take the expectation with respect to the corresponding approximations. Note that in the case of the credal set approach this will lead to interval valued assessment for expected similarity. Formally, this leads to the following two approaches.

Maximum entropy assessment: Compute the expected similarity with respect to the maximum entropy measure $\mathbb{P}^*_\alpha$. We arrive at the following *real-valued* assessment:

$$m^*_{G_n}(S) := \mathbb{E}_{\mathbb{P}^*_\alpha}(Y_S) = \sum\nolimits_{\underline{R}\in\mathcal{R}^n} Y_S(\underline{R}) \cdot \mathbb{P}^*_\alpha(\{\underline{R}\}) \tag{10}$$

Compared to the credal assessment defined below, this assessment clearly has the attractive feature of being represented by *one single* number. Of course, whether one really sees this is an advantage, depends on how convincing one finds the entropy based argument sketched before.

Credal assessment: Compute the expectation with respect to the set $\mathcal{M}_\alpha$, that is the interval of all expectations that are compatible with the probabilistic information available. We arrive at the following *interval-valued* assessment:

$$M_{G_n}(S) := [\underline{M}_{G_n}(S), \overline{M}_{G_n}(S)] := \Big[\inf_{\pi\in\mathcal{M}_\alpha} \mathbb{E}_\pi(Y_S), \sup_{\pi\in\mathcal{M}_\alpha} \mathbb{E}_\pi(Y_S)\Big] \tag{11}$$

Clearly, it holds that $m_{G_n}(S) \in M_{G_n}(S)$ and $m^*_{G_n}(S) \in M_{G_n}(S)$ and thus both the true expected similarity value and its maximum entropy assessment are contained in the credal interval. The smaller the width of the credal interval, the less ambiguity is involved and, consequently, the more reliable is the maximum entropy assessment.

For the computation of the quantities (10) and (11), 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 in our study in section 4. The validity of the proposition can easily be checked by simple computation which is not explicitly given here.

Proposition 2. *For the maximum entropy assessment and the credal assessment defined in (10) and (11), the following equations hold, respectively:*

i) $m^*_{G_n}(S) = \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \frac{1}{|A_n^{-1}(k_j)|} \sum_{\underline{R}\in A_n^{-1}(k_j)} Y_S(\underline{R})\Big)$

ii) $\underline{M}_{G_n}(S) = \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \min_{\underline{R}\in A_n^{-1}(k_j)} Y_S(\underline{R})\Big)$

iii) $\overline{M}_{G_n}(S) = \sum_{j=1}^{\xi}\Big(\alpha_j \cdot \max_{\underline{R}\in A_n^{-1}(k_j)} Y_S(\underline{R})\Big)$ □

2.7. Elicitation of homogeneity class probabilities

In real world applications, the homogeneity class probabilities α will typically be unknown and, accordingly, an estimate $\hat{\alpha} := (\hat{\alpha}_1, \ldots, \hat{\alpha}_\xi)$ has to be obtained. This can be achieved by (at least) three different approaches. Firstly, one can draw on expert knowledge, i.e. ask one or more experts from the investigated field for their probability estimates. If more than one expert is involved, one could receive an estimate by using either an weighted average of the experts' estimates or directly working with the credal set containing all of them.[14] Secondly, one can collect data. For this purpose, we can construct a questionnaire to assess the ordering of $q := |C|$ alternatives by preference at d time-points, or more practically, d parallel items simultaneously as surrogates. Every group member participates in the survey. Then, each item produces a preference profile of the group under consideration and, therefore, we receive a collection of d preference profiles $\underline{R}_1, \ldots, \underline{R}_d$. For each of these profiles we compute the homogeneity measure and receive data $\underline{x} := (x_1, \ldots, x_d)$, where $x_s := A_n(\underline{R}_s)$ for $s = 1, \ldots, d$. We then estimate the class probabilities by computing the relative frequencies, that is

$$\hat{\alpha}_j := \frac{1}{d} \cdot \sum_{s=1}^{d} \mathbb{1}_{\{k_j\}}(x_s) \tag{12}$$

Finally, one can think of a *Bayesian approach*: Note that a preference homogeneity measure $A_n : \mathcal{R}^n \to \mathbb{R}$ defines a categorically distributed random variable taking values in $\{k_1, \ldots, k_\xi\}$. Specifically, since $\alpha_j := \mathbb{P}(A_n = k_j)$ for all $j = 1, \ldots, \xi$, we have that $A_n \sim \mathrm{Cat}(\alpha)$. If we choose a *Dirichlet distribution* with parameter vector $\gamma \in \mathbb{R}_+^\xi$ as a prior for α, formally $\alpha \sim \mathrm{Diri}(\gamma)$, we can use $\underline{R}_1, \ldots, \underline{R}_d$ from above and compute the posterior distribution of α given $\underline{x}$. As the two distribution families are conjugate to each other, the posterior is again a Dirichlet distribution, however, with parameter $\gamma|\underline{x} := (\gamma_1|\underline{x}, \ldots, \gamma_\xi|\underline{x})$, where $\gamma_j|\underline{x} := \gamma_j + \sum_{s=1}^{q} \mathbb{1}_{\{k_j\}}(x_s)$ for $j = 1, \ldots, \xi$.[15] As an estimate for α we then can, for instance, use the posterior expectation given by

$$\hat{\alpha}_j := \frac{\gamma_j|\underline{x}}{\sum_{l=1}^{\xi} \gamma_l|\underline{x}} \tag{13}$$

Clearly, which approach to follow, depends on the situation: If q is large and the homogeneity measure can potentially attain lots of different values, taking the relative frequencies will often fail, since it requires too many data points. In such cases, the Bayesian approach has certain advantages. However, this approach needs to specify a hyperparameter γ. If available, γ can be specified by expert knowledge. If this is not feasible, one instead could choose a near-vacuous prior model like the *Imprecise Dirichlet Model (IDM)* (see Walley (1996) for the original work or Bernard (2005) for an overview). Note that, when it comes to elicitation, the advantages of the proposed framework become perfectly clear: Instead of directly asking experts on 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 $\{k_1, \ldots, k_\xi\}$. Due to its very intuitive interpretation, the distribution α is much easier to elicitate: How homogeneous do you think the considered group is in probability?

[14] For an detailed overview on expert elicitation methodology see for instance (Augustin et al., 2014, Chapter 15).

[15] For further details concerning the *Dirichlet-Categorical Model* (and techniques from Bayesian Statistics in general), see for instance Gelman et al. (2004).

3. Aggregation rules investigated in the study

In this section, we briefly recall five common preference aggregation procedures typically discussed in Social Choice Theory and demonstrate how they straightforwardly can be extended to a preference aggregation function.[16] All aggregation rules to be recalled here are well-investigated concerning their behavior with respect to the axiomatic criteria known from Social Choice Theory (such as for instance the *avoidance of Condorcet losers* or *resistance to manipulability*[17]). Contrarily, in the focus of our investigation is the performance of these rules with respect to the criterion introduced in the previous section 2.5. Afterwards, in section 3.2, we shortly describe a novel aggregation method, recently proposed in Schollmeyer (2016), which is based on a generalized concept of quantiles on complete lattices.

3.1. Full order variants of some common aggregation procedures

In the following, we list five common aggregation rules and briefly describe how these can be extended into a full preference aggregation function.

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 the relation R.[18] The *mean rank* aggregation function is given by

$$\text{MR} : \mathcal{R}^n \to \mathcal{Q} \ , \ \underline{R} \mapsto \text{MR}(\underline{R}) \tag{14}$$

where we have $(a,b) \in \text{MR}(\underline{R})$ iff $\sum_{i=1}^{n} \big(\text{rank}_{R_i}(a) - \text{rank}_{R_i}(b)\big) > 0$. 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. Consequently, two alternatives a and b are equivalent with respect to $\sim_{\text{MR}(\underline{R})}$ iff they achieve coinciding average rank within the profile $\underline{R}$.

Condorcet's method: Another well-investigated aggregation procedure is *Condorcet's method*. The preference aggregation function induced by this rule has the form

$$\text{CO} : \mathcal{R}^n \to \mathcal{Q} \ , \ \underline{R} \mapsto \text{CO}(\underline{R}) \tag{15}$$

where we have $(a,b) \in \text{CO}(\underline{R})$ iff $(c_{\underline{R}}(a,b) > c_{\underline{R}}(b,a) \wedge c_{\underline{R}}(a,b) > e_{\underline{R}}(a,b))$. That is, 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. Here, alternatives a and b are equivalent with respect to $\sim_{\text{CO}(\underline{R})}$ iff either at least half of the group is indifferent between them or equally many persons prefer a before b and vice versa.

Instant runoff (Hare's method):[19] *Instant runoff* is an example for a *sequential* aggregation procedure: In the first step, all alternatives with the fewest number of first place votes are excluded from C. These form the alternatives that are least preferred by the group, 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, which the group prefers second least. Again, between these alternatives the group is indifferent, but each of them

[16] Most of the common aggregation procedures are intended to only select one *best* alternative rather than a *complete ranking*. However, the ideas underlying these often easily extend to a ranking of all alternatives.

[17] See for instance Grofman and Feld (2004) for further details.

[18] Formally, we have $\text{rank}_R(a) := |\{b \in C : (a,b) \in R\}| + \frac{1}{2}|\{b \in C : (a,b) \in \sim_R \wedge a \neq b\}| + 1$

[19] Originally proposed in Hare (1857); for an overview on runoff methods see Colomer (2004).

is preferred to every alternative excluded in the first step. Successively keeping on repeating this procedure, we end up with a set of optimal options having the same number of first place votes. Informally, this describes a preference aggregation function

$$\mathrm{IR} : \mathcal{R}^n \to \mathcal{Q} \quad , \quad \underline{R} \mapsto \mathrm{IR}(\underline{R}) \tag{16}$$

A formal description of the method can be found in appendix A2. Note that two alternatives are equivalent with respect to $\sim_{\mathrm{IR}(\underline{R})}$ iff they are excluded at the same stage of the procedure.

Coomb's rule:[20] The basic idea of *Coomb's rule* is very similar to that of instant runoff voting as it also is 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 instead of the ones with the minimal number of first place votes in every step. We denote the corresponding preference aggregation function by

$$\mathrm{CM} : \mathcal{R}^n \to \mathcal{Q} \quad , \quad \underline{R} \mapsto \mathrm{CM}(\underline{R}) \tag{17}$$

where we have $(a, b) \in \mathrm{CM}(\underline{R})$ iff a is excluded at a later stage than b. Moreover, a and b are equivalent with respect to $\sim_{\mathrm{CM}(\underline{R})}$ iff they are excluded at the same stage. A formal definition of the method can be given analogously as for instant runoff voting (see again appendix A2).

Dictatorship: The idea of the dictatorship aggregation function is pretty simple to state: The whole group has to prefer whatever the dictator prefers. If we for instance nominate group member i_0 to be the dictator, we receive the following preference aggregation function:

$$\mathrm{DI}_{i_0} : \mathcal{R}^n \to \mathcal{Q} \ , \ \underline{R} \mapsto R_{i_0} \tag{18}$$

Clearly, this does not seem like the fairest way of aggregating preferences (at least for individuals other than i_0). Nevertheless, it might be worth investigating how dictatorship performs in comparison to the well-investigated aggregation functions recalled above.

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

The comparison in section 4 will contain, next to the extensions of the well-known rules just discussed, also an aggregation procedure that initially arose from an attempt to generalize concepts of centrality and outlyingness of observations to partially ordered data: *commonality sharing*[21]. Opposed to the other aggregation methods investigated here, this aggregation procedure does not locally look at different alternatives or pairs of alternatives, but takes into account the whole orderings of all persons and embeds these 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 (2016), is used to select one or more orderings of persons who are most centered in the population. Concretely, one applies the following procedure: For a given minimum size k, one looks at every possible sub-population M_k^i consisting of at least k persons. Then, one considers the set Q_k of all persons 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

[20] A discussion of Coomb's rule can be found in Coomb and Cohen (1984), a comparison with Hare's method is done in Grofman and Feld (2004).

[21] For a more detailed description, see Schollmeyer (2016).

common should also be edges of every ordering q_k^j in Q_k). The set Q_k of persons 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 ordering q, the smaller the smallest k such that Q_k still contains q, the more central is the ordering 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 arbitrary from the set Q_k or apply some further aggregation rule to the orderings in Q_k. (In the study of section 4 we apply the first approach.) Note that the *commonality sharing* consensus rule is in fact a non-local rule in the sense that if for example two persons 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 persons interchange with each other their orderings within the set $\{a, b, c\}$ while remaining their orderings within the set $\{d, e, f\}$, so the whole orderings play in fact a role as a whole. This is a main difference to the aggregation rules of section 3.1 (except dictatorship), where e.g. it does not matter from which person an alternative gets its score in the mean rank aggregation or where edges for pairs of alternatives in Condorcets method are counted without differentiating between edges belonging to the same person and edges belonging to different persons. Note also that the commonality sharing rule can be computed in a much simpler way as one would expect from the conceptual description of the aggregation procedure. In fact, computing the commonality sharing rule can be done in $\mathcal{O}(n \cdot |C|^2)$ time.

4. A study for the case $n = 8$ and $|C| = 4$

In this section, we apply the evaluation framework developed in section 2 by means of a study for groups under varying degree of preference homogeneity. More precisely, we investigate to which extent the performance of the preference aggregation functions recalled in section 3 depends on the degree of homogeneity of the group under consideration for the case of a group consisting of $n = 8$ personss that have to rank $|C| = 4$ alternatives.

4.1. Setup of the study

Assume that $|C| = 4$. For sake of simplicity, we assume that within the individual preferences there is no indifference, i.e. we consider only profiles in which every person either strictly ranks a better than b or vice versa whenever $a \neq b$ and $a, b \in C$. For measuring homogeneity, we choose the measure δ_8 from equation (4) restricted to the domain $\mathcal{H}^8 \subset \mathcal{R}^8$, where $\mathcal{H}$ is the set of all relations $R \in \mathcal{R}$ satisfying $\sim_R = \mathrm{diag}(C^2)$. Clearly, in this case the definition of the measure δ_8 given in equation (4) reduces to

$$\delta_8 : \mathcal{H}^8 \to [0, 1] \quad , \quad \underline{R} \mapsto \frac{\sum_{(a,b)\in R_0} \max\Big\{c_{\underline{R}}(a, b), c_{\underline{R}}(b, a)\Big\}}{48} \tag{19}$$

For this setting, we investigate how the six preference aggregation functions recalled in section 3 perform for groups constituting different degrees of preference homogeneity, formalized by different choices of the homogeneity class weights α. Specifically, the comparison of the aggregation procedures is done using both assessment approaches proposed, the maximum entropy approach defined

in equation (8) and the credal set approach defined in equation (9).

First, note that the measure δ_8 takes exactly the values $\frac{24}{48}, \frac{25}{48}, \ldots, \frac{48}{48}$, where a value of $\frac{1}{2}$ indicates minimum homogeneity, whereas a value of 1 indicates perfect homogeneity (and therefore that all group members share identical preferences due to (S1)). Accordingly, it is obvious that the weight vector $\alpha_{min} := (1, 0, \ldots, 0)$ represents a lower degree of group homogeneity than the weight vector $\alpha_{max} := (0, \ldots, 0, 1)$. Generalizing this idea, we can construct a sequence $\alpha_0 := \alpha_{min}, \alpha_1, \ldots, \alpha_{49}, \alpha_{50} := \alpha_{max} \in \Delta^{24}$ of weight vectors representing increasing degree of group homogeneity by setting

$$\alpha_{ij} := \mathrm{Bin}\big(24, \tfrac{i}{50}\big)(\{j\}) = \tbinom{24}{j} \cdot (\tfrac{i}{50})^j \cdot (1 - \tfrac{i}{50})^{24-j} \tag{20}$$

for $i = 0, \ldots, 50$ and $j = 0, \ldots, 24$, where α_{ij} denotes the jth component of the ith weight vector and $\mathrm{Bin}(n, p)$ denotes the binomial distribution with parameters n and p. Using the constructed sequence then allows for analyzing the performance of a preference aggregation function for varying degrees of group consensus, i.e. varying degree of group homogeneity. Due to proposition 2, this is possible without computing the whole assessment for every single weight vector.

4.2. Discussion of the results

The results of the study described in the previous section 4.1 are visualized in the figures 1 and 2. *Figure 1* shows the assessments for expected similarity for all aggregation functions considered in section 3 and all weight vectors $\alpha_0, \ldots, \alpha_{50}$. In each of the pictures, the x-axis ranges from 0 to 50, where an x-value of i represents weight vector α_i. The y-axis ranges from 0 to 48, where 0 is the minimal and 48 is the maximal reachable expected similarity value. The red line shows the maximum entropy assessment for the expected similarity value along increasing degree of group homogeneity. The grey shaded region represents the interval-valued credal assessment. If, for instance, the group's homogeneity is represented by α_{20} the maximum entropy assessment for mean rank aggregation value equals approximately 32, whereas the credal assessment ranges in $[15, 33]$.

Clearly, the width of the interval given by the credal assessment (and therefore the strength of ambiguity underlying the situation) depends on the degree of homogeneity of the group as well as on the choice of the aggregation rule. If we again consider the mean rank rule, we see that for groups with high or low degree of homogeneity, the credal intervals are rather narrow, whereas for groups of medium homogeneity the credal intervals are relatively wide. If we compare mean rank rule with Condorcet's method, we see that for almost every degree of homogeneity the latter involves less ambiguity, since it produces narrower credal intervals. Note that the comparison of the width of the credal intervals constitutes a highly relevant information as it indicates how *sensitive* the evaluation of an aggregation procedure reacts on choosing one (possibly wrong) approximating measure from the credal set: If we wrongfully evaluate an aggregation function by its maximum entropy assessment, the average error we make will be higher for aggregation functions whose evaluation reacts very sensitive to wrong model choice. Taking solely into account this *sensitivity* towards the choice of an inappropriate probabilistic model, we see that the commonality sharing rule performs best, followed by Condorcet's method. Moreover, both dictatorship and mean rank aggregation are medium sensitive towards wrong model choice, whereas Coomb's rule and instant-runoff voting perform pretty poorly.

For larger settings, the esimates need to be computed simulation based. First results of such simulation studies indicated that the ranking of the aggregation procedures remains unchanged. The

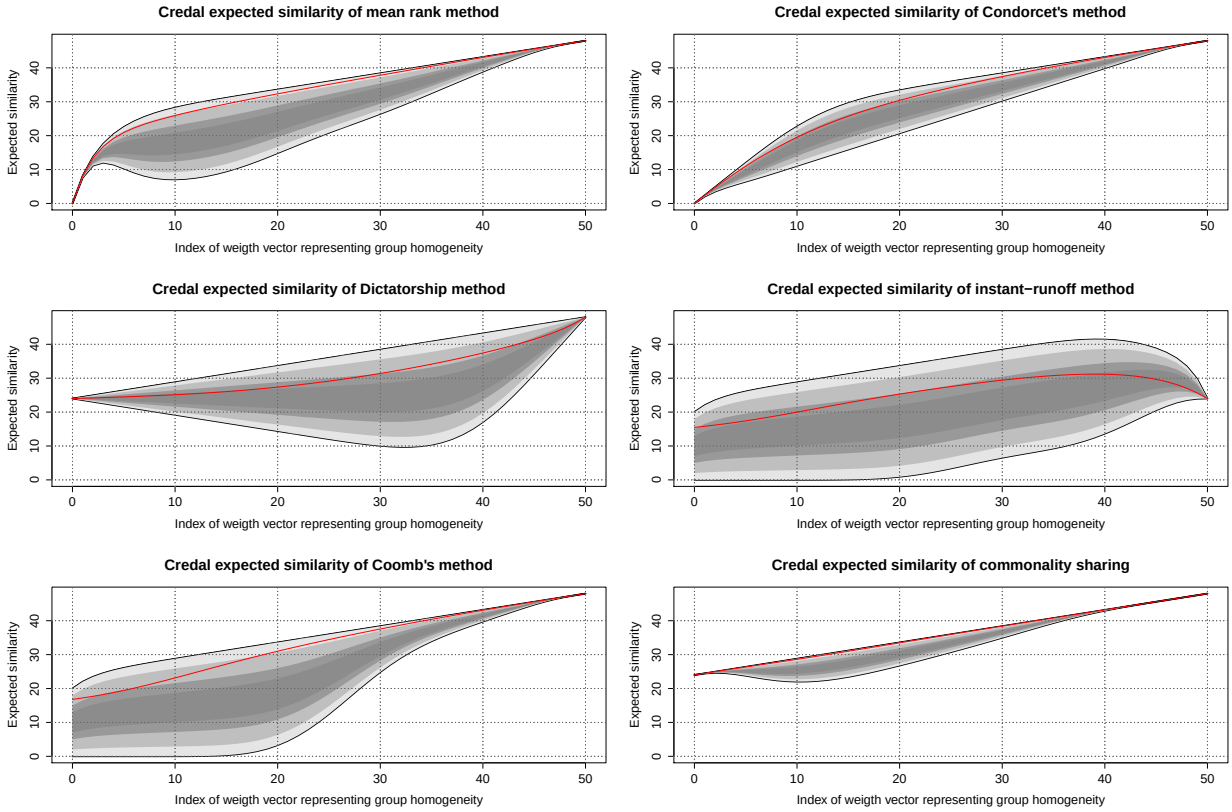

Figure 1: Results for the aggregation procedures from section 3.

exact design of the simulation that was applied is described in appendix A3.

Figure 2 consists of two different pictures: The upper picture shows the lower expected similarity value (the lower bound of the credal interval) along increasing degree of group consensus for all aggregation procedures under consideration. The first (maybe surprising) fact to note is that the commonality sharing rule outperforms all other aggregation rules for *no matter which* weight vector the underlying degree of group consensus is represented by. Further, we see that for rather inhomogeneous groups (represented by weight vectors α_i, where $i \leq 16$) choosing a dictatorship performs better than all other methods except from commonality sharing. For medium to completely homogeneous groups (represented by α_i, where $i \geq 16$), Condorcet's rule turns out to be the second best performing aggregation method behind commonality sharing, however, for very homogeneous groups ($i \geq 30$) very closely followed by mean rank aggregation and Coomb's rule. Moreover, it is interesting to note that Coomb's rule outperforms instant runoff voting for no matter what degree of homogeneity. The same holds true for commonality sharing and dictatorship. Comparing the maximum entropy assessments instead, paints us a very similar picture. Again, commonality sharing is superior to all the other aggregation methods. However, mean rank aggregation now outperforms Condorcet's method and is superior to an dictatorship already for groups with a very low homogeneity value ($i \geq 8$). In contrast to the upper picture, the different methods are closer together and not as easy to compare, when the maximum entropy assessment is considered.

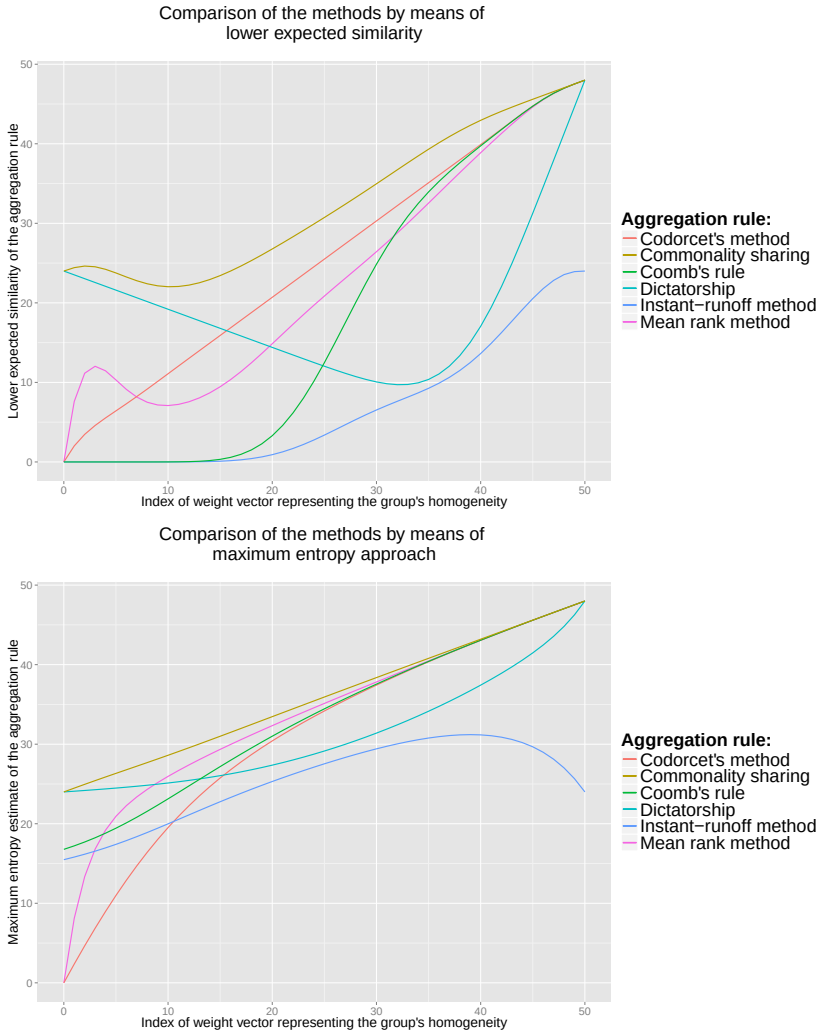

Figure 2: Comparison of the lower expected similarity values (upper picture) and the maximum entropy assessment (lower picture) of the different aggregation functions along increasing degree of group homogeneity.

5. Summary, concluding remarks and discussion

In this paper, we introduced a statistical criterion for evaluating the quality of a preference aggregation function if probabilistic information on the homogeneity structure of the group members' preferences is available. The proposed criterion is fundamentally based on the concept of a preference homogeneity measure, for which we gave both a minimal axiomatization as well as a concrete proposal. Moreover, some conceptual differences of the proposed homogeneity measure and measures already known from literature (such as Kendall's and Smith's W or Herfindahl's index) are discussed. Subsequently, we introduced two different approaches for assessing the value $m_{G_n}(S)$ of the criterion for given group G_n and aggregation function S: The *maximum entropy assessment* is motivated by a criterion originating from information theory, while the *credal assessment* provides an ambiguity-robust approach applying the theory of imprecise probabilities. Comparing these assessments, we investigated the performance of five common aggregation procedures as well as the

recently proposed commonality sharing rule by means of a small study for groups along varying degree of homogeneity. Specifically, we could show that the optimality of a preference aggregation function for a fixed group might indeed depend on the group's homogeneity structure.

Of course, in future research the framework needs to be applied to real world data. Specifically, we plan to try out the proposed elicitation procedures and evaluating framework for survey data on political opinions and investigate whether groups of significantly differing degree of homogeneity can be identified in empirical studies. Additionally, several refinements/extensions/improvements of the evaluation framework presented in this paper can be considered in future research. In the following, we briefly mention the two aspects seeming most promising to us:

Axiomatic foundations: The axioms that have been proposed in section 2.3 are to be understood as *minimal* requirements for measures of preference homogeneity. However, they are rather weak, since 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. Moreover, in the light of the discussion in section 2.5, it should not remain unmentioned that our choice of the similarity measure (6) as an indicator rather relies on intuition and heuristics, still waiting for a rigorous axiomatic justification, where the main difficulties include keeping the definitorical freedom of a pure categorical approach.

Efficient algorithms for simulation: In section 4, we presented a study for a group of $n = 8$ members ranking $|C| = 4$ alternatives. For this setting, the assessments for expected similarity could be computed analytically. However, for larger settings this becomes computer intensive and simulation based assessment has to be applied instead. A proposal for a simulation design is given in appendix A3. More efficient designs, comparable to the MCMC-driven approaches already used in the statistical analysis of networks, are planned to be investigated in future research. This would also allow for much deeper evaluation and comparison of the different aggregation procedures. In particular, one could clarify whether the commonality sharing rule still outperforms the other investigated rules in larger settings.

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.

Arrow, K. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):555–566.

Augustin, T., Coolen, F., de Cooman, G., and Troffaes, M., editors (2014). *Introduction to Imprecise Probabilities*. Wiley and Sons.

Bacharach, M. (1975). Group decisions in the face of differences of opinion. *Management Science*, 22(2):182–191.

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.

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.

Can, B., Ozkes, A. I., and Storcken, T. (2015). Measuring polarization in preferences. *Mathematical Social Sciences*, 78:76–79.

Colomer, J. M., editor (2004). *The Handbook of Electoral System Choice*. Palgrave MacMillan.

Coomb, 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 probabilite des decisions rendues a la pluralite des voix*. Paris.

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.

French, S. (1988). *Decision Theory: An Introduction to the Mathematics of Rationality*. Wiley and Sons.

Gehrlein, W. V. (1981). Non-profile specific measures of social homogeneity. In *Proceedings of the Southeast American Institute for Decision Sciences*, pages 167–169.

Gehrlein, W. V. and Lepelley, D. (2011). Condorcet efficiency and social homogeneity. In Gehrlein, W. V. and Lepelley, D., editors, *Voting Paradoxes and Group Coherence*, pages 157–198. Springer.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman and Hall.

Goodhart, C. A. E. (1975). Problems of monetary management: The UK experience. *Papers in Monetary Economics*, 46:479–502.

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.

Haddaway, P. and Ha, V. (2003). Similarity of personal preferences: theoretical foundations and empirical analysis. *Artificial Intelligence*, 32(2):149–173.

Hare, T. (1857). *The Machinery of Representation*. Maxwell.

Inada, K. (1964). A note on the simple majority decision rule. *Econometrica*, 32(4):525–531.

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. G. (1959). Mathematics without numbers. *Quantity and Quality*, 88(4):577–591.

Kendall, M. G. and Smith, B. B. (1939). Project euclid. *The Annals of Mathematical Statistics*, 10(3):275–287.

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.

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.

Regenwetter, M., Grofman, B., Popova, A., Messner, W., Davis-Stober, C. P., and Cavagnaro, D. R. (2009). Behavioural social choice: a status report. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1518):833–843.

Schollmeyer, G. (2016). Application of lattice-valued quantiles to ranking data: analyzing outlyingness of preference orderings in profiles. working paper, current version available under: `http://www.statistik.lmu.de/~gschollmeyer/working_paper_ranking_data.pdf`.

Shannon, C. E. (1949). A mathematical theory of communication data. *The Bell System Technical Journal*, 27:379–423.

Tsetlin, I., Regenwetter, M., and Grofman, B. (2003). The impartial culture maximizes the probability of majority cycles. *Social Choice and Welfare*, 21(3):387–398.

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.

Wu, N. (1996). *The Maximum Entropy Method*. Springer.

Young, H. P. and Levenglick, A. (1978). A consistent extension of Condorcet's election principle. *SIAM Journal of Applied Mathematics*, 35:285–300.

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 $\sim_{R_0} = \mathrm{diag}(C^2)$, 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 $\mathrm{Im}(\delta_n) \subset [0,1]$. Hence, δ_n is well-defined.

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

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

(S3): Let $\underline{R} := (R_1,\ldots,R_n) \in \mathcal{R}^n$ be any preference profile such that *exactly* $k \in \{\lfloor \frac{n}{2} \rfloor, \ldots, n-1\}$ group members share identical preferences. W.l.o.g., assume it holds that $R_1 = \cdots = R_k =: R^*$ (otherwise we can rearrange the profile in this way due to (S2)). For all pairs $(a,b) \in C^2$ with $a \neq b$ it then holds that

$$f_{\underline{R}}(a,b) := \max\big\{c_{\underline{R}}(a,b), c_{\underline{R}}(b,a), e_{\underline{R}}(a,b)\big\} \geq k \tag{21}$$

since each pair (a,b) is identically ranked within the orders $R_1,\ldots,R_k$. Choose an arbitrary index $j_0 \in \{k+1,\ldots,n\}$ and define $\underline{Q} := (Q_1,\ldots,Q_n)$ to be the profile that arises from $\underline{R}$ by exchanging order R_{j_0} by 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 $Q_1,\ldots,Q_k$ 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: $\forall\, j \in \{k+1,\ldots,n\} : R^*_{\{a_0,b_0\}} \neq (R_j)_{\{a_0,b_0\}}$.

This implies that $(R_{j_1})_{\{a_0,b_0\}} = (R_{j_2})_{\{a_0,b_0\}}$ for all $j_1, j_2 \in \{k+1,\ldots,n\}$ (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 $Q_1,\ldots,Q_k$ and Q_{j_0}.

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

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 showed 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. □

A2: *A mathematical description of the aggregation function induced by instant-runoff voting*

To formally state the aggregation function induced by instant-runoff voting, some additional notation is needed: For $R \in \mathcal{R}$ and $L \subset C$, we denote by $R^{-L} \subset (C \setminus L)^2$ the restriction of R

on L.[22] Denote by $\mathcal{R}^{-L}$ the set of all such relations. Moreover, denote by $\max(R^{-L})$ the set of undominated elements w.r.t. R^{-L}. Then, for a profile $\underline{R}^{-L} := (R_1^{-L}, \ldots, R_n^{-L}) \in (\mathcal{R}^{-L})^n$ we set $L(\underline{R}^{-L}) := \text{argmin}_a|\{i : a \in \max(R_i^{-L})\}|$, that is all alternatives with the fewest number of first place votes in $\underline{R}^{-L}$. For fixed $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$, we recursively define the sets

$$L_{k+1}(\underline{R}) := L\Big(\underline{R}^{-\cup_{s=0}^{k} L_s(\underline{R})}\Big) \tag{22}$$

for $k = 0, \ldots, k_{\max} - 1$, where $L_0 := \emptyset$ and $k_{\max}$ is the smallest integer satisfying $\cup_{s=0}^{k_{\max}} L_s(\underline{R}) = C$. Now we can state the instant runoff aggregation function as

$$\text{IR} : \mathcal{R}^n \to \mathcal{Q} \quad , \quad \underline{R} \mapsto \text{IR}(\underline{R}) \tag{23}$$

where $(a, b) \in \text{IR}(\underline{R})$ iff $\exists k_1 > k_2 \in \{0, \ldots, k_{\max}\}$ such that $a \in L_{k_1}(\underline{R}) \wedge b \in L_{k_2}(\underline{R})$. Using this rule, the group is indifferent between a and b, if there exists $\exists k \in \{0, \ldots, k_{\max}\}$ with $a, b \in L_k(\underline{R})$. Moreover, for $\underline{R} \in \mathcal{R}^n$, $\text{IR}(\underline{R})$ is an asymmetric and negatively transitive relation again.

A3: *A possible simulation design for larger settings*

The simulation is done in the following way: Draw N random samples $\underline{R}_1, \ldots, \underline{R}_N$ from the space of profiles $\mathcal{R}^n$. For all $j = 1, \ldots, \xi$, 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 $\alpha := (\alpha_1, \ldots, \alpha_\xi)$ and aggregation function S, we use the characterization of the assessments given in Proposition 2 and receive

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

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

A simulation design producing less 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 $\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$. For $\underline{R} := (R_1, \ldots, R_n) \in \mathcal{R}^n$ and $\phi \in \Phi$, we set $\underline{R}_\phi := (R_{\phi(1)}, \ldots, R_{\phi(n)})$ and define an equivalence relation $\sim_\Phi$ 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 $\mathcal{R}^n_{\sim_\Phi}$ denote the quotient space produced by $\sim_\Phi$ and let $f : \mathcal{R}^n_{\sim_\Phi} \to \mathcal{R}^n$ be any choice function satisfying $f(\mathcal{C}) \in \mathcal{C}$ for all $\mathcal{C} \in \mathcal{R}^n_{\sim_\Phi}$. Further, for every possible homogeneity value k_j, where $j = 1, \ldots, \xi$, we define the set $\mathcal{L}_j := \{\mathcal{C} \in \mathcal{R}^n_{\sim_\Phi} : \; A_n(f(\mathcal{C})) = k_j\}$ 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 are

[22] Precisely, we have $(a, b) \in R^{-L}$ iff $((a, b) \in R \wedge a, b \in C \setminus L)$. Note that the elements of $\mathcal{R}^n$ are again asymmetric and negatively transitive binary relations, however, defined on the set $C \setminus L$.

constant on every $\mathcal{C} \in \mathcal{R}^n_{\sim_\Phi}$ (as they are invariant under permutations of the inserted profile), one easily verifies the following identities:

$$m^*_{G_n}(S) \;= \sum_{j=1}^{\xi} \Big(\alpha_j \cdot \frac{\sum_{\mathcal{C} \in \mathcal{L}_j} Y_S(f(\mathcal{C})) \cdot |\mathcal{C}|}{\sum_{\mathcal{C} \in \mathcal{L}_j} |\mathcal{C}|} \Big)$$

$$\underline{M}_{G_n}(S) = \sum_{j=1}^{\xi} \Big(\alpha_j \cdot \min_{\mathcal{C} \in \mathcal{L}_j} Y_S(f(\mathcal{C})) \Big)$$

$$\overline{M}_{G_n}(S) = \sum_{j=1}^{\xi} \Big(\alpha_j \cdot \max_{\mathcal{C} \in \mathcal{L}_j} Y_S(f(\mathcal{C})) \Big)$$

Using the above identities allows to apply a similar simulation design as proposed before, however, instead of drawing samples from the space $\mathcal{R}^n$, we can now sample from the smaller space $\mathcal{R}^n_{\sim_\Phi}$. 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 $\mathcal{R}$.