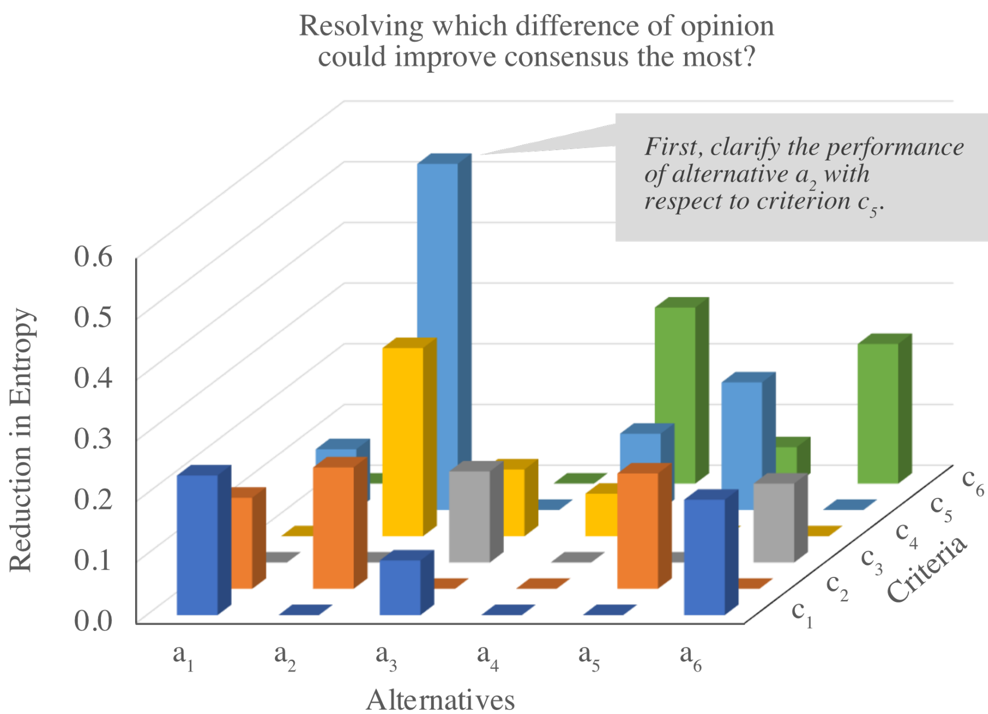
Graphical Abstract

Combinatorial Multi-Criteria Acceptability Analysis: A Decision Analysis and Consensus-Building Approach for Cooperative Groups

Jana Goers, Graham Horton



Highlights

Combinatorial Multi-Criteria Acceptability Analysis: A Decision Analysis and Consensus-Building Approach for Cooperative Groups

Jana Goers, Graham Horton

- Acceptability analysis can be adapted for use by cooperative, rational groups to achieve more detailed analytical insights.
- Judgements and preferences are markers for mental models that group members can share in order to promote consensus.
- Acceptability analysis is extended to include a greedy consensus-building heuristic.
- Consensus-building in a multi-criteria decision problem can be interpreted as an interactive entropy-minimization task.
- Group consensus can be achieved in a small number of steps with a variety of decision models, even if many differences of opinion remain.

Combinatorial Multi-Criteria Acceptability Analysis: A Decision Analysis and Consensus-Building Approach for Cooperative Groups

Jana Goers[a,*], Graham Horton[a]

[a] *Otto-von-Guericke-University Magdeburg, Universitaetsplatz 2, Magdeburg, 39106, , Germany*

Abstract

In group decisions, the decision-makers' judgements and preferences will inevitably conflict. These discrepancies are caused by differing mental models, and a cooperative, rational group can usually arrive at a correct decision by sharing these mental models and agreeing on the most convincing one. Acceptability analysis was introduced as a decision analysis tool in cases where input data is missing or ambiguous and is therefore represented by random variables. It computes a statistical index that describes each alternative's performance. We present a new type of acceptability analysis for group decisions called Combinatorial Multicriteria Acceptability Analysis (CMAA). CMAA is not itself a new decision method, but a framework to be used in conjunction with the decision method of choice. It generates combinations of user inputs, resulting in a discrete computational problem. CMAA identifies the issues where unshared mental models prevent consensus and yields detailed insights into the dependence of acceptability scores on individual user inputs. Information entropy is used to identify the discrepancies whose resolution will improve consensus most quickly, which supports an interactive consensus-building function for a facilitator. An example demonstrates the new analytical and consensus-building possibilities offered by the combinatorial approach.

Keywords: decision analysis, acceptability analysis, information entropy, shared mental models, multi-criteria group decision-making

---

*Corresponding author

Email addresses: `jana.goers@ovgu.de` (Jana Goers), `graham.horton@ovgu.de` (Graham Horton)

1. Introduction

In this paper, we introduce Combinatorial Multicriteria Acceptability Analysis, a decision analysis and consensus-building approach for cooperative, rational groups. In this context, a cooperative group is one whose members have a common objective, and 'rational' means that each member will deliver the same evaluation, if given the same information. Examples of groups that match or come close to this ideal are a team of engineers and designers deciding on the next set of features to be added to a product, a hiring committee for a tenured university position, and physicians on a tumor board selecting a treatment for a cancer patient.

In organisations, there are two principal reasons for making decisions in groups. First, a heterogeneous group of decision-makers has the potential to make a better decision than any of its individual members, if they pool their different areas of expertise effectively, as findings by Stasser and Titus (1985), Kim and Kim (2008) and Nijstad et al. (2014) indicate. Second, a strong consensus can strengthen the decision-makers' commitment to a decision and thus improve a project's chances of success (Moscovici and Doise, 1994; de Vreede et al., 2013).

Stochastic Multi-Criteria Acceptability Analysis (Lahdelma et al., 1998) is a decision-analysis tool for situations in which criteria weights are ambiguous or missing. SMAA performs an inverse weight space analysis that identifies weak alternatives that can potentially be eliminated, and it provides an indication of where more certainty in the input data is needed.

In a group of decision-makers, differences of opinion are inevitable. These discrepancies are a consequence of unshared mental models: different interpretations or items of information that are held only by subsets of the group members. Hidden Profile studies, which originated with the work of Stasser and Titus (1985), have shown that the decision-makers' ability to reach a correct, unanimous decision depends on how effectively they share their mental models. This work utilizes these insights in a multi-criteria context in order to minimize the amount of sharing needed to achieve consensus.

Consensus-building in Multi-criteria Group-Decision-Making (MCGDM) describes iterative processes to improve a consensus metric by successively reducing disparities in user inputs. This is usually achieved by computing a desirable opinion and asking dissenting decision-makers to modify their evaluations towards the desired one (del Moral et al., 2018). Thus, the iterations are driven by a numerical objective, rather than by utilizing minority dissent (Nijstad et al., 2014) or the shared mental model approach. By contrast, the approach presented in this paper is neutral with respect to the resolution of input disparities and does not influence the users in favor of any particular judgement or preference, so that the consensus achieved represents an actual agreement between the decision-makers.

In this paper, we present a new approach to acceptability analysis that we call *Combinatorial Multicriteria Acceptability Analysis* (CMAA). The term 'combinatorial' refers here to the treatment of differing user inputs: instead of fusing them together to obtain a single decision problem, CMAA retains them separately and solves the set of decision problems thus obtained by completely or partially enumerating combinations of the inputs. CMAA is not a decision method *per se*, but a framework that is used in conjunction with the decision method of choice.

The first contribution made by CMAA is concerned with its application as a decision analysis tool. By treating user inputs separately, it is possible to identify the specific issues of dissent with the highest sensitivity for the acceptability indices, enabling detailed insights into the decision problem.

The second contribution is the use of information entropy as a measure of the separation of the alternatives into weak and strong. By computing the entropies that will result from each potential resolution of any discrepancy, we can identify the ones that can deliver the most progress and recommend the corresponding discrepancy to the decision-makers for consideration.

We choose a non-compensatory decision model with a small number of judgement equivalence classes to present and discuss an extended example problem. We also illustrate the application of the combinatorial approach in conjunction with a variety of commonly used models.

2. Scientific Background

2.1. Distributed mental models

A mental model is the representation a person has in their mind of a particular issue. In a heterogeneous group of decision-makers, it is likely that mental models on a particular issue will differ; we call this a *distributed mental model (DMM)*. Decision methods require numerical inputs such as ratios or criteria weights. Each such value represents the overall evaluation of a mental model reduced to a single number, but is not equivalent to it. Distributed mental models can lead to differing inputs by the decision-makers, which we will refer to as *discrepancies*.

Stasser and Titus (1985) used the concept of distributed mental models to study why groups often fail to make better decisions than individuals, even though the group as a whole possesses a wider range of knowledge and information than any of its individual members. They initiated a field of research known as Hidden Profiles. A review of the first 25 years of Hidden Profile research can be found in Lu et al. (2011).

The Hidden Profile paradigm is based on the assumption that the decision group is cooperative and rational. A cooperative group is one whose members have a common objective, rather than different (and often conflicting) objectives. 'Rational' in this context means that all decision-makers will agree on an evaluation of an issue, if they share their mental models of that issue. Note that 'cooperative' does not imply that all members' initial judgements and preferences are similar. In fact, the opposite is preferable: the broader the variety of perspectives available, the greater the likelihood of uncovering the most convincing arguments.

One consequence of the mental model paradigm in MCDM is that decision models with a small number of judgement equivalence classes are preferable. A small number of values is more intuitive and requires less cognitive effort for the decision-makers (Mandler, 2020). Ideally, a decision model will use linguistic variables from a Likert scale with distinct semantics such as 'very poor', 'satisfactory', or 'good', because the semantic gap is smaller than for an abstract, numerical value such as '7 points out of 10'.

We use two examples to illustrate the application of the CMAA method.

Example 1 (Research output). In a hiring committee for a university position, one of the evaluation criteria is 'the candidate has high-quality research output'. Some members of the committee have researched the h-index of a particular candidate and evaluated their research output as 'satisfactory'. Other members have observed that all of this candidate's publications had appeared in top-quality scientific journals, and have submitted the judgement 'very good'. The overall mental model concerning this question is based on two different items of information that are distributed between two disjunct sets of group members and have resulted in a judgement discrepancy.

Example 2 (Departmental priorities). In the same hiring committee, two members know that the president of the university has set the department the goal of improving its research profile. One of them interprets 'high profile' financially and gives criterion 'The candidate has the potential to win large research grants' the highest priority and the criterion 'The candidate has a high-quality research output' second-highest priority. The other committee member also assigns the top two priorities to these criteria, but interprets 'high profile' as the quality of research as measured by scientific peers and assigns these two priorities in the reverse order. Their distributed mental model is based on different interpretations and results in a preference discrepancy.

For a cooperative, rational group, the correct solution to a decision problem is defined as the choice that all members would make independently, once they have shared their mental models of all the relevant issues. Groups that successfully share their mental models are more likely to discover the correct solution to a decision problem (Kim and Kim, 2008; Schulz-Hardt et al., 2006). Different evaluations are a sign that different mental models are present; therefore, dissent within decision groups can improve decision quality (Schulz-Hardt et al., 2006). Nijstad et al. (2014) emphasize the importance of minority dissent for a group's decision-making performance, since a decisive argument may be held by just one member. Schulz-Hardt and Mojzisch (2012) conclude that the ability to share and correctly process minority information is critical to a group's decision-making competence.

In the following, we will refer to the discussion in which a group shares its mental models of a particular issue and decides on a common response as a *clarification conference*. Clarification conferences are proposed by the facilitator in order to resolve discrepancies and may last only a few minutes.

2.2. Use of entropy in MCDM

Information entropy (Shannon, 1948) is a measure of the information content of – or, equivalently, the uncertainty in – a dataset. The data is treated as if it had been generated by a discrete random variable, where the probabilities for each outcome are represented by a probability vector $(p_1, \ldots, p_m)$. The entropy h is then defined as

$$h = -\sum_{i=1}^{m} p_i \cdot \log_2(p_i). \tag{1}$$

The entropy has a maximum value of $h = -\log_2(p_i)$ when $p_i = 1/m \; \forall i$ and a minimum value of $h = 0$ when $p_i = 1$ for some i. The lower the value of h, the greater the separation of the values p_i, which corresponds to a higher information content or a lower uncertainty.

The Entropy Weight Method is used to determine criteria weights without using information from the decision-makers (Zeleny, 1982). It is based on the observation that criteria for which the variability in the performance judgements is low contributes little to the separation of the alternatives (Hwang and Yoon, 1981). The usefulness of each criterion is measured using the entropy of the normalized performance judgements with respect to each criterion, and the low-entropy criteria are assigned a greater weight.

In the context of group decisions, Yue (2017) has applied entropy weighting to determine weights for decision-makers. The entropy of each individual judgement matrix is used to upgrade or downgrade the contributions of each decision-maker to the aggregated result; decision-makers whose judgements have greater distinguishing power are awarded more weight than those whose judgements display less variability.

Ciomek et al. (2017) used information entropy for problems with no criteria information. The decision-maker is repeatedly prompted to declare which of two alternatives they prefer, and the responses are used to reduce the space of permissible criteria weights. The selection of alternative pairs is based on the entropy of their acceptabilities, with the goal of minimizing the number of prompts needed. Van Valkenhoef and Tervonen (2016) developed a similar approach using reference alternatives in order to simplify the comparisons for the decision-maker.

2.3. Consensus-building in MCDM

In Collaboration Engineering, consensus is the level of commitment that a group of stakeholders brings to a proposal (de Vreede et al., 2021). This consensus is important, because it can play a key role in the success of a project (Briggs et al., 2005; Moscovici and Doise, 1994).

On the other hand, in Multicriteria Group Decision-Making, consensus is a numerical quantity that is computed from the user inputs. Hard consensus means unanimous agreement among the decision-makers (Del Moral et al., 2018). Soft consensus means that the computed level of agreement attains a certain numerical threshold (Zhang et al., 2017). It is generally assumed that full unanimity is difficult or impossible to achieve, and soft consensus is therefore usually sought instead.

Until now, group MCDMs have aggregated user inputs to single values or functions, for example the arithmetic mean for additive models or the geometric mean for the Analytic Hierarchy Process. For SMAA, Lahdelma and Salminen (2001) proposed the mean, intersection or union of the continuous user input functions. However, if DMMs are to be represented, no kind of aggregation that fuses the inputs to a single entity can be used, because this would cause their individual identities to be lost.

Discrepancies in the initial preferences and judgements prevent immediate consensus. Consensus-building is typically an iterative process towards a higher degree of consensus by detecting the judgements that are most distant from a desired value and determining a 'correction' trajectory (Pérez et al., 2018). The decision-makers are then encouraged by the algorithm to modify their 'incongruent' judgements along the proposed trajectory (Zhang et al., 2018; Grošelj and Stirn, 2017), or the modification is performed by the algorithm autonomously (Dong and Xu, 2016). In other approaches, Wu et al. (2015) upgrade decision-makers according to their presumed authority, and judgements supplied by decision-makers who are deemed to be 'non-cooperative' can be downgraded by the algorithm (Dong et al., 2016).

Horton and Goers (2021) call consensus-building algorithms of this type 'authority-driven': the algorithm is given the authority to influence or even to override the decision-makers in order to achieve a computational consensus. This contradicts the DMM paradigm, which recognizes the importance of minority dissent. Instead, we claim that an argument-driven approach is needed, in which the decision algorithm is neutral with respect to the outcome and does not manipulate inputs or influence the decision-makers.

2.4. Stochastic multicriteria acceptability analysis

SMAA (Lahdelma et al., 1998) is a tool for analysing group decision problems in which no criteria weight information is available. It has subsequently been adapted to accommodate partial criteria information caused by uncertainty or inaccuracy in criteria measurements, or, in the case of groups, missing or conflicting inputs from the decision-makers (Lahdelma and Salminen, 2001). It has also been combined with various decision methods, including ordinal criteria (Lahdelma et al., 2003), the Analytic Hierarchy Process (Durbach et al., 2014), ELECTRE (Tervonen et al., 2007), PROMETHEE (Corrente et al., 2021) and TOPSIS (Okul et al., 2014). Pelissari et al. (2020) provide a survey of the first 20 years of SMAA research.

SMAA requires continuous, real-valued inputs for both performance measurements and criteria weights. The analysis is based on estimating the fraction of the weight space that results in a particular ranking position for each alternative. The computational model consists of integral expressions in a high-dimensional space. Since these cannot be evaluated analytically, Monte Carlo simulation is used.

The principal SMAA indicator is the rank acceptability index b_i^r. This is an estimate of the proportion of the continuous judgement and preference space that results in rank r for alternative a_i. The rank 1 acceptability of an alternative is often referred to simply as its *acceptability*. In principle, acceptability indices can be used to rank the alternatives or to select a single, preferred alternative. However, this is not the primary purpose of SMAA. Instead, it is used as an analysis tool to provide insight into the relationships between preference vectors and the resulting ranking of the alternatives.

In group decisions, SMAA methods reduce the multiplicity of user inputs to single mathematical objects. However, by aggregating the judgements to a single function, it is no longer possible to compute their individual contributions to the acceptability index of each alternative.

3. Combinatorial Multicriteria Acceptability Analysis

The transition from stochastic to combinatorial acceptability analysis is achieved by a different interpretation of aggregation. In SMAA, criteria preferences and performance judgements are probability distributions, which are fused by the aggregation operator to a single probability distribution function. This function is then evaluated using Monte Carlo simulation. By contrast, user inputs in CMAA are retained side-by-side, resulting in a discrete combinatorial problem. The essential difference is not related to the stochastic nature of the judgements and preferences in SMAA, but to the fact that it fuses different inputs to a single entity.

The CMAA framework is compatible with both compensatory and non-compensatory decision models. Both criteria preferences and the performance of alternatives may be expressed using commonly used data types such as ordinal priorities, scalar weights, intervals or probability distributions. The framework also accommodates the various requirements of these decision models such as the total ordering of ordinal criteria.

A multi-criteria decision problem consists of n criteria, denoted by c_j, m alternatives, denoted by a_i and decision-makers denoted by DM_d.

3.1. Performance judgements

A decision problem is composed of *performance judgement tasks* and *criteria preference tasks*. The evaluation of a judgement task (c_j, a_i) will be referred to as a *judgement* $\lambda_k(c_j, a_i)$, where index k is local to each task. Judgements are subjective, even if objective measurements are available: in Example 1, an h-index of 25 might be judged 'acceptable' by one decision-maker and 'insufficient' by another. We have a *judgement discrepancy* for the judgement task (c_j, a_i), when two or more judgements differ: $\lambda_{k1}(j,i) \neq \lambda_{k2}(j,i)$. Since CMAA makes no assumptions about the decision model, any type of judgement can be used, including scalar or fuzzy numbers, linguistic values or probability distributions.

We can arrange the judgements of each decision-maker into (individual) judgement matrices and aggregate these matrices to create an *aggregated judgement matrix* $\mathbf{A}_{n \times m}$. Each element of the aggregated matrix is the set formed by the union of the corresponding individual judgements:

$$\mathbf{A}_{ji} = \bigcup_k \lambda_k(j,i) \ . \tag{2}$$

	DM$_1$						Aggregated					
	a_1	a_2	a_3	a_4	a_5	a_6	a_1	a_2	a_3	a_4	a_5	a_6
c_1	A	A	B	B	A	B	AB	A	AB	AB	A	B
c_2	B	AB	B	B	B	A	AB	ABX	AB	AB	B	A
c_3	A	A	AB	B	B	A	A	AB	AB	B	B	AB
c_4	B	A	X	X	A	A	B	ABX	ABX	BX	A	A
c_5	X	B	A	-	B	B	X	B	A	A	AB	BX
c_6	A	X	X	A	A	B	A	BX	X	A	AB	B

Table 1: Individual matrix (left) and example aggregated judgement matrix (right)

When the meaning is unambiguous, we will use shorthand expressions such as $\lambda_1\lambda_2$ in place of $\{\lambda_1, \lambda_2\}$. We will refer to the cardinality of the set in Eqn. (2) as the *valency* ϕ_{ji} of the discrepancy.

Table 1 shows an example individual judgement matrix and aggregated judgement matrix. Permissible judgements and their semantics are:

- A: "The alternative meets the criterion very well."
- B: "The alternative meets the criterion satisfactorily."
- X: "The alternative meets the criterion barely, or not at all."

When judgements have discrete values, we may adopt a useful simplification, in that multiple occurrences of a value can be unified. For example, if three decision-makers submit the judgements $\lambda_1 =$ '1', $\lambda_2 =$ '1 or 2' and $\lambda_3 =$ '2 or 3' for a particular judgement task, then the aggregated judgement {'1', '1 or 2', '2 or 3'}, can be abbreviated to {'1', '2', '3'}. In such cases, we assume that the decision-maker who submitted $\lambda_2 =$ '1 or 2' holds distinct arguments in favor of each of the two values, but cannot decide between them, as opposed to having a single argument in favor of the composite '1 or 2'. The aggregated value indicates that mental models are present in the group that result in the judgements '1', '2' and '3'.

Example 1 (continued). Using the ABX model, the judgements regarding the candidate's research are aggregated to {A,B}, or simply AB.

If five decision-makers submit the judgements $\{\lambda_1, \lambda_1, \lambda_1, \lambda_1, \lambda_2\}$ for a particular judgement task, where $\lambda_2 \neq \lambda_1$, the aggregated judgement is $\{\lambda_1, \lambda_2\}$. In other words, judgements do not function as votes. The purpose of the aggregated judgement is not to model the majority opinion or the center of gravity of the judgements. This is an important (and, as far as we know, unique) attribute of the combinatorial approach.

3.2. Criteria preferences

Criteria are treated in an analogous manner to performance judgements. Each preference task requires the decision-makers to provide their subjective evaluations of the importance of a criterion, called a *criteria preference*. Decision-maker preferences for the priority/weight of criterion c_j are denoted by $\mu_k(j)$, where the index k is local to each criterion. Table 2 shows an example of individual and aggregated preferences.

Example 2 (continued). Using the notation 1, 2, 3, ... to denote criteria priorities in decreasing order, one of the members submits the preferences $\mu_1(2) = 1$, $\mu_1(3) = 2$, and the other submits $\mu_2(2) = 2$, $\mu_2(3) = 1$, where criterion c_2 refers to the potential to win large research grants and criterion c_3 refers to the quality of research output.

When more than one preference for a given criterion is submitted, we have a *preference discrepancy*. Note that although the discrepancy is represented as a set of different numerical values, its significance is that different mental models regarding the importance of a particular criterion are present, which may need to be shared between the members of the group.

Individual preferences can be aggregated using the set union operator to form the *aggregated preference vector* $\mathbf{P}$:

$$\mathbf{P}_j = \bigcup_k \mu_k(j) \ .$$

When the μ_k have discrete values, we may choose to simplify the preference vector. If, for example, a decision-maker submits $\mu_1(1) =$ '1 or 2', and another submits $\mu_2(1) =$ '2 or 3', we can represent their aggregated preference as $\{1, 2, 3\}$, instead of {'1 or 2', '2 or 3'}. The reasoning for this is analogous to the case of judgements: a decision-maker has two mental models that lead to the preferences '1' and '2', but cannot decide between them, rather than a single mental model for the composite preference '1 or 2'.

A decision model may place conditions on the preferences or judgements. For example, the criteria weights may be required to form a probability vector. In CMAA, the original user inputs are never modified; any modifications demanded by the model are made when problem instances are generated. Thus, in a decision problem with three criteria, a user preference of 4 for

	DM_1	Aggregated
c_1	3	{1,2,3,4,5}
c_2	1	{1,2,3}
c_3	2	{1,2,3,6}
c_4	5 or 6	{1,2,4,5,6}
c_5	4	{1,2,4,5,6}
c_6	5 or 6	{4,5,6}

Table 2: Individual preference vector for DM_1 and aggregated preference vector

criterion c_1 would be normalized to 0.5 in an instance with preference vector (4, 2, 2) and to 0.4 in an instance with preference vector (4, 3, 3). In a second example, lexicographic methods require the criteria to be totally ordered, which restricts the preference vector to those that represent a bijective mapping between criteria and priorities. In a third example, preserving the expressed order of criteria weights may be desired. The preferences (1, 2, 3) from one decision-maker and (4, 5, 6) from another both express the same desired order of the criteria. This would limit the set of permissible combinations to {(1, 2, 3), (4, 5, 6), (1, 5, 6), (1, 2, 6)}. We call a preference vector that meets the conditions specified by the decision model *feasible*.

3.3. Resolution of discrepancies

Discrepancies are identified by differing numerical values submitted for a particular judgement or preference task. However, in CMAA, their primary function is to flag the presence of differing mental models. *Resolving* a discrepancy means that the decision-makers share their mental models of the issue and agree on one. The external effect of the resolution is that the set of differing numerical values is replaced by a single one.

The resolution of a judgement discrepancy results in the selection of one of its component judgements $\lambda_k(j,i)$ and the deletion of all the others:

$$\mathbf{A}_{ji} \leftarrow \lambda_k(j,i). \tag{3}$$

A judgement matrix that has been modified by resolving the discrepancy for the judgement task (c_j, a_i) to the value $\lambda_k(j,i)$ is written as $\mathbf{A}|\lambda_k(j,i)$.

Instead of selecting one of the inputs as the resolution of a discrepancy, decision-makers may choose a compromise value. They may also fail to resolve a discrepancy, or only achieve a partial resolution.

An aggregated judgement matrix is *completely resolved*, when resolutions have been selected for every discrepancy contained in it. We call the resulting matrix an *instance* of the aggregated matrix, and we denote the number of instances of an aggregated judgement matrix by $||\mathbf{A}||$. $||\mathbf{A}||$ grows exponentially with the number of discrepancies that it contains:

$$||\mathbf{A}|| = \prod_{j,i} \phi_{ji} \ . \tag{4}$$

Similarly, the resolution of a preference discrepancy for criterion c_j results in the selection of one of the preferences values and deleting all the others. A preference vector in which the discrepancy for criterion c_j has been resolved to $\mu_k(j)$ is written $\mathbf{P}|\mu_k(j)$, and the following update is made:

$$\mathbf{P}_j \leftarrow \mu_k(j) \ .$$

A preference vector that no longer contains any discrepancies defines one instance of that vector. We denote the number of instances generated by an aggregated preference vector by double vertical bars: $||\mathbf{P}||$.

We denote the decision problem by $[\mathbf{P}; \mathbf{A}]$. The instance space of the problem is obtained by enumerating all combinations of instances of $\mathbf{P}$ and $\mathbf{A}$. The number of instances in the decision problem K is then given by

$$K = ||\mathbf{P}|| \cdot ||\mathbf{A}||. \tag{5}$$

3.4. Analysis algorithm

The CMAA acceptability analysis algorithm is shown in Algorithm 1. The algorithm begins by creating the aggregated judgement matrix and preference vector from the decision-makers' inputs (line 1). It then creates a new instance of the decision problem (line 2) and solves the decision problem for that instance, resulting in a ranking of the alternatives (line 3). Counter variables that register the rank acceptabilities from the solved decision problem and various attributes of the current instance are updated accordingly (line 4). Lines 2, 3 and 4 are repeated for each instance of the decision problem. Finally, the output analytics are computed (line 6).

If the decision model is stochastic, then each problem instance in line 3 of Algorithm 1 is a stochastic decision problem to which Monte Carlo simulation might be applied. However, this does not mean that CMAA is intrinsically a stochastic method.

The counter variables in line 4 are:

- B_i^r, the number of instances in which alternative a_i achieved rank r,

Algorithm 1: Combinatorial acceptability analysis

Given: Decision model and decision algorithm
Input: Decision-maker judgements λ and preferences μ

```
1 Construct aggregated preference vector P and judgement matrix A;
2 for each instance of [P; A] do
3     Apply the decision algorithm to this instance;
4     Update counter variables;
5 end
6 Compute statistics from the counter variables;
```

Output: Acceptability indices and other statistics

- $Q_i(\lambda_k(j, i2))$, the number of instances in which alternative a_i achieved rank 1 when the judgement task (c_j, a_{i2}) was equal to $\lambda_k(j, i2)$:

$$Q_i(\lambda_k(j, i2)) = B_i^1[\mathbf{P}; \mathbf{A}] \mid \lambda_k(j, i2), \tag{6}$$

- $S_i(\mu_k(j))$, the number of instances in which alternative a_i achieved rank 1 when the preference for c_j was equal to $\mu_k(j)$:

$$S_i(\mu_k(j)) = B_i^1 [\mathbf{P}; \mathbf{A}] \ \mid \mu_k(j),$$

- $Y_i(j1, j2)$, the number of instances in which alternative a_i achieved rank 1 when criterion c_{j1} had a higher priority/greater weight than criterion c_{j2} in the preference vector:

$$\begin{aligned} Y_i(j1, j2) &= B_i^1 [\mathbf{P}; \mathbf{A}] \ \mid c_{j1} \succ c_{j2} \ \text{ or} \\ Y_i(j1, j2) &= B_i^1 [\mathbf{P}; \mathbf{A}] \ \mid c_{j1} > c_{j2}. \end{aligned}$$

  This variable can be used to understand the effects of pairwise criteria relations on alternatives.

- If pairwise comparisons are used as an input format for ordinal criteria, criteria preferences $\mu_k(j1, j2)$ take the form $c_{j1} \succ c_{j2}$. We can then use the notation $Y_i(\mu_k(j1, j2))$ to represent the number of instances in which $\mu_k(j1, j2)$ was valid and alternative a_i achieved rank 1:

$$Y_i(\mu_k(j1, j2)) = B_i^1 [\mathbf{P}; \mathbf{A}] \ \mid \mu_k(j1, j2) \ .$$

The analytical statistics computed in line 6 correspond to those known from SMAA, plus three new ones.

The most important statistic is the rank acceptability index

$$b_i^r = B_i^r / K \ ,$$

the proportion of instances in which alternative a_i achieved rank 1.

If the decision algorithm generates rankings without ties, we will have

$$\sum_{i=1}^{m} b_i^r = 1 \tag{7}$$

for each r. However, if ties result in some instances of the decision, then Eqn. (7) will not hold, since B_i^r will be incremented by each tied alternative at rank r, and the value of B_i^{r+1} will be correspondingly depleted.

The holistic acceptability and pairwise winning indices from SMAA can be computed in an analogous manner from the b_i^r. However, they are not needed when a consensus process follows the initial analysis, and they will not be considered further here. The central weight vector of SMAA has no analog in CMAA, because it is predicated on a continuous weight space.

We now introduce three new analysis variables: the *judgement acceptability index*, the *preference acceptability index* and the *pairwise preference acceptability index*. All three variables give insights into how different options affect the acceptabilities and thus function as sensitivity measures.

The judgement acceptability index $q_i(\lambda_k(j, i2))$ is defined as the proportion of instances in which alternative a_i achieved rank 1 when the judgement discrepancy (c_j, a_{i2}) was set to $\lambda_k(j, i2)$:

$$q_i(\lambda_k(j, i2)) = Q_i(\lambda_k(j, i2))/K \ .$$

A large difference between the values for different judgements of a particular discrepancy indicates that the acceptability of the alternative will vary considerably, depending on which value that discrepancy is resolved to. Conversely, if the values are equal, the acceptability of the alternative is independent of how that discrepancy is resolved. A value of 0 means that the alternative cannot achieve rank 1 if that resolution is chosen.

The judgement acceptability indices for a given alternative can be arranged into a *judgement acceptability matrix*. The values for each discrepancy add up to the rank 1 acceptability of the alternative. Empty entries represent judgement values that were not submitted or unanimous judgements that did not generate a discrepancy.

The preference acceptability index $s_i(\mu_k(j))$ is defined as the proportion of instances in which alternative a_i achieved rank 1 when $\mu_k(j)$ was chosen to resolve the preference discrepancy at criterion c_j:

$$s_i(\mu_k(j)) = S_i(\mu_k(j))/K \ .$$

This is a decision analysis quantity that offers insight how the decision-makers' preferences affect the acceptability of each alternative. An index value $s_i(\mu_k(j)) = 0$ means that alternative a_i can never achieve rank 1 if the preference $\mu_k(j)$ is chosen.

The pairwise preference acceptability $y_i(j1, j2)$ for alternative a_i and criteria c_{j1} and c_{j2} is defined as the proportion of times alternative a_i achieved rank 1 when the criterion c_{j1} had a higher priority/greater weight than criterion c_{j2} in the preference vector:

$$y_i(j1, j2) = Y_i(j1, j2)/K \ .$$

This variable shows how the relative importance of two criteria affects the acceptability of each alternative. An index value $y_i(j1, j2) = 0$ means that alternative a_i never achieved rank 1 when criterion c_{j1} had a higher priority/greater weight than criterion c_{j2}.

If pairwise preference inputs are being used for ordinal criteria, the pairwise preference acceptabilities can be associated with the corresponding preferences $\mu_k(j1, j2)$:

$$y_i(\mu_k(j1, j2)) = Y_i(\mu_k(j1, j2))/K \ .$$

If both $\mu_{k1}(j1, j2) = (c_{j1} \succ c_{j2})$ and $\mu_{k2}(j1, j2) = (c_{j2} \succ c_{j1})$ are present as decision-maker inputs, then we have a pairwise preference discrepancy.

The complexity $\mathcal{C}$ of the analysis algorithm is given by

$$\mathcal{C} = \mathcal{O}(\mathcal{D} \cdot K) \ ,$$

where $\mathcal{D}$ is the complexity of the decision algorithm for one instance of the decision problem (line 3 in Algorithm 1). For many commonly used methods, $\mathcal{D} = \mathcal{O}(m \cdot n)$ operations. Since m and n will usually be small in practice, the complexity essentially depends on the number and cardinality of the discrepancies in the input, according to Eqn. (5).

A large number of judgement discrepancies can lead to prohibitive computation times. To avoid this, Monte Carlo simulation can be used: instead of enumerating the entire set of instances, a set of $K \ll ||\mathbf{A}|| \cdot ||\mathbf{P}||$ randomly sampled instances is used. Line 2 of Algorithm 1 becomes

2 **for** K independent random instances of $[\mathbf{P};\ \mathbf{A}]$ **do**.

Each instance of the decision problem is generated by selecting a resolution of each judgement discrepancy λ and preference discrepancy μ randomly, independently and with equal probability.

3.5. Consensus iteration

The initial acceptability computation will likely reveal many discrepancies. Some of these will have a greater influence on the results than others. An efficient consensus process should address the most influential discrepancies first. We will achieve this by computing the rank 1 acceptabilities that would arise from each potential resolution in $[\mathbf{P};\mathbf{A}]$.

We define the *potential* judgement acceptability index $\hat{q}$ for alternative a_i, judgement task (c_j, a_{i2}) and resolved value $\lambda_k(j, i2)$ as the rank 1 acceptability that a_i would obtain, if that resolution were to be chosen:

$$\hat{q}_i(\lambda_k(j,i2)) = b_i^1[\mathbf{P};\mathbf{A}|\lambda_k(j,i2)].$$

These values are computed as follows:

$$\hat{q}_i(\lambda_k(j,i2)) = Q_i(\lambda_k(j,i2))/K_{ji}^J\ ,$$

where $K_{ji}^J = K/\phi_{ji}$ is the resulting number of instances when the discrepancy at (c_j, a_{i2}) is resolved. In other words, the potential judgement acceptabilities are obtained simply by re-scaling the current values to account for the reduction in $||\mathbf{A}||$ caused by the elimination of the rejected judgements.

Similarly, the potential preference acceptability $\hat{s}$ and potential pairwise preference acceptability $\hat{y}$ are obtained by re-scaling s and y to account for the new, smaller number of instances:

$$\begin{aligned}\hat{s}_i(\mu_k(j)) &= S_i(\mu_k(j))/K_{jk}^P \\ \hat{y}_i(\mu_k(j1,j2)) &= Y_i(\mu_k(j1,j2))/K_{j1,j2}^{PP}\ ,\end{aligned}$$

where K_{jk}^P is the resulting number of instances when the preference discrepancy at criterion c_j is resolved to $\mu_k(j)$, and K_{jk}^{PP} is the resulting number of instances if a pairwise preference discrepancy for criteria c_{j1} and c_{j2} is present and is resolved to $\mu_k(j1,j2)$.

We use the Shannon information entropy to measure the separation of acceptability indices. The entropy resulting from a proposed judgement resolution is obtained by substituting $\hat{q}_i(\lambda_k(j,i_2))$ into Eqn. (1). These values may first have to be normalized to a probability vector, because rank 1 acceptabilities may not sum to 1 if there are instances of the decision problem that return alternatives with shared first ranks.

For a given $\mathbf{P}$ and $\mathbf{A}$, we define the *judgement entropy* $h(\lambda_k(j,i))$ for a given judgement resolution $\lambda_k(j,i)$ as the entropy in the normalized rank 1 acceptabilities when that resolution is applied:

$$h(\lambda_k(j,i)) = h[\mathbf{P};\mathbf{A}|\lambda_k(j,i)]\ .$$

We can arrange the judgement entropies in a *judgement entropy matrix*.

For a given $\mathbf{P}$ and $\mathbf{A}$, we define the *preference entropy* $h(\mu_k(j))$ for preference resolution $\mu_k(j)$ as the entropy in the normalized rank 1 acceptabilities that would result if that resolution were to be chosen:

$$h(\mu_k(j)) = h[\mathbf{P}|\mu_k(j); \mathbf{A}] \ .$$

The pairwise preference entropy is defined analogously:

$$h(\mu_k(j1, j2)) = h[\mathbf{P}|\mu_k(j1, j2); \mathbf{A}] \ .$$

The smallest entropy value identifies the resolution that would provide the greatest separation of the acceptabilities and increase in information. The entropy-minimizing resolution therefore tends to push acceptability values away from the average and towards the extremes 0 and 1.

Based on the judgement and preference entropies, we can now define an iterative, heuristic algorithm to guide the decision-makers from the initial acceptabilities towards a unanimous decision by iteratively reducing the entropy to 0. Alternatively, we might set a threshold $h_{crit} > 0$ for the entropy that allows the decision-makers to terminate the iteration with a soft consensus when the acceptabilities have been sufficiently separated.

The algorithm is shown in Algorithm 2. It receives as input the output from Algorithm 1 and an entropy consensus threshold h_{crit} and returns updated acceptability indices with a lower entropy.

Algorithm 2: CMAA consensus iteration

Given: h_{crit}
Input: Output from Algorithm 1
1 **repeat**
2   select a judgement or preference discrepancy;
3   propose a clarification conference to the DMs;
4   update decision problem according to resolved discrepancy;
5   update judgement and preference entropies;
6   recompute h;
7 **until** $h \leq h_{crit}$ **or** *user-stop*;
Output: Updated acceptability indices

In line 2, the discrepancy is selected that contains the judgement or preference discrepancy with the lowest-entropy resolution, i.e. that yields the smallest value of any $h(\lambda_k(j, i))$ or $h(\mu_k(j))$. The facilitator should not suggest to the decision-makers that they choose the entropy-minimizing resolution – they should always return the resolution that results from their shared mental model of the issue.

The termination condition *user-stop* is included to cover two edge cases in which h_{crit} is not achieved. Certain instances of the decision problem may yield multiple winners, which will prevent all the acceptability accruing to a single alternative, or the decision-makers may fail to resolve or only partially resolve one or more discrepancies in a manner that prevents consensus being reached. Thus, the decision-makers may decide to terminate the iteration and accept the better, but not completely separated, alternatives.

Algorithm 2 implements a greedy, or myopic heuristic: it selects the discrepancy with the greatest potential for entropy improvement in a single step. An algorithm that looks ahead several steps is also conceivable; it would generate a tree of possible paths and select the branch that potentially leads to a hard consensus in the smallest number of steps.

Algorithm 2 is interactive: computations alternate with clarification conferences between the decision-makers. It can be used as a digital assistant that provides support to a human facilitator or as a fully automatic digital facilitator, for example if the decision-makers are distributed geographically and communicate via the internet.

We could apply Algorithm 2 in an authority-driven manner by encouraging the group to agree on the resolution that produces the greatest reduction in entropy. This approach would likely reach a (computational) consensus in a small number of steps, but at the cost of potentially alienating group members whose opinions were overridden in the process. Instead, we follow an argument-driven approach, which is for the digital facilitator to be neutral with regard to the outcome of each discrepancy resolution.

4. Test problem

4.1. Decision problem

We consider a hiring committee tasked with selecting a candidate for a tenured position at a university. There were six applicants $a_1, \ldots, a_6$, and the following six criteria were to be applied:

- c_1: The candidate has previous experience from respected institutions.
- c_2: The candidate has the potential to win large research grants.
- c_3: The candidate's research output is of high quality.

- c_4: The candidate has an attractive teaching portfolio.
- c_5: The candidate made a good impression on the students.
- c_6: The candidate's potential for cooperation with other departments of the university and research centers in the city is high.

The committee consisted of eight decision-makers $\mathrm{DM}_1, \ldots, \mathrm{DM}_8$. A lexicographic decision model was used with the three equivalence classes A, B and X for performance judgements that were defined in Section 3.1.

Table 1 shows the individual judgement matrix for decision-maker DM_1 and the aggregated judgement matrix $\mathbf{A}$ for the test problem. The judgements of the other decision-makers are omitted for the sake of brevity. Decision-maker DM_1 did not provide a judgement for (c_5, a_4), so the aggregated value is determined by the other decision-makers. He/she was uncertain about the judgement tasks (c_2, a_2) and (c_3, a_3) and therefore submitted ambiguous judgements. The aggregated judgement matrix contains three trivalent discrepancies and 14 bivalent discrepancies, which generate a total of $||\mathbf{A}|| = 3^3 \cdot 2^{14} = 442,368$ instances according to Eqn. (4).

Table 2 shows the criteria preferences for decision-maker DM_1 and the aggregated preference vector $\mathbf{P}$ for the test problem. DM_1 had clear preferences for the top four priorities and was indifferent about the ordering of the lowest two. As a group, the decision-makers were quite ambivalent: there was no agreement for any criterion, and for three criteria there were five different proposals. The aggregated preference vector generates $||\mathbf{P}|| = 88$ different feasible preference vectors. The total number of configurations of the decision problem according to Eqn. (5) is therefore $K = 38,928,384$.

4.2. Results of the decision analysis

Algorithm 1 was implemented in the C programming language on a 2.3 GHz MacBook Pro notebook computer. The computation time for a single instance of the test problem was less than $1\mu s$, and the overall computation time was about 28 seconds.

Table 3 shows the rank acceptabilities. The rank 1 acceptabilities deliver the ranking $a_3 \succ a_6 \succ a_1 \succ a_2 \succ a_4 \succ a_5$. Alternative a_5 performs very weakly and would be a candidate for elimination, based on this data alone.

Table 4 shows the judgement acceptability matrix for alternative a_6. Similar matrices exist for the other alternatives. For each discrepancy (c_j, a_i), the three values show the contributions of the corresponding judgements $\lambda_k(j,i) =$ A, B or X to the acceptability $b_6^1 = 0.24$.

	b_i^1	b_i^2	b_i^3	b_i^4	b_i^5	b_i^6
a_1	0.17	0.16	0.16	0.15	0.15	0.20
a_2	0.14	0.15	0.16	0.20	0.15	0.21
a_3	0.25	0.18	0.16	0.14	0.14	0.14
a_4	0.11	0.15	0.15	0.15	0.19	0.24
a_5	0.09	0.18	0.21	0.23	0.20	0.08
a_6	0.24	0.18	0.16	0.13	0.17	0.13

Table 3: Performance results for the test problem

	a_1			a_2			a_3			a_4			a_5			a_6		
	A	B	X	A	B	X	A	B	X	A	B	X	A	B	X	A	B	X
c_1	0.11	0.13	-	-	-	-	0.11	0.12	-	0.12	0.12	-	-	-	-	-	-	-
c_2	0.10	0.14	-	0.06	0.09	0.09	0.10	0.14	-	0.11	0.13	-	-	-	-	-	-	-
c_3	-	-	-	0.11	0.13	-	0.11	0.13	-	-	-	-	-	-	-	0.16	0.08	-
c_4	-	-	-	0.07	0.08	0.08	0.06	0.09	0.09	-	0.12	0.12	-	-	-	-	-	-
c_5	-	-	-	-	-	-	-	-	-	-	-	-	0.12	0.12	-	-	0.12	0.11
c_6	-	-	-	-	0.12	0.12	-	-	-	-	-	-	0.12	0.12	-	-	-	-

Table 4: Judgement acceptability matrix for alternative a_6

Judgement acceptabilities provide sensitivity information, since a large difference in values means that the acceptability is strongly dependent on which resolution is chosen. In the example, the greatest sensitivity for alternative a_6 is located at (c_3, a_6), where the judgement A delivers twice as much acceptability as the judgement B. By contrast, equal values (for example for the discrepancy at (c_6, a_5)) mean that the acceptability of a_6 does not depend on how this discrepancy is resolved.

Table 5 shows the preference acceptabilities for alternative a_6. Each row and column of the table adds up to the rank 1 acceptability $b_6^1 = 0.24$. This alternative is not particularly sensitive to the priority of c_6, but it is highly sensitive to the priority of c_1: it is preferred four times as often when criterion c_1 is located at priority 5, compared to priority 2, and it is inefficient if c_1 is located at priority 1. Similarly, it is preferred more than four times as often when criterion c_5 is located at priority 6 compared to priority 2, and it is inefficient if c_5 is located at priority 1. Analogous datasets exist for the other alternatives, which provide similar insights.

Table 6 shows the pairwise preference acceptabilities for alternative a_6. Of interest are the mirror pairs (c_{j1}, c_{j2}) and (c_{j2}, c_{j1}) which sum to $b_6^1 =$

	1	2	3	4	5	6
c_1	0.00	0.02	0.07	0.07	0.08	-
c_2	0.10	0.07	0.07	-	-	-
c_3	0.05	0.05	0.09	-	-	0.04
c_4	0.09	0.08	-	0.03	0.02	0.02
c_5	0.00	0.02	-	0.06	0.07	0.09
c_6	-	-	-	0.07	0.07	0.09

Table 5: Preference acceptabilities for alternative a_6

Superior	Inferior criterion					
criterion	c_1	c_2	c_3	c_4	c_5	c_6
c_1	-	0.01	0.05	0.06	0.19	0.20
c_2	0.23	-	0.16	0.12	0.23	0.24
c_3	0.18	0.08	-	0.08	0.19	0.20
c_4	0.17	0.11	0.15	-	0.20	0.21
c_5	0.05	0.01	0.05	0.04	-	0.13
c_6	0.04	-	0.04	0.03	0.11	-

Table 6: Pairwise preference acceptabilities for alternative a_6

0.24. Alternative a_6 depends strongly on the relative priorities of c_1 and c_2: almost all the acceptability stems from instances in which $c_2 \succ c_1$. Similar data exists for the other alternatives, enabling analogous conclusions.

Example 2 (continued). The acceptability of candidate a_6 is twice as high, when grant potential (criterion c_2) has a higher priority than research quality (criterion c_3) compared to the reversed case.

The level of detail offered by these three new analysis variables is one of the principal advantages of the combinatorial approach; the decision-makers are given a great amount of detailed information about the contributions of their various judgements and preferences to the acceptability indices.

4.3. Results of the consensus-building iteration

The initial entropy in the normalized rank 1 acceptabilities b_i^1 according to Eqn. (1) is $h = 2.49$, which is close to the maximum possible value of $h = 2.59$, indicating a high level of uncertainty.

	a_1			a_2			a_3			a_4			a_5			a_6		
	A	B	X	A	B	X	A	B	X	A	B	X	A	B	X	A	B	X
c_1	2.44	2.45	-	-	-	-	2.39	2.53	-	2.50	2.45	-	-	-	-	-	-	-
c_2	2.47	2.44	-	2.42	2.42	2.38	2.32	2.51	-	2.51	2.37	-	-	-	-	-	-	-
c_3	-	-	-	2.48	2.40	-	2.36	2.51	-	-	-	-	-	-	-	2.43	2.50	-
c_4	-	-	-	2.50	2.48	2.47	2.40	2.50	2.52	-	2.50	2.49	-	-	-	-	-	-
c_5	-	-	-	-	-	-	-	-	-	-	-	-	2.51	2.45	-	-	2.48	2.50
c_6	-	-	-	-	2.49	2.49	-	-	-	-	-	-	2.49	2.49	-	-	-	-

Table 7: Judgement entropies for the test problem

	Position in preference vector					
	1	2	3	4	5	6
c_1	2.28	2.48	2.27	2.19	2.17	-
c_2	2.21	2.42	2.50	-	-	-
c_3	1.95	2.36	2.36	-	-	2.19
c_4	1.60	2.11	-	2.44	2.41	2.30
c_5	1.42	2.32	-	2.41	2.37	2.13
c_6	-	-	-	2.54	2.52	2.30

Table 8: Preference entropies for the test problem

Table 7 shows the judgement entropies after the initial analysis. All values are close to the starting value of $h = 2.49$, indicating that no significant improvement in the separation of the acceptabilities can be achieved by resolving any of these discrepancies. The smallest entropy value is 2.32, which would be obtained by resolving the AB discrepancy at (c_2, a_3) to A.

Table 8 shows the entropies that result from each possible preference resolution after the initial acceptability computation. The smallest value is $h = 1.42$, which would be obtained if criterion c_5 were to be assigned priority 1. This would represent a substantial improvement in the separation of acceptabilities over the initial $h = 2.49$.

Table 9 shows an example consensus iteration for the test problem. Step 0 shows the results from the initial decision analysis, as previously described. The greatest reduction in entropy would be obtained if the preference discrepancy $\mathbf{P}_5 = \{1, 2, 4, 5, 6\}$ were to be resolved to priority 1, so the algorithm proposes a clarification conference for this discrepancy.

In Step 1, the group agrees on priority 6 for criterion c_5, which only provides a modest reduction in entropy to $h = 2.13$. Alternatives a_1, a_2 and

Step	Chosen resol.	b_1^1	b_2^1	b_3^1	b_4^1	b_5^1	b_6^1	h	Min. entropy resolution
		Rank 1 acceptabilities							
0	-	0.17	0.14	0.25	0.11	0.09	0.24	2.49	$(c_5, 1)$
1	$(c_5, 6)$	0.27	0.17	0.12	0.04	0.02	0.38	2.13	$(c_4, 1)$
2	$(c_4, 1)$	0.00	0.09	0.08	0.00	0.00	0.83	0.82	$(c_3, a_6) =$A
3	$(c_3, a_6) =$AB	0.00	0.09	0.08	0.00	0.00	0.83	0.82	$(c_4, a_2) =$B or X
4	$(c_4, a_2) =$BX	0.00	0.00	0.08	0.00	0.00	0.92	0.41	$(c_4, a_3) =$B or X
5	$(c_4, a_3) =$B	0.00	0.00	0.00	0.00	0.00	1.00	0.00	-

Table 9: Example consensus iteration for the test problem

a_6 are strengthened, while the others are weakened. The smallest entropy is now to be found if criterion c_4 were to be assigned to priority 1, so the facilitator proposes a clarification conference for the discrepancy $\mathbf{P}_4 = \{1, 2, 4, 5\}$.

In Step 2, the group assigns priority 1 to c_4, achieving the maximum entropy improvement. This eliminates a_1, a_4 and a_5, and the acceptability of a_6 increases to 0.83. The greatest entropy reduction can now be achieved by resolving the AB judgement discrepancy at (c_3, a_6) to A.

Example 1 (continued). In Step 3, the group is unable to resolve the discrepancy at (c_3, a_6) concerning the quality of candidate a_6's research output: the aggregated judgement therefore remains AB, and the acceptabilities remain unchanged.

Since the discrepancy was not resolved, the discrepancy containing the second-largest reduction in entropy is considered. This is the ABX discrepancy at (c_4, a_2), if the resolution B or X is chosen.

In Step 4, the group can only reach a partial resolution of this discrepancy: they agree on the evaluation BX. Nevertheless, this results in the elimination of a_2 and the further strengthening of a_6 to 0.92. The algorithm now determines that resolving the ABX discrepancy at (c_4, a_3) to either B or X will result in a hard consensus with a_6 as the winning alternative, and the group resolves to B in the final clarification conference.

Consensus was achieved in five steps, in which the group chose the minimal-entropy resolution twice, the non-minimizing resolution once, failed to resolve a discrepancy once and managed only a partial resolution once. The method is quite resilient, achieving consensus quickly, even though the decision-makers chose a sub-optimal resolution in three out of five cases. At termination, $||\mathbf{P}|| = 4$ and $||\mathbf{A}|| = 98,304$, so the number of instances remaining in the decision problem was 393,216.

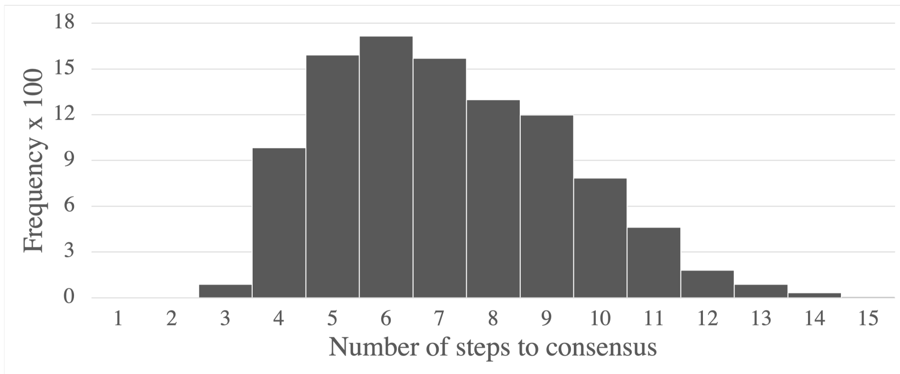


Figure 1: Distribution of consensus path lengths for the test problem

Example 1 (conclusion). Consensus has been reached, and candidate a_6 has been selected, even though the committee did not resolve the question of the quality of his/her research output.

Example 2 (conclusion). Consensus has been reached, and a candidate has been selected, even though the relative priorities of criteria c_2 (grant potential) and c_3 (research quality) remained unresolved.

5. Analysis and application of the algorithms

5.1. Distribution of consensus path length

The example consensus iteration for the test problem in the previous section was designed to illustrate the various outcomes of a clarification conference. However, it is unclear how representative this particular example is. We therefore performed a simulation study of consensus path lengths. If we consider the decision-makers' resolution of each discrepancy proposed by the consensus algorithm to be a random event, then we can treat the number of steps needed to achieve consensus as a random variable.

In a Monte Carlo simulation experiment, the test problem was solved 10,000 times, randomly resolving to either a $\lambda_k(j,i)$ or a $\mu_k(j)$ at each step, where the probabilities for choosing each resolution were uniformly distributed. The results of this experiment are shown in Figure 1. The shortest paths found consisted of only three steps. The longest paths required 15 steps to reach consensus. The mean path length was 7.14 steps. We conclude that the example path in Section 4.3 is shorter than average, but with a probability of 0.16 is not unreasonable.

If, in the consensus iteration of Section 4, the group had chosen the entropy-minimizing resolution at each step, then five steps would have been needed to reach consensus. The greedy heuristic therefore performed better than the average random path, but did not find the shortest path. Note also, that in this example, the failure to select the entropy-minimizing resolution at several steps did not increase the number of steps needed for consensus beyond the entropy-minimizing number.

5.2. Application of CMAA with other decision models

Consensus-building using the combinatorial approach benefits from a non-compensatory decision model, because it enables alternatives to be eliminated quickly. However, the approach is not limited to this choice, and in this section we apply in conjunction with a variety of well-known decision models using larger and more complex decision problems.

A random number generator was used to simulate the input of three decision-makers for a decision problem of size $m = 10$, $n = 6$. Permissible criteria weights and performance scores were integers in the range $[1, 5]$, except in the case of the ABX model. Parameters were chosen so that $||\mathbf{P}|| \approx 200$ and about 50% of the judgement tasks in $\mathbf{A}$ were discrepancies. The resulting random judgement matrices $\mathbf{A}$ and preference vectors $\mathbf{P}$ were used to generate decision problems using six different decision models:

- SAW: Simple additive weighting.
- WPM: Weighted Product Method. This approach is similar to Simple Additive Weighting, but the addition and multiplication operations are replaced by multiplication and exponentiation, respectively.
- PROMETHEE: PROMETHEE is an outranking method based on pairwise comparisons of performance values (Behzadian et al., 2010).
- TOPSIS: TOPSIS (Hwang and Yoon, 1981) evaluates alternatives based on their distance to hypothetical best and worst alternatives.
- SAW-NC: The preferences are taken to be on an ordinal scale, where $n \succ n - 1$. This results in a non-compensatory decision model.
- ABX-LEX: The decision model used for the test problem of Section 4.1. Here, only the three judgements A, B and X are permitted. However, the density of discrepancies was the same as for the other models.

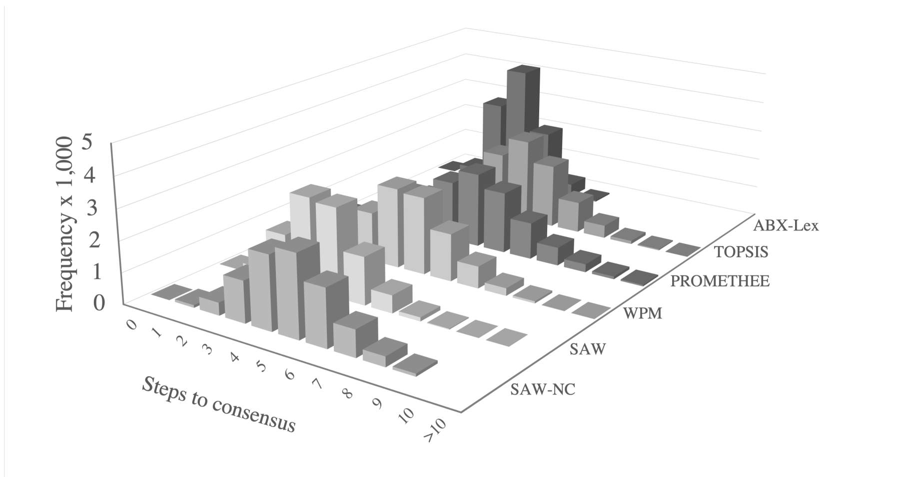


Figure 2: Entropy-minimizing consensus path lengths for six decision models

For each decision model, 10,000 random decision problems with the specified parameters were generated. For the four compensatory decision models and SAW-NC, synchronized random numbers were used to ensure that each of the 10,000 sample problems was identical, and criteria weights were normalized. Each of the decision problems was solved using a Monte Carlo simulation with 10,000 independent replications, and the consensus path selected the entropy-minimizing resolution at each step.

Figure 2 shows the distributions of the entropy-minimizing consensus path lengths for each of the six decision models. The lowest mean number of steps to achieve consensus was found for the non-compensatory ABX-Lex model. For three of the compensatory methods, a very small number of paths of length 10 or more (fewer than 0.5% of the total) were found. We believe these are rare event anomalies due to the Monte Carlo simulation. These results suggest that the entropy-based combinatorial acceptability method converges quickly for all six models.

6. Summary and outlook

6.1. Summary

In this paper, we have introduced Combinatorial Multicriteria Acceptability Analysis as a decision analysis tool for cooperative groups. In the CMAA paradigm, identifying and sharing distributed mental models is the key to achieving consensus on the correct decision. CMAA explicitly represents each decision-maker's input in its data structures and uses the set-theoretic union as the aggregation operator for user judgements and preferences. This leads to a combinatorial computational task which is solved either by enumerating the entire space of combinations, or by sampling a random subset in the Monte Carlo manner.

This approach allows us to compute the sensitivity of each alternative's acceptability to each judgement and preference. The decision-makers can learn which individual inputs strengthen or weaken each alternative, and which do not affect them at all. In this, CMAA provides a greater level of insight into the decision problem than was previously possible.

A further contribution of this paper is the use of information entropy as a metric for the separation of the acceptabilities. By computing the entropy change that can be achieved by each discrepancy resolution, the issue with the greatest potential for improving consensus can be identified, yielding a simple, greedy consensus-building heuristic.

This heuristic makes an interactive software agent possible that guides the decision-makers towards consensus by successively suggesting discrepancies for them to resolve. It can be implemented either as a digital assistant for a human facilitator or as a fully automated digital facilitator.

We recommend the CMAA approach to practitioners, if they seek fast and straightforward consensus for a multi-criteria decision in a cooperative group. CMAA provides improved insights into the decision and can guide the group to a unanimous, hard consensus with a small number of discrepancy resolution discussions. Consensus-building with CMAA is efficient for several well-known decision models. However, for maximum efficiency, a non-compensatory decision model with a small number of judgement equivalence classes should be used, which may be unfamiliar to the decision-makers.

6.2. Outlook

The computational complexity of CMAA grows exponentially with the number of discrepancies generated by the decision problem. For very large problems, the complete enumeration of all instances can be replaced by Monte Carlo simulation. The behaviour and performance of the Monte Carlo variant of the method have yet to be studied.

The consensus-seeking heuristic employed here is greedy: it selects the discrepancy for resolution that has the greatest potential for entropy reduction, but does not necessarily generate the shortest overall path. It seems attractive to develop and study a variant with a greater degree of look-ahead.

CMAA demonstrated some resilience for the test problem: it can improve consensus, even if the decision-makers fail to resolve discrepancies. For this reason, it might also be applicable when the decision-makers are not cooperative, such as in public and political decision-making.

The combinatorial approach to acceptability analysis is not limited to the ordinal decision model used for illustration in this paper. However, the details of the application to the various decision models that are in widespread use and the behavior of the resulting methods – in particular the convergence speed of the consensus iteration and the resilience with respect to non-optimal and failed resolutions – are topics for future research.

7. Acknowledgement

The authors thank the anonymous reviewers for their comments, which have helped them to significantly improve several aspects of this paper.

References

Behzadian, M., Kazemzadeh, R., Albadvi, A., and Aghdasi, M. (2010). Promethee: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, 200(1):198–215.

Briggs, R. O., Kolfschoten, G. L., and Vreede, G.-J. d. (2005). Toward a theoretical model of consensus building. In *Eleventh Americas Conference on Information Systems*, AMCIS, pages 101 – 110.

Ciomek, K., Kadziński, M., and Tervonen, T. (2017). Heuristics for selecting pair-wise elicitation questions in multiple criteria choice problems. *European Journal of Operational Research*, 262(2):693–707.

Corrente, S., Greco, S., Leonardi, F., and Slowinski, R. (2021). The hierarchical SMAA-PROMETHEE method applied to assess the sustainability of European cities. *Applied Intelligence*, 51:1–19.

de Vreede, G.-J., Briggs, R. O., and Kolfschoten, G. L. (2021). Collaboration engineering for group decision and negotiation. In Kilgour, D. M. and Eden, C., editors, *Handbook of Group Decision and Negotiation*, pages 751–776. Springer Nature Switzerland, Cham, Switzerland.

de Vreede, T., Reiter-Palmon, R., and de Vreede, G.-J. (2013). The effect of shared mental models on consensus. In *2013 46th Hawaii International Conference on System Sciences*, pages 263–272. IEEE.

del Moral, M. J., Chiclana, F., Tapia, J. M., and Herrera-Viedma, E. (2018). A comparative study on consensus measures in group decision making. *International Journal of Intelligent Systems*, 33(8):1624–1638.

Del Moral, M. J., Tapia, J. M., Chiclana, F., Al-Hmouz, A., and Herrera-Viedma, E. (2018). An analysis of consensus approaches based on different concepts of coincidence. *Journal of Intelligent & Fuzzy Systems*, 34(4):2247–2259.

Dong, Y. and Xu, J. (2016). *Consensus Building in Group Decision Making*. Springer, Singapore.

Dong, Y., Zhang, H., and Herrera-Viedma, E. (2016). Integrating experts' weights generated dynamically into the consensus reaching process and its applications in managing non-cooperative behaviors. *Decision Support Systems*, 84:1–15.

Durbach, I., Lahdelma, R., and Salminen, P. (2014). The analytic hierarchy process with stochastic judgements. *European Journal of Operational Research*, 238(2):552–559.

Grošelj, P. and Stirn, L. Z. (2017). Soft consensus model for the group fuzzy AHP decision making. *Croatian Operational Research Review*, 8(1):207–220.

Horton, G. and Goers, J. (2021). ABX-LEX: An argument-driven approach for the digital facilitation of efficient group decisions. *International Journal of Information Technology & Decision Making*, 20(01):137–164.

Hwang, C. and Yoon, K. P. (1981). *Multiple Attribute Decision Making, Methods and Applications: A State-of-the-art Survey*. Springer.

Kim, H. and Kim, D. (2008). The effects of the coordination support on shared mental models and coordinated action. *British Journal of Educational Technology*, 39(3):522–537.

Lahdelma, R., Hokkanen, J., and Salminen, P. (1998). SMAA-stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, 106(1):137–143.

Lahdelma, R., Miettinen, K., and Salminen, P. (2003). Ordinal criteria in stochastic multicriteria acceptability analysis (smaa). *European Journal of Operational Research*, 147(1):117–127.

Lahdelma, R. and Salminen, P. (2001). SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. *Operations Research*, 49(3):444–454.

Lu, L., Yuan, Y. C., and McLeod, P. L. (2011). Twenty-five years of hidden profiles in group decision making: A meta-analysis. *Personality and Social Psychology Review*, 16(1):54–75.

Mandler, M. (2020). Coarse, efficient decision-making. *Journal of the European Economic Association*, 18(6):3006–3044.

Moscovici, S. and Doise, W. (1994). *Conflict and consensus: A general theory of collective decisions*. SAGE Publications Ltd.

Nijstad, B. A., Berger-Selman, F., and De Dreu, C. K. (2014). Innovation in top management teams: Minority dissent, transformational leadership, and radical innovations. *European Journal of Work and Organizational Psychology*, 23(2):310–322.

Okul, D., Gencer, C., and Aydogan, E. K. (2014). A method based on smaa-topsis for stochastic multi-criteria decision making and a real-world application. *International Journal of Information Technology & Decision Making*, 13(05):957–978.

Pelissari, R., Oliveira, M. C., Amor, S. B., Kandakoglu, A., and Helleno, A. L. (2020). Smaa methods and their applications: a literature review and future research directions. *Annals of Operations Research*, 293(2):433–493.

Pérez, I. J., Cabrerizo, F. J., Alonso, S., Dong, Y., Chiclana, F., and Herrera-Viedma, E. (2018). On dynamic consensus processes in group decision making problems. *Information Sciences*, 459:20–35.

Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R., and Frey, D. (2006). Group decision making in hidden profile situations: dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology*, 91(6):1080.

Schulz-Hardt, S. and Mojzisch, A. (2012). How to achieve synergy in group decision making: Lessons to be learned from the hidden profile paradigm. *European Review of Social Psychology*, 23(1):305–343.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(4):623–666.

Stasser, G. and Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(6):1467–1478.

Tervonen, T., Lahdelma, R., Dias, J. A., Figueira, J., and Salminen, P. (2007). SMAA-TRI. In Linkov, I., Kiker, G. A., and Wenning, R. J., editors, *Environmental security in harbors and coastal areas*, pages 217–231. Springer.

Van Valkenhoef, G. V. and Tervonen, T. (2016). Entropy-optimal weight constraint elicitation with additive multi-attribute utility models. *Omega*, 64:1–12.

Wu, J., Chiclana, F., and Herrera-Viedma, E. (2015). Trust based consensus model for social network in an incomplete linguistic information context. *Applied Soft Computing*, 35:827–839.

Yue, C. (2017). Entropy-based weights on decision makers in group decision-making setting with hybrid preference representations. *Applied Soft Computing*, 60:737–749.

Zeleny, M. (1982). *Multiple Criteria Decision Making*. McGraw-Hill.

Zhang, H., Dong, Y., and Chen, X. (2017). The 2-rank consensus reaching model in the multigranular linguistic multiple-attribute group decision-making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(12):2080–2094.

Zhang, H., Palomares, I., Dong, Y., and Wang, W. (2018). Managing non-cooperative behaviors in consensus-based multiple attribute group decision making: An approach based on social network analysis. *Knowledge-Based Systems*, 162:29–45.