Decision Making: Applications in Management and Engineering
Vol. 6, Issue 2, 2023, pp. 828-852.
ISSN: 2560-6018
eISSN: 2620-0104
DOI: https://doi.org/10.31181/dmame622023783

PROJECT SELECTION IN A BIOTECHNOLOGY STARTUP USING COMBINATORIAL ACCEPTABILITY ANALYSIS

Jana Goers[1*] and Graham Horton[1]

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

Received: 3 May 2022;
Accepted: 17 July 2023;
Available online: 23 August 2023.

Original scientific paper

Abstract: *Combinatorial Multi-Criteria Acceptability Analysis (CMAA) is a new algorithmic framework that enables the use of standard (i.e., single-user) multicriteria decision-making methods by groups. In this paper, we present the first application of CMAA to a real-life decision. Our objectives were to study the performance of the method in a real-life setting and to test two hypotheses concerning the application of the method. Three founders of a biotechnology startup had to choose a product development project. We describe the decision problem and the consensus path taken by the founders, and we illustrate some of the analytical possibilities offered by the method. Of the 25 evaluation conflicts contained in the initial input, only eight needed to be resolved in order to achieve a hard consensus. A simulation experiment showed that the expected value for this size problem is 7.6 resolution steps. The method generates a very large state space, so complete enumeration can become prohibitively expensive. A computational experiment confirmed our assumption that 10,000 random samples are sufficient if Monte Carlo simulation is used instead. A third simulation experiment provided support for the hypothesis that consensus-building with the non-compensatory decision model used is more efficient than with a more typical compensatory model. We conclude that the CMAA method is well-suited for multi-criteria group decisions; it provides a wealth of analytical detail, and its entropy-based heuristic can guide the group to consensus in a small number of steps.*

Keywords: *Decision analysis, acceptability analysis, group decision-making, shared mental models, consensus-building.*

1. Introduction

In this paper we present the first application of the Combinatorial Multicriteria Acceptability Analysis (CMAA) framework (Goers & Horton, 2023) to a group decision.

* Corresponding author.
E-mail addresses: jana.goers@ovgu.de (J. Goers), graham.horton@ovgu.de (G. Horton)

Our objective is to validate the appropriateness of CMAA for practical decision-making and to develop some guidelines for practitioners. In the case study, the founders of a biotechnology startup had to select a product development project.

Typical multi-criteria group decision-making (MCGDM) methods are based on distance measures – usually between the user inputs and the group average (Moral et al., 2018; Tapia et al., 2023). Hard consensus means this distance is zero, but this is generally assumed to be unrealistic, so a soft consensus with a non-zero distance is accepted instead (Guo et al., 2023). Consensus-building consists of identifying the inputs that are to be adjusted using an identification rule, determining a desired correction direction for those inputs using a direction rule, and requesting the affected decision-makers to modify their inputs in the required direction (Zhang et al., 2019). Such approaches assume that decision-makers have different – and possibly conflicting – objectives, and that modifying their evaluations will incur a cost (Guo et al., 2023; Zhang et al., 2019). However, input averaging approaches have the drawbacks that they result in an evaluation that corresponds to nobody's opinion and that they suppress outlying, but potentially important judgements.

Combinatorial Multicriteria Acceptability Analysis was developed to support cooperative, rational groups in achieving quick decisions with a high degree of consensus. CMAA is a framework that can be integrated with any standard multi-criteria decision method such as SAW, AHP, TOPSIS or PROMETHEE. Instead of aggregating decision-maker inputs to an average value, CMAA generates combinations of them, and the proportion of combinations that return an alternative as the preferred alternative is its acceptability. The information entropy in the vector of acceptabilities is used as consensus measure, and a consensus-building is an entropy minimization task. Since the entropy of the acceptabilities makes no reference to decision-maker inputs, it is an output metric, but it is not a `coincidence among solutions´ (Herrera-Viedma et al., 2014) since it does not compare the scores or ranks produced by individual decision-makers. The entropy-based consensus-building process has been shown to converge considerably faster than a typical input averaging approach (Goers & Horton, 2023).

A startup is a company that has been recently founded or is in the process of being founded and meets two conditions: it is introducing an innovation to the market, and it has the intent to grow rapidly (Tech, 2018). Startup founders therefore have to make decisions that are based on assumptions and are subject to a large amount of uncertainty. Indeed, one influential definition of a startup is "a human institution, founded to create a new product or service in the conditions of extreme uncertainty" (Ries, 2011). Technological uncertainty is concerned with whether the innovation can be made to meet performance, manufacturability, usability and other technical criteria, and market uncertainty is centered on whether the market will accept the new and unfamiliar solution. Bortolini et al. (2021) state: "In environments and situations of great uncertainty, complexity and speed of change, especially for new business ventures or startups, many scholars believe that success comes from the speed at which the organization can conduct tests and experiments". This is particularly true for early-stage startups, which have very limited time and resources to achieve product-market fit.

In a group decision involving a high degree of uncertainty, the probability is high that the decision-makers will submit differing judgements or preferences on any given question, because they hold varying information about it or interpret it differently. This *unshared mental model* can explain why a group fails to reach a correct decision, even though as a whole it has greater knowledge than any of its individual members (Stasser & Titus, 1985). To reveal the correct decision, decision-makers must share their individual mental models to form a *shared mental model* (De Vreede et al., 2013; Schulz-Hardt & Mojzisch, 2012).

We assume that the decision-making group is cooperative and rational. In this context, 'cooperative' means that its members are all pursuing the same objective and that hidden agendas and politics do not play a role in the decision, and 'rational' means that all members will submit the same judgement if they all have access to the same information. They are an example of a 'consensus-committed group' (DeSanctis & Gallupe, 1987). In a cooperative, rational group, decision-makers modify their initial evaluations willingly and quickly when they hear new information or stronger arguments from the other members of the group. In our experience, both startup founders and innovation teams in established companies are cooperative and rational decision-makers.

The contributions of this work are as follows:

- to describe the first application of CMAA to a real-life group decision, and illustrate some of the analytical details it can provide,
- to define active, inactive and pivot discrepancies, which aid understanding of the consensus-building process,
- to validate the acceptance of the method by independent decision-makers,
- to demonstrate that CMAA is more efficient when a non-compensatory decision model is used compared to a compensatory model, and
- to show that 10,000 random samples are sufficient when the analysis is performed by Monte Carlo simulation.

2. Background

2.1. Distributed and shared mental models

Since a group possesses a wider range of knowledge and information than any of its members, it should, in principle, be able to make better decisions than any of them individually. However, this is often not the case in practice, and (Stasser & Titus, 1985) initiated a field of research known as Hidden Profiles (Lu et al., 2012) in order to study this phenomenon.

Hidden Profile research is based on the concept of a *mental model* – the set of information held by a person or group of people on a particular issue. In a heterogeneous group of decision-makers, members will have different knowledge and experience, and the mental model of the issue is *unshared* or *distributed*. In a group decision, a distributed mental model (DMM) will lead to differing performance judgements and criteria preferences, which we will refer to as *judgement discrepancies* and *preference discrepancies*, respectively.

In this context, there exists a correct decision, which is defined as the alternative that all members of the group would choose, once their mental models have been shared. Hidden Profile studies (Kline, 2005; Schulz-Hardt et al., 2006) have shown that a group can discover and agree on the correct decision, if they are successful at sharing their mental models. Clearly, initial dissent within the group is necessary for decision quality (Dreu & West, 2001; Nijstad et al., 2014), and the ability to share and correctly process minority information is critical to the decision-making competence of a group (Schulz-Hardt & Mojzisch, 2012).

The task of a consensus-building algorithm should therefore be to identify those discrepancies whose resolution would improve consensus the most and present these to the group for discussion. We call these discussions *clarification conferences*, in which the distributed mental model becomes a *shared mental model,* and the decision-makers agree on a judgement or preference, thereby resolving the discrepancy. (Yahaya & Abu-Bakar (2007) describe the clarification conference as follows: "[The group member articulates] his basic assumption, his frame of reference and his mental model that leads him to believe in something about the uncertainties. When [these] are exchanged and challenged, they are tested until the best one prevails." Goers & Horton (2023) have shown that, using the CMAA approach, clarification conferences may only be needed for a fraction of the total number of discrepancies generated by the initial decision-maker input.

2.2. Treatment of groups in MCDM

Multi-criteria decision methods are designed for use by a single decision-maker. Therefore, when used in group decision-making, the differing judgements and preferences must be reduced to single values. This process is known as aggregation and is generally realized by averaging the inputs using the arithmetic or geometric mean, or, more generally, an ordered weighted average operator (Moral et al., 2018). Linear decision models such as Simple Additive Weighting use the arithmetic mean, while multiplicative models such as the Analytic Hierarchy Process and the Weighted Product Method use the geometric mean.

One disadvantage of this approach is that the averaged values may misrepresent the situation and lead to an incorrect decision. For example, three decision-makers might judge the performance of an alternative with respect to a given criterion with scores of 1 (very poor), 3 (mediocre) and 5 (very good). A decision method that uses the arithmetic mean would treat the performance as mediocre. However, the decision-maker who submitted the score of 1 (or of 5) may possess a critical piece of information that is unknown to the other two decision-makers, and which would convince them to also submit a score of 1 (or 5), if they were to learn of it.

There are many case studies in the literature where judgements vary widely across the decision-makers. In one example(Bairagi, 2022; Zhang et al., 2019), four decision-makers submitted the performance judgements \`Fair´, \`Medium Good´, \`Good´, and \`Very Good´ for the alternative/criterion pairs (c_5, a_3) and (c_{15}, a_2). These were then converted to fuzzy intervals, averaged, and finally converted to crisp numbers to be used in the SAW decision method. The criteria were based on objective factors, and the decision-makers were stated to be rational. Under these circumstances, it seems likely that by sharing their mental models of these questions, the decision-makers could have produced a more accurate and more unanimous evaluation of the two issues.

Most consensus metrics measure the differences between the decision-maker evaluations or between their individual rankings. Consensus-building methods improve this metric iteratively by identifying the inputs that are to be adjusted, and then either prompting the group to discuss the issue, encouraging the affected decision-makers to adjust their input in the desired direction (Zhang et al., 2019), or even adjusting the inputs automatically and without their originators' knowledge (Xu, 2009). In each case, the goal is to 'correct' outlying values in a desired direction. In most cases, the desired direction is towards the group average. These steps are repeated until the consensus metric reaches a pre-defined threshold, and a soft consensus is reached. The evaluations of decision-makers who refuse to adjust their evaluations may be penalized (Chao et al., 2021; Dong et al., 2016), or even ignored entirely (Kacprzyk & Fedrizzi, 1988).

Moral et al. (2018) showed that the choice of distance metric, including Euclid, Manhattan and Cosine, had a significant effect on the convergence speed of the consensus process. Cai et al. (2023a) provide a survey of collaborative decision methods that covers determination of criteria weights and decision-maker weights as well as consensus measuring approaches. Tapia et al. (2023) study the Gini index as a dispersion-based, rather than distance-based consensus metric and conclude that it is appropriate for use in the MCGDM context.

Many approaches assign different weights to decision-makers. A survey is provided by Boix-Cots et al. (2023). Subjective approaches include estimates of decision-maker experience or competence (Janković & Popović, 2019). Objective approaches are based on input values (Koksalmis & Kabak, 2019), for example similarity to group average or consistency of ratio judgements. Yue derives weights for each decision-maker based on the entropy of their judgement matrices (Yue, 2017): the lower the entropy of the judgements, the greater their power to discriminate. Therefore, decision-makers with low-entropy judgements should be assigned a greater weight.

Some authors assume that input adjustments have a cost. (Zhang et al., 2019) present different efficiency metrics for the consensus-building process and derive optimal strategies for the judgement adjustments. Guo et al. (2023) consider the tolerance level of decision-makers towards adjusting their inputs.

The driving force towards consensus with these approaches is a model that postulates that arithmetical compromise best represents a set of varying evaluations. However, for consensus-committed groups (DeSanctis & Gallupe, 1987), this conflicts with the observation that minority arguments can be essential to the decision-making ability of a group (Nijstad et al., 2014). We conclude that – for a cooperative, rational group at least – a multi-criteria group decision method should not make any assumptions about the desirability of any particular decision-maker input, nor should it assume that a certain function of the inputs will produce an appropriate recommendation.

3. Combinatorial Multicriteria Acceptability Analysis

In the following, a multicriteria decision problem consists of m alternatives, denoted by a_i, n criteria, denoted by c_j and d decision-makers denoted by DM_k.

3.1. Motivation and basic approach

The goal of Combinatorial Multi-Criteria Acceptability Analysis is to provide a means for a group to use their preferred (single-user) multi-criteria decision method. The CMAA framework has been shown to work with commonly used decision models such as Simple Additive Weighting, TOPSIS and PROMETHEE (Goers & Horton, 2023).

The key attribute of CMAA, which distinguishes it from other approaches, lies in the aggregation of user inputs: instead of fusing them into a single mathematical entity such as an arithmetic mean, it retains them separately. It then generates instances of the decision by combining inputs from different decision-makers to form single-input decisions. The algorithm generates a discrete state space consisting of all instances of the problem, applying the decision algorithm to each instance, and updating various statistical quantities accordingly. One instance corresponds to the correct decision; this is the one in which each discrepancy had been resolved to the judgement or preference that results from the shared mental model for that issue. In practice, however, it is not necessary to find this instance, because a large number of instances yield the same preferred alternative, and only a comparatively small number of discrepancies must be resolved.

CMAA provides both decision analysis based on the current state of the user input and forward-looking guidance for an iterative consensus-building process. It only requires that the decision algorithm returns the preferred alternative for a given instance of the decision problem, or a complete ranking, if acceptabilities for lower ranks are of interest.

Decision-maker judgements and preferences are combined to form the *aggregated judgement* matrix A and *aggregated preference vector* P, respectively. These structures show where decision-makers agree or disagree about the performance of an alternative or the importance of a criterion. Examples of judgement and preference aggregation from the case study are shown in Table 3 and Table 4 in Section 4.3.

The overall CMAA decision model is written [P; A], where every instance of P is combined with every instance of A. The number of instances of the aggregated judgement matrix ||A|| is equal to the number of combinations of resolutions of each judgement discrepancy, and the number of instances of the aggregated preference vector ||P|| is equal to the number of valid preference vectors that can be generated from it. The overall number of instances for a given decision problem is $K = \|\mathrm{P}\| \cdot \|\mathrm{A}\|$. The computational complexity of the analysis grows with the number of discrepancies the problem contains and can become extremely large. Unacceptably long computation times can be avoided by Monte Carlo simulation, i.e., by only evaluating a random sample set of the instances.

Table 1. CMAA example

P	A			
{3}	{3, 4}	{4, 3}	{2, 3}	{3, 2}
{1}	{3}	{2, 1}	{3, 1}	{1, 2}

We illustrate the CMAA approach using the very small group decision with m=4, n=d=2 shown in Table 1. Values in curly brackets are the decision-maker inputs. The decision-makers submit the same judgement at (c_2, a_1) and the same preferences for both c_1 and c_2. All other judgements are discrepancies. We thus have $\|\mathrm{P}\| = 1$ and $\|\mathrm{A}\| = 2^7 = 128$. The decision model is Simple Additive Weighting.

3.2. Acceptability

The principal analysis variable in CMAA is the *rank acceptability index* b_i^r. This is the relative frequency within the space of problem instances with which alternative a_i achieves rank r. In this paper, we will only consider the rank 1 acceptability indices b_i^1, which we refer to simply as *acceptabilities*. Acceptability values returned by the CMAA algorithm yield a sum greater than 1 when the decision algorithm returns multiple preferred alternatives. Rank acceptability was introduced by Lahdelma & Salminen (2001) in the context of Stochastic Multicriteria Acceptability Analysis, where it was computed using samples from a continuous distribution that represents uncertainty in the inputs.

An alternative with a high acceptability has a large amount of support in the state space, and one with a low acceptability has little support. We will characterize these as 'strong' and 'weak' alternatives, respectively. It is important to recognize that these characterizations do not express the probability of selection of an alternative: under the DMM assumption, the result of the decision is already fixed, and the consensus iteration is a deterministic search algorithm.

In order to enable decision analysis and consensus-building, the acceptability of an alternative can be refined with two new analysis variables: the *judgement acceptability* and the *preference acceptability*. Their purpose is to measure the effect of each individual decision-maker input. They are defined as the proportion of rank 1 occurrences of each alternative within the search space that are due to each individual judgement or preference. Each variable can be computed to reflect either the current state or the potential future state of the decision. Table 6 in Section 4.4 shows examples of current judgement acceptabilities, and Table 11 in Section 5.1 show examples of potential preference acceptabilities from the case study.

In Table 1, the acceptabilities of alternatives a_1, a_2, a_3 and a_4 are 96/128, 32/128, 8/128 and 0, respectively. Based on these decision-maker inputs, a_1 is the strongest, because 75% of all instances return it as the preferred alternative. The acceptabilities sum to 136/128, because eight instances return a two-way tie for the preferred alternative.

3.3. Entropy-based consensus-building heuristic

Consensus-building in CMAA is based on identifying the resolution that advances consensus the furthest by using information entropy (Shannon, 1948). The entropy h of a probability vector p of dimension m is given by

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

Information entropy expresses the amount of uncertainty inherent in a set of probability values. The greater the separation of the p_i, the lower the entropy; the maximum value of $h = \log_2(1/m)$ is obtained when all p_i are equal, and the minimum value $h = 0$ is reached when $p_i = 1$ for some i, and all other values are 0. Thus, substituting the normalized acceptabilities b_i^1 into Eqn. (1) yields a measure of the degree of separation of the alternatives into those with weak and those with strong support.

Substituting the potential judgement acceptabilities gives the *judgement entropies*, and substituting the potential preference acceptabilities gives the *preference entropies*. These show how each resolution would affect the separation of the acceptabilities. Since p must be a probability vector, the acceptabilities must first be normalized, if the analysis contained instances that returned multiple rank 1 placements. The smallest of these values identifies the judgement or preference resolution that will lead to the greatest separation of rank 1 acceptabilities. The facilitator then recommends the discrepancy containing this resolution to the decision-makers for clarification. If the decision-makers agree on a resolution that reduces the entropy, they will have moved closer to consensus. Each resolution reduces the number of instances in the search space and may render one or more active discrepancies inactive (see Section 3.4). The process terminates, when:

- the entropy reaches 0, indicating that a hard consensus has been achieved,
- the entropy falls below a pre-determined threshold, indicating a soft consensus,
- no resolution will result in a change in the acceptabilities b_i^1, or
- an acceptability reaches a pre-determined threshold such as 0.95 and is considered 'strong enough' to determine the preferred alternative.

This entropy-based heuristic is greedy (myopic) because it only looks ahead one step. Goers & Horton (2023) have shown that it does not always find the shortest path to consensus. The heuristic is also optimistic, because it is oriented towards the most advantageous discrepancy resolution at each step. However, the decision-makers can, of course, resolve the discrepancy differently, which may reduce the entropy in the acceptabilities by a smaller amount, or even increase it.

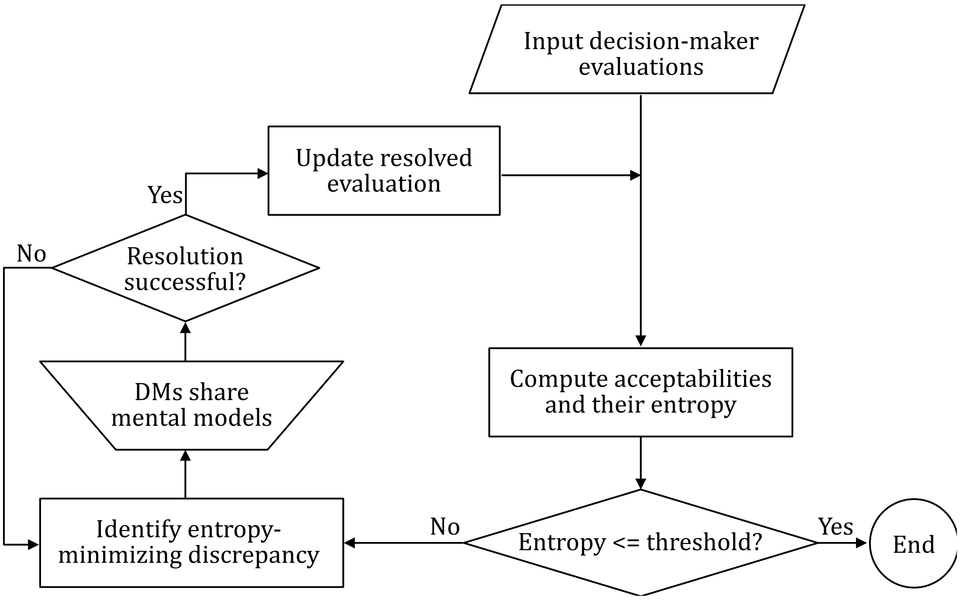

Figure 1. Flowchart of the CMAA consensus-building heuristic

Figure 1 shows the CMAA consensus-building process. In each iteration, the discrepancy containing the resolution resulting in the lowest entropy is identified. This discrepancy is discussed by the decision-makers, who (in the ideal case) return a unanimous evaluation; if they are unsuccessful, the next-best discrepancy is selected. The acceptabilities are re-computed for the new, simpler decision. These steps are repeated until the entropy of the acceptabilities reaches a specified threshold.

3.4. Inactive, active and pivot discrepancies

The key idea behind consensus-building in CMAA is to determine the impact of each discrepancy resolution on the acceptability of each alternative. In this context, we can distinguish between three types of discrepancy that we call 'active', 'inactive' and 'pivot'.

An *inactive discrepancy* is one whose resolution has no effect on the outcome of the decision and therefore can be ignored safely. For example, if a decision is determined solely by performances with respect to most important criteria, then discrepancies associated with low criteria weights or priorities are inactive. During consensus-building, an active judgement discrepancy (c_j, a_i) becomes inactive when alternative a_i is eliminated from the decision. The efficiency of the combinatorial consensus process stems from the fact that discrepancies can be converted from active to inactive status rapidly.

The various decision-maker judgements for criterion c_j and alternative a_{i2} are represented by $\lambda_k(j, i2)$, where the index k is local to each criterion. For example, in Table

3, the judgement discrepancy (c_1, a_4) has $\lambda_1(1,4) = B$ and $\lambda_2(1,4) = X$. The judgement discrepancy for criterion c_j and alternative a_{i2} is inactive when all resolution possibilities $\lambda_k(j, i2)$ for that discrepancy deliver the same potential acceptability for all alternatives.

The various decision-maker preferences for criterion c_j are represented by $\mu_k(j)$, where the index k is local to each criterion. For example, the preference discrepancy for criterion c_4 in Table 4 has the four possible resolutions $\mu_1(4) = 1$, $\mu_2(4) = 4$, $\mu_3(4) = 5$ and $\mu_4(4) = 6$. A preference discrepancy is inactive, when all resolution possibilities deliver the same potential acceptability for all alternatives.

Since all its resolutions lead to the same acceptability, resolving an inactive discrepancy would not advance the consensus-building process. Therefore, in the interest of efficiency, it should be ignored.

An *active discrepancy* is one in which different resolutions will lead to different acceptabilities for one or more alternatives. Every discrepancy is either active or inactive. The consensus-building process is steered by the sequence of active discrepancies that are selected for resolution. Consensus is complete when no active discrepancies remain.

The judgement discrepancy $(j, i2)$ is a *pivot discrepancy* (or simply a *pivot*) for alternative a_i, if it contains a resolution $\lambda_k(j, i2)$ that reduces the acceptability of a_i to 0. Similarly, the preference discrepancy for criterion c_j is a pivot for alternative a_i, if there is a resolution $\mu_k(j)$ that reduces the acceptability of a_i to 0. If the decision-makers select such a judgement or preference in a clarification conference, the affected alternatives are eliminated from the decision, and any associated discrepancies become inactive. This reduces the complexity of the decision and thereby accelerates the process towards convergence. Knowledge of the current active, inactive and pivot discrepancies helps the facilitator in monitoring and guiding the consensus-building process.

In Table 1, the {3,4} judgement discrepancy at (c_1, a_1) is a pivot because resolution to 4 results only in instances in which a_1 is the preferred alternative, eliminating the other three. In this case, hard consensus would be achieved in just one step. The two discrepancies associated with alternative a_4 are inactive because no combination of judgements can make this the preferred alternative. The {2,1} discrepancy at (c_2, a_2) is inactive, because regardless of how it is resolved, a_2 will be preferred if the discrepancy at (c_1, a_1) is resolved to 3, and it will not be preferred if it is resolved to 4. The remaining three discrepancies are active.

4. The product-development decision problem

4.1. The decision context

A group of three post-doctoral researchers at a leading public research institute in Germany was planning to found a biotechnology startup. Their company would develop and market two innovative technologies for cultivating viruses for use in vaccine manufacturing and gene therapy. The founders met the conditions for a cooperative, rational group, and none of them had any previous experience with formal decision-making methods.

Their first invention was a bioreactor for cultivating viruses in host cells. It was based on a novel design which promised faster throughput and eliminated the risk of virus mutations, which are highly undesirable in an industrial manufacturing process. The second invention was a purification device for separating the virus product from the cultivation medium. It used a new filter material that made the device cheaper and easier to use and gave a higher yield than existing technologies.

All three founders had PhDs in Bioprocess Engineering. Each of the first two founders was responsible for developing one of the technologies, and the third was responsible for business development. The decision was taken one year into a two-year government-funded development and entrepreneurship program, after which the founders would have to have acquired venture capital funding and launched their company. The founders had identified six possible laboratory experiments that would help them to develop these products further. The decision task was to select the experiment that delivered the greatest value.

As with any technology-based startup, the founders were confronted with a high degree of uncertainty. There was technological uncertainty whether the devices would perform as expected and be scalable from laboratory prototypes up to industrial dimensions. On the market side, it was unknown whether potential customers such as gene therapy research institutes and vaccine manufacturers would accept the new and unfamiliar technologies. For these reasons, the new devices would have to achieve significantly better performance than existing solutions and demonstrate successful application by independent test users.

4.2. The decision problem

The decision consisted of six alternatives, denoted by a_i, and six criteria, denoted by c_j. The three decision-makers are denoted by DM_d. The alternatives, which have been partially anonymized for reasons of confidentiality, were as follows:

- a_1: *Test whether the new virus can be cultivated in the current bioreactor using cell-type A.*
- a_2: *Test whether the new virus can be cultivated in the current bioreactor using cell-type B.*
- a_3: *Design, build and evaluate the performance of an updated version of the bioreactor.*
- a_4: *Test whether the new filter material can achieve comparable performance to the current material.*
- a_5: *Test whether the modified filter membrane improves performance.*
- a_6: *Design, build and evaluate a scaled-up version of the filter with four times the throughput.*

The alternatives can be categorized in different ways. For example, a_1, a_2 and a_3 were concerned with the bioreactor and were therefore competitors for the attention of the first technical co-founder, while a_4, a_5 and a_6 were concerned with the filter product and competed for the attention of the second technical co-founder. Concerning the time frames, alternatives a_1 and a_2 could be carried out with existing hardware, a_4 and a_5 required modifications to existing hardware, and a_3 and a_6 required building new hardware from scratch.

The 'new virus' in alternatives a_1 and a_2 refers to a COVID-19 vaccine candidate. The bioreactor was being developed for cultivating the adeno-associated virus, which is widely used in gene therapy. However, the COVID-19 pandemic broke out during the project, and COVID vaccine manufacturing suddenly became an important and unexpected new application possibility for the two technologies. Thus, these two alternatives were concerned with a new and potentially very important application of the technology, whereas the other four alternatives represented planned applications.

The six performance criteria supplied by the decision-makers were:

- c_1: *If successful, the product will attract more potential customers.* This is important to accelerate the early growth of the startup.
- c_2: *If successful, the idea underlying the resulting product will be patentable.* Patent protection is essential in order to prevent copycat products.
- c_3: *The experiment has a high probability of success.* 'Success' means that the experiment will deliver the hoped-for result.
- c_4: *The experiment will yield a result in four weeks or less.* 'Result' here means a definitive learning, be it positive or negative.
- c_5: *The experiment is relevant for developing an MVP.* A minimum viable product (MVP) is an early or partial version of a product that is used to gather user feedback. It is an essential component of the 'lean' approach to startup planning (Moogk, 2012).
- c_6: *A positive result will increase the chances of raising venture capital.* The founders knew they would need a large amount of venture capital in order to bring their products to market successfully.

None of the criteria are directly mutually contradictory. The opposite is often the case in product development, for example *The product is highly innovative/The product can be developed quickly.* However, they do address different areas of concern and different time frames. Criteria c_3 and c_4 are short-term issues, c_5 is concerned with validation, c_2 and c_6 are concerned with the attractiveness to investors, and c_1 is concerned with post-launch growth, which at the time was at least 18 months into the future.

The decision model was lexicographic, with three equivalence classes for the performance judgements:

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

This decision model is called 'ABX-Lex' (Horton & Goers, 2021). This is an ordered linguistic term set, which is preferable to numerical judgements for qualitative decisions (Li et al., 2020).

A non-compensatory decision model with ordinal criteria was used, because in an entrepreneurial context, a strong performance in a single high-priority criterion can in and of itself justify the inclusion of an alternative in the choice set. For example, if raising venture capital had top priority for the founders, then an alternative that performs strongly for it would be more attractive than others that score poorly for it, regardless of their strength with respect to lower-priority criteria.

Table 2. Individual performance judgements

	DM_1						DM_2						DM_3					
	a_1	a_2	a_3	a_4	a_5	a_6	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	A	B	B	A	A	A	B	X	X	A	A	A	B	B	X	B
c_2	B	B	A	A	B	A	A	A	A	A	A	B	A	A	A	A	B	B
c_3	B	B	A	B	B	A	B	B	B	A	A	A	B	B	B	B	B	A
c_4	B	B	B	A	A	B	B	B	B	A	A	X	X	B	A	B	B	B
c_5	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A
c_6	A	A	A	B	X	B	A	A	B	B	B	B	A	A	B	B	X	B

4.3. Decision-maker input

The decision-makers submitted their initial performance judgements and criteria preferences via email without conferring. Their performance judgements are shown in Table 2. The decision-makers were generally optimistic: only seven of the $3 \cdot 36 = 108$ judgements were judged to have performance X. We assume this is because the decision-makers were unlikely to propose alternatives that generally performed badly against their own criteria.

Table 3. Aggregated performance judgements

	a_1	a_2	a_3	a_4	a_5	a_6
c_1	A	A	AB	BX	BX	AB
c_2	AB	AB	A	A	AB	AB
c_3	B	B	AB	AB	AB	A
c_4	BX	B	AB	AB	AB	BX
c_5	A	A	A	A	AB	A
c_6	A	A	AB	B	BX	B

Table 3 shows the aggregated performance judgements. Of the 36 judgement tasks, 19 were discrepancies. All discrepancies were bivalent (AB or BX), so the total number of judgement instances generated by this input was $||A|| = 2^{19} = 524{,}288$.

Table 4. Individual (left) and aggregated (right) preference vectors

	DM_1	DM_2	DM_3	P
c_1	{3,4}	{4,5}	{1,2}	{1,2,3,4,5}
c_2	{4,5,6}	{2,3}	{3}	{2,3,4,5,6}
c_3	{4,5}	{1,2}	{1,2,3,4,5,6}	{1,2,3,4,5,6}
c_4	{1}	{4,5,6}	{5,6}	{1,4,5,6}
c_5	{2,3}	{1,2}	{3,4}	{1,2,3,4}
c_6	{4,5,6}	{3,4}	{2,3}	{2,3,4,5,6}

Table 4 shows the criteria preferences of the three decision-makers, together with the aggregated preference vector P. The decision-makers did not agree on a priority for any of the criteria, resulting in six preference discrepancies. Only in two cases was a decision-maker certain about their preferred priority of a criterion: DM_1 with $(c_4, 1)$ and DM_3 with $(c_2, 3)$. DM_3had no opinion about the priority for criterion c_3 and therefore submitted all possible priorities. The number of feasible criteria vectors for each decision-maker was quite small, namely 4, 2 and 2 for DM_1, DM_2 and DM_3, respectively. For example, the top four criteria positions for DM_3 are, in fact, uniquely determined. Because the priorities must be totally ordered, the aggregated preference vector generates only $||P|| = 200$ feasible criteria vectors. The total number of instances of the decision problem was therefore $K = 104{,}857{,}600$.

Table 5. Initial acceptability indices (before normalization)

a_1	a_2	a_3	a_4	a_5	a_6	Σ
0.12	0.21	0.40	0.11	0.05	0.19	1.08

4.4. Initial acceptability analysis

Table 5 shows the (rank 1) acceptabilities computed by the CMAA analysis algorithm, which yielded the ranking $a_3 \succ a_2 \succ a_6 \succ a_1 \succ a_4 \succ a_5$. The sum of the values is greater than 1, because some of the instances produced multiple winners. The entropy in the (normalized) acceptabilities according to Eqn. (1) was $h = 2.33$, which is close to the maximum possible value of 2.58, corresponding to the greatest uncertainty. The algorithm was coded in the C programming language without compiler optimizations and executed on a 2.3 GHz MacBook Pro. The computation time for analyzing the 104,857,600 instances was 76 seconds.

Alternative a_3 (*Design, build and evaluate the performance of an updated version of the bioreactor*) dominates by a large margin; it would have become the preferred alternative, if the decision had terminated at this point without a subsequent consensus-building phase. However, this conclusion would have been incorrect, as Section 5.1 will show.

Current acceptabilities give the decision-makers insight into how their existing judgements and preferences affect the acceptabilities of each alternative. The analysis provides sensitivity information and highlights the most influential discrepancies. This is particularly important if consensus-building is unlikely to succeed completely – for example when decision-makers represent different stakeholders who are unlikely to agree on criteria preferences. The decision-makers know which issues can be ignored and concentrate on resolving those that can provide most clarity.

Table 6. Initial current judgement acceptabilities for alternative a_3.

	a_1			a_2		a_3		a_4			a_5			a_6		
	A	B	X	A	B	A	B	A	B	X	A	B	X	A	B	X
c_1	-	-	-	-	-	0.28	0.12	-	0.20	0.20	-	0.20	0.20	0.19	0.22	-
c_2	0.20	0.21	-	0.20	0.21	-	-	-	-	-	0.20	0.20	-	0.20	0.21	-
c_3	-	-	-	-	-	0.28	0.12	0.20	0.21	-	0.20	0.20	-	-	-	-
c_4	-	0.20	0.20	-	-	0.26	0.14	0.19	0.21	-	0.19	0.21	-	-	0.20	0.20
c_5	-	-	-	-	-	-	-	-	-	-	0.20	0.20	-	-	-	-
c_6	-	-	-	-	-	0.25	0.16	-	-	-	-	0.20	0.20	-	-	-

Table 6 shows the initial current judgement acceptabilities for alternative a_3: *Build and test the updated bioreactor.* There are several discrepancies such as (c_6, a_5), where the sensitivity is 0 – both judgements B and X make equal contributions to this alternative's acceptability, and (at this point during the process) resolving the discrepancy would not help to clarify the status of this alternative. By contrast, at the discrepancy (c_1, a_3), the judgement A contributes more than twice as much support than the judgement B.

Analogous data exists for the other alternatives and for the criteria preferences at each step of the consensus iteration, from which similar observations can be made and conclusions drawn. The authors are not aware of any other approach to multicriteria group decision making that provides this amount of analytical detail.

CMAA is intended for use by consensus-committed groups in an iterative consensus-building process. However, there may be occasions where complete consensus is unlikely to be achieved. In such situations, the algorithm can still be used to establish partial successes, answer 'what-if' questions and provide guidance when individual clarification attempts fail:

- 'Unfortunately, we were not able to agree on the performance of alternative a_3: *Test updated bioreactor* with respect to criterion c_4: *Result in four weeks or* *less*. However, evaluating this alternative with respect to c_3: *The experiment has a high probability of success* also has a high potential to reduce the uncertainty in our decision.'

- 'Alternatives a_4: *Test the new filter material* and a_5: *Test the modified filter membrane* are the two weakest alternatives based on our initial input. If we could agree that criterion c_1: *Will attract more potential customers* has the highest priority, we could eliminate both of these alternatives immediately.'
- 'The attractiveness of both alternatives concerning the new virus depends strongly on the relative priorities of criteria c_4: *Result in four weeks* and c_5: *Relevant for MVP*. Perhaps we should discuss that question first.'

5. Consensus-building and decision-maker feedback

5.1. Consensus-building

In order to move forward from the initial acceptability analysis to a consensual decision, the authors facilitated a face-to-face meeting with the three founders. The consensus-building algorithm was used as an interactive digital assistant and to visualize the intermediate results.

Table 7. Initial judgement entropies

	a_1			a_2		a_3		a_4			a_5			a_6		
	A	B	X	A	B	A	B	A	B	X	A	B	X	A	B	X
c_1	-	-	-	-	-	2.07	2.44	-	2.35	2.31	-	2.35	2.32	2.30	2.28	-
c_2	2.36	2.23	-	2.27	2.32	-	-	-	-	-	2.35	2.32	-	2.32	2.33	-
c_3	-	-	-	-	-	2.04	2.45	2.35	2.28	-	2.36	2.30	-	-	-	-
c_4	-	2.37	2.23	-	-	2.02	2.50	2.32	2.25	-	2.40	2.21	-	-	2.33	2.33
c_5	-	-	-	-	-	-	-	-	-	-	2.35	2.31	-	-	-	-
c_6	-	-	-	-	-	2.20	2.42	-	-	-	-	2.34	2.32	-	-	-

The consensus process is driven by the entropies obtained for each possible discrepancy resolution. Table 7 shows the entropies for each resolution of a judgement discrepancy based on the original input, and Table 8 shows the initial entropies for each resolution of a preference discrepancy. At this stage, no judgement resolution could have had a significant effect on the entropy: the greatest improvement (i.e., reduction) would have been obtained by resolving (c_4, a_3) to A. The smallest value from either table is located at $(c_3, 1)$ among the preference entropies; the value $h = 1.56$ represents a substantial improvement in the separation of the alternatives. It is not surprising that selecting the dictator criterion would yield the greatest entropy improvement, owing to its dominating influence in the lexicographic decision model.

Table 8. Initial preference entropies

	Position in preference vector					
	1	2	3	4	5	6
c_1	1.92	2.32	2.30	2.26	2.18	-
c_2	-	2.28	2.31	2.34	2.34	2.35
c_3	1.56	2.16	2.32	2.23	2.15	2.13
c_4	1.86	-	-	2.21	2.15	2.19
c_5	2.16	2.33	2.34	2.34	-	-
c_6	-	2.18	2.30	2.31	2.27	2.24

The first clarification conference therefore considered the question, *Which priority should criterion c_3 have?* The decision-makers chose to answer the alternative question, *Which criterion should have top priority?* If, after exchanging their mental models for this question, the decision-makers had agreed on priority 1 for criterion c_3, the entropy in the rank 1 acceptabilities would have improved from 2.33 to 1.56, while resolution to other priorities would have resulted in much smaller improvements. Throughout the consensus-building process, the effects of the various resolutions on the entropy were not revealed to the decision-makers, in order to avoid the possibility of influencing their choice.

Table 9. Consensus path chosen by the decision-makers.

S	Res	b_1^1	b_2^1	b_3^1	b_4^1	b_5^1	b_6^1	h	Res_{EO}	h_A
-	-	0.12	0.21	0.40	0.11	0.05	0.19	2.33	$(c_3, 1)$	1.56
1	$(c_4, 1)$	0.02	0.05	0.51	0.28	0.17	0.02	1.86	$(c_4, a_3) = A$	0.58
2	$(c_4, a_3) = B$	0.05	0.11	0.10	0.47	0.31	0.03	2.04	$(c_4, a_4) = A$	0.49
3	$(c_4, a_4) = A$	0.00	0.00	0.00	0.91	0.11	0.00	0.49	$(c_4, a_5) = B$	0.00
4	$(c_4, a_5) = A$	0.00	0.00	0.00	0.82	0.22	0.00	0.74	$(c_1, a_4) = B$	0.48
5	$(c_4, a_4) = B$	0.00	0.00	0.00	0.92	0.11	0.00	0.48	$(c_3, a_4) = A$	0.20
6	$(c_3, a_4) = A$	0.00	0.00	0.00	1.00	0.03	0.00	0.20	$(c_1, a_5) = X$	0.00
7	$(c_1, a_5) = B$	0.00	0.00	0.00	1.00	0.06	0.00	0.32	$(c_2, a_5) = B$	0.00
8	$(c_1, a_5) = B$	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-	-

Table 9 shows the consensus path taken by the decision-makers. Shown for each step S are the resolution chosen by the decision-makers Res, the resulting acceptability indices for each alternative b_i^1, the entropy h in the acceptability indices, the resolution that would result in the greatest reduction in entropy at the next step Res_{EO}, and the resulting smallest achievable entropy h_A. Note that the raw acceptability values are shown, whose sum can be greater than 1. At steps 4, 5, 6 and 7, there was more than one resolution that minimized the entropy; in each case, the decision-makers selected the discrepancy to discuss they thought would be easiest to resolve. The preferred alternative at consensus in Step 8 was a_4: *Test whether the new filter material can achieve comparable performance to the current material.*

At step 1, the decision-makers agreed that criterion c_4: *The experiment will yield a result in four weeks or less* should have the highest priority. The founders were faced with many uncertainties regarding the manufacturability and performance of their devices, and it was important to reduce them quickly in order to avoid spending time and money on the wrong products. By choosing the resolution $(c_4, 1)$, the founders made learning speed their top priority.

In steps 1, 2, 4 and 7, the decision-makers did not choose the resolutions with the greatest entropy improvement, resulting in an entropy increase in three cases. In Steps

4 and 7, the entropy-optimal resolution of the pivot would have led directly to a hard consensus.

The dramatic improvement in entropy in step 3 results from a large increase in the rank 1 acceptability for alternative a_4 and the elimination of alternatives a_1, a_2, a_3 and a_6. The procedure could have been terminated at this point, declaring a_4 as the decision result. This consensus, although technically still soft, could perhaps be regarded as "very firm".

The significance of Step 3 can be understood by inspecting the aggregated performance judgements in Table 3. After Step 2, the decision-makers had selected the ability to yield results in four weeks (criterion c_4) as their top priority, and resolved the discrepancy at (c_4, a_3) to B. In this situation, resolving (c_4, a_4) to A at Step 3 immediately eliminated alternatives a_1, a_2, a_3 and a_6 (since a performance of A outranks performances B and BX), leaving only a_5 as a competitor. Thus, after only three resolution steps, it was already clear that the winning alternative would relate to the filter and not to the bioreactor. Significantly, alternatives a_4 and a_5 were the two weakest alternatives based on the initial input. This shows how misleading the initial acceptability indices can be: only three distributed mental models needed to be shared in order to overturn the initial result.

Steps 4 to 8 merely served to reduce the acceptability of a_5, the single remaining competitor to a_4, from 0.11 to 0.0. In fact, from Step 6 on, alternative a_5 could no longer overtake a_4 – it could at best share top rank. This can be concluded from the raw acceptabilities. Alternative a_4 receives an acceptability of 1.0 in step 6, which means it achieves rank 1 in every instance, whereas a_5 only achieves rank 1 in 3% of all instances.

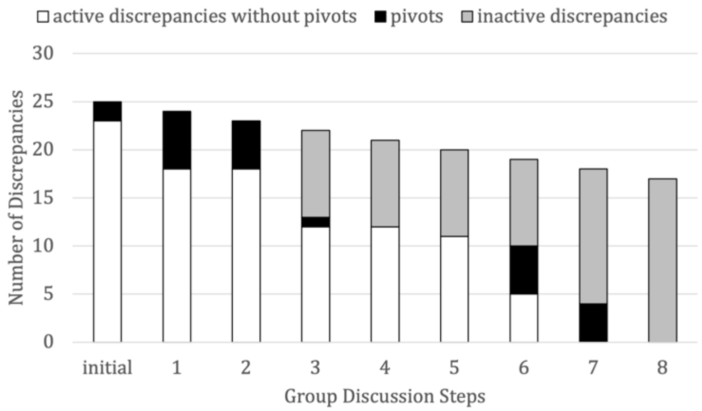

Figure 2. Number of judgement and preference discrepancies of each type during consensus-building

Figure 2 shows the number of each type of judgement and preference discrepancies at each step of the consensus-building process. The initial decision-maker input contained 19 judgement discrepancies and six preference discrepancies, all of which are active. The large number of discrepancies that switched from active to inactive in Step 3 is a result of the group choosing the entropy-optimal resolution that eliminated the four alternatives a_1, a_2, a_3 and a_6. At Step 3, only two alternatives remained, and a resolution to B would have led to an immediate consensus. However, in the clarification conference, the decision-makers chose the judgement A, so the process continued.

This situation repeated in Step 6. Finally, in Step 7, the decision-makers chose in favor of the judgement that eliminated the last remaining rival to alternative a_4. After Step 8, all remaining judgement discrepancies are (by definition) inactive. All the remaining instances returned a_4 as the preferred alternative.

Table 10 shows the aggregated judgement matrix **A** and preference vector **P** after consensus was reached. Unique values that resulted from the clarification of discrepancies are underlined. 12 judgement discrepancies remained, yielding $||\mathbf{A}|| = 2^{12} = 4{,}096$. These discrepancies were all, by definition, inactive and could safely be ignored. In the aggregated preference vector, only one discrepancy needed to be resolved: criterion c_4: *The experiment will yield a result in four weeks or less* was assigned to the top priority. The number of instances of the final preference vector was $||\mathbf{P}|| = 54$. The total number of instances remaining at consensus was thus 221,184.

Table 10. Aggregated judgement matrix A and preference vector P at consensus

	A						**P**
	a_1	a_2	a_3	a_4	a_5	a_6	
c_1	A	A	AB	<u>B</u>	<u>B</u>	AB	{2,3,4,5}
c_2	AB	AB	A	A	<u>B</u>	AB	{2,3,4,5,6}
c_3	B	B	AB	<u>A</u>	AB	A	{2,3,4,5,6}
c_4	BX	B	<u>B</u>	<u>A</u>	<u>A</u>	BX	<u>{1}</u>
c_5	A	A	A	A	AB	A	{2,3,4,5}
c_6	A	A	AB	B	BX	B	{2,3,4,5}

During the consensus-building phase, the potential acceptabilities show the decision-makers how each resolution would affect the acceptabilities in the next step. This information is used to compute the entropies and can provide facilitator guidance at critical junctures.

Table 11. Initial potential preference acceptabilities for alternative a_2

	Position in criteria vector					
	1	2	3	4	5	6
c_1	0.368	0.196	0.150	0.106	0.079	-
c_2	-	0.185	0.205	0.198	0.197	0.187
c_3	0.000	0.075	0.208	0.293	0.307	0.284
c_4	0.050	-	-	0.266	0.257	0.219
c_5	0.283	0.160	0.160	0.139	-	-
c_6	-	0.331	0.258	0.171	0.129	0.093

For example, Table 11 shows the potential preference acceptabilities for alternative a_2: *Test whether the new virus can be cultivated in the current bioreactor using cell-type B* after the initial input. The most beneficial preference resolution from the point of view of this alternative is $(c_1, 1)$, which would improve its acceptability from 0.213 to 0.368. The discrepancy at c_3 is a pivot since the resolution $(c_3, 1)$ would eliminate a_2 from the decision. During the actual decision process, the decision-makers chose $(c_4, 1)$, which weakened a_2 considerably.

Analogous data exists for the other alternatives and for the performance judgements at each step of the consensus iteration, from which similar observations can be made and conclusions drawn.

5.2. Feedback from the decision-makers

One objective with this study was to record any reservations by the decision-makers about the method and the results obtained. To this end, we conducted a debriefing that consisted of the following five questions:

- Q: *Were you happy with the number and meaning of the three judgement categories A, B and X?*
  A: *Yes, they were sufficient. A larger number of categories would have increased the cognitive load.*
- Q: *1) Did you understand the dictator property of non-compensatory criteria and 2) did it concern you?*
  A: *1) Yes; 2) No*
- Q: *Does the DMM approach (i.e., the assumption that discrepancies are caused by differing mental models and that they can be resolved by sharing them) seem applicable?*
  A: *Yes, but it requires the group to be cooperative.*
- Q: *Was the consensus-building transparent and understandable?*
  A: *Yes.*
- Q: *How do you feel about the fact that there were unresolved discrepancies remaining after consensus had been reached?*
  A: *It was surprising at first, but understandable and not a concern.*

We conclude that the CMAA approach and the decision model were appropriate and acceptable and achieved the desired transparency.

A second objective with this study was to uncover opportunities for improvement of the CMAA algorithm. We observed that the format used to represent the decision-makers' criteria preferences was not appropriate. This had two negative consequences: The first was that decision-makers' input contained redundancies which they were not aware of, as described in Section 4.3. The second negative consequence was that the cognitive load for stating preferences and for resolving preference discrepancies was too high. The tasks *Which criterion should occupy this position in the criteria vector?* or *Which of the following positions in the criteria vector could this criterion occupy?* are non-atomic questions that involve multiple options and are not in the spirit of multi-criteria decision models. In the debriefing session, all three decision-makers stated that they would have preferred to express their preferences in the form $c_{j1} \succ c_{j2}$ (*Criterion c_{j1} is more important to me than criterion c_{j2}*), rather than name absolute priorities for each criterion. They stated this would have been easier, faster, and more intuitive.

We assume that it is more likely that a decision-maker will have a mental model that provides an argument for *Criterion c_{j1} is more important than criterion c_{j2}* than one for *Criterion c_{j1} might belong at priority p*. For this reason, the input becomes more intuitive, and a clarification conference for the question *Which of the criteria c_{j1} and c_{j2} is more important?* is more straightforward than *What priority should criterion c_{j1} have?* or *Which criterion should have priority p?*

6. Simulation experiments

6.1. Accuracy of the Monte Carlo simulation

If the number of decision instances is very large, Monte Carlo simulation can be used instead of sampling the entire state space. This raises the question of whether a limited number of random samples can still achieve sufficient accuracy.

Table 12. Acceptability indices for full sampling and random sampling

	a_1	a_2	a_3	a_4	a_5	a_6	h
10,000 samples	0.123	0.203	0.415	0.111	0.057	0.193	2.411
Full problem	0.124	0.213	0.404	0.112	0.054	0.191	2.414

To investigate this question, we performed a CMAA analysis of the decision problem defined by Table 3 and Table 4, using Monte Carlo simulation with 10,000 random instances. This represents about 0.01% of the overall state space.

Table 12 compares the acceptabilities obtained from the complete analysis (see Table 5) with those obtained with the Monte Carlo simulation, as well as their respective entropies h (after normalization). All acceptabilities are within 5% of the ground truth values. The difference in the entropies is about 0.1%, and both variants generated the same consensus path. Even if a small deviation in the entropy did cause the Monte Carlo-based method to select a different discrepancy to be resolved, the expected consensus path length will almost certainly be unaffected.

We conclude that – for this decision problem at least – 10,000 random samples would have been sufficient. The computation time for the analysis could have been reduced to less than one second.

6.2. Consensus path length

The decision-makers needed eight clarification steps to achieve consensus. We now consider the questions whether this was a representative number of steps for this particular problem and whether the decision problem itself was typical in this regard.

We performed a Monte Carlo simulation in which 10,000 consensus paths were generated for the decision problem. At each step, the discrepancy that promised the greatest decrease in entropy was selected, and the decision-makers were simulated by choosing resolutions randomly with equal probability. In Figure 3. Distribution of path lengths for the decision problem and similarly dimensioned random problems, the darker columns show the distribution of path lengths needed to achieve consensus, using the primary vertical axis. The shortest path had only four steps – one of these selects the entropy-minimizing resolution at each step. The values tail off rapidly, so the probability of a long path is small. The mean number of steps to consensus is 7.7. We conclude that the number of steps needed to achieve consensus in the real-life case was representative. In other words, it was not a lucky result of the specific resolutions chosen by the decision-makers: on average, other choices would have led to a similar number of steps.

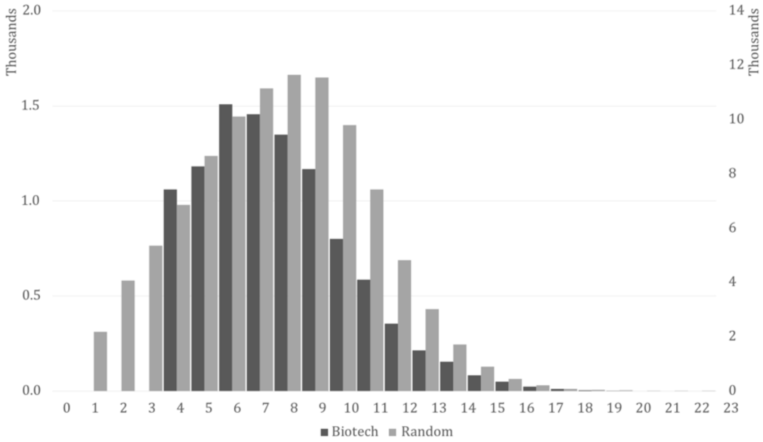

Figure 3. Distribution of path lengths for the decision problem and similarly dimensioned random problems

The lighter columns in Figure 3 show the distribution of consensus path lengths for a set of 1,000 randomly generated decision problems of size $m = n = 6$ using the same decision model and solution method and with similar values for ||A|| and ||P||. For each decision problem, 100 random consensus paths were simulated, resulting in 100,000 simulated decisions overall. Its values correspond to the secondary vertical axis. The general shape of the distribution is very similar to that obtained using the real-life input, with the exception that there were decisions where fewer than four clarification steps were needed. The mean number of steps to achieve consensus is 7.6. We conclude that the specific 6×6 problem in the case study was typical, in the sense that the number of steps to consensus was very close to the expected value for comparable problems.

6.3. Comparison with compensatory decision models

We hypothesize that CMAA will reach consensus in fewer steps using the non-compensatory ABX-Lex decision model than with a compensatory one. This is because of its ‘dictator’ property: in descending order of criterion priority, all alternatives with a less-than-highest score with respect to that criterion are eliminated. Thus, any resolution of a discrepancy during the consensus-building process may eliminate several alternatives simultaneously.

We tested this hypothesis with a simulation experiment using three compensatory decision models: Simple Additive Weighting (SAW), TOPSIS and Weighted Product Model (WPM). The decision-maker inputs of Table 2 and Table 3 were re-interpreted as numerical performance scores and criteria weights. The judgements A, B and X were replaced by scores 3, 2 and 1, respectively, and the criteria priorities 1...6 were replaced by the weights 6...1.

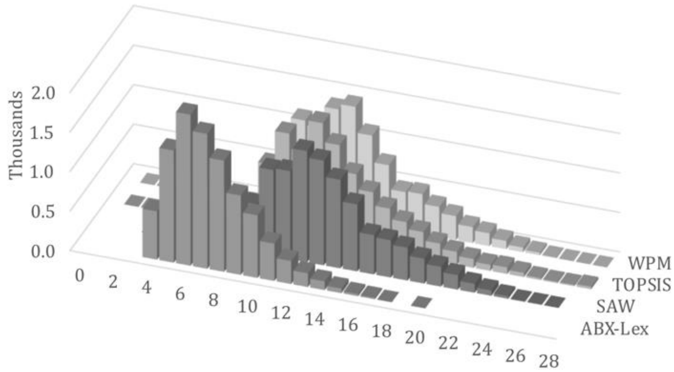

Figure 4. Distribution of CMAA consensus path lengths for ABX-Lex and three compensatory models

10,000 consensus paths were generated, whereby each discrepancy was resolved randomly. Figure 4 shows the distributions of the consensus path lengths for each model. The histogram for ABX-Lex is clearly shifted to the left compared to the other three. The mean path length obtained with ABX-Lex is 7.53, compared to 13.55, 13.17 and 13.45 for SAW, TOPSIS and WPM, respectively, providing support for the hypothesis.

7. Conclusions

7.1. Summary

Combinatorial Multicriteria Acceptability Analysis is an algorithmic framework that permits standard multicriteria decision methods to be used by groups without aggregating individual evaluations to a single value such as the arithmetic mean. It provides a detailed decision analysis based on acceptability variables and a guided consensus-building process that attempts to minimize their information entropy.

CMAA was used to facilitate a product development decision by the founders of a biotechnology startup. This is the first application of CMAA to a real-world decision. The entropy-optimal consensus path would have required only four steps. However, the decision-makers diverged from this shortest path four times, leading to eight resolution steps in total, each of which lasted only a few minutes. The strongest alternative based on the initial input did not become the consensus alternative, highlighting the importance of including a clarification process to identify the correct alternative. The decision-makers were satisfied with the result and reported no problems or concerns with the model or the process.

Three computational experiments were performed. The first studied the consensus path length for the decision, yielding an expected value of 7.7 resolution steps. The second experiment showed that Monte Carlo simulation with 10,000 samples provided sufficient accuracy, even though the complete decision contained more than 100,000,000 instances. The third confirmed our expectation that CMAA reaches consensus more quickly with a non-compensatory decision model than with a compensatory one.

Our conclusions are based on single experiments, and further validation is needed. The case study was based on a comparatively small decision with only six alternatives and criteria and three decision-makers; the efficiency of the consensus-building and acceptance by the decision-makers should be tested for larger dimensions.

The decision-makers were a consensus committed, cooperative and rational group, and all discrepancies were resolved quickly and without controversy. The CMAA approach is tolerant of failed or partial resolutions, which can result in longer paths to consensus, and it would be instructive to observe its performance with a less ideal group of decision-makers.

The absolute values used for the criteria preferences led to a higher cognitive load than pairwise comparisons (such as are used in AHP) would have done, and they also resulted in redundant input combinations. If ordinal preferences are to be used, then pairwise comparisons would be more appropriate.

Various future developments of the CMAA algorithm suggest themselves. Entropy could be computed across multiple ranks if a complete ranking is to be determined. In its current form, the consensus-building algorithm is myopic – it only considers the next resolution step; it may be possible to discover discrepancies with a greater entropy-reducing potential by extending the look-ahead to more than one step. Similarly, it might be more efficient to select the discrepancy to be resolved based on the expected value of the resulting entropy, rather on than the minimum value. Finally, a fully automated digital facilitator could be developed for use in conjunction with an internet-based communication tool.

7.2. Recommendations

We recommend Combinatorial Multicriteria Acceptability Analysis as a multicriteria group decision-making method if three conditions can be met.

First, the decision is being made by a rational, cooperative group that is committed to consensus. This implies that decision-makers will likely be able to resolve their differences once they have shared their information and interpretations of each issue.

Second, performance is subjective: even if objective measurements are available, decision-makers will be required to make value judgements of them, for example using a Likert scale, or even just a pass/fail choice.

Third, if a non-compensatory decision model is to be used, the decision-makers should be aware of the consequences: an alternative that is not assigned to the highest value/performance category with respect to the highest-priority criterion can no longer be 'rehabilitated' by a high score in a lower-priority criterion. If this feature is not desired, a more familiar model such as Simple Additive Weighting, AHP or TOPSIS can be used instead, possibly at the cost of a longer consensus process.

If these conditions are fulfilled, CMAA offers the following benefits to a decision-making group:

- Hard consensus can be achieved in a small number of steps, without having to resolve all discrepancies, even when there is a large amount of initial disagreement.
- The cognitive load is low, since most discrepancies do not need to be addressed at all.
- The method encourages the sharing of mental models, leading to a common understanding of critical issues within the group.
- The method delivers detailed insights into the effect of individual inputs on the strength of each alternative.

Acknowledgements: The authors would like to thank Julián López-Meza, Pavel Marichal-Gallardo and Felipe Tapia for supplying the decision problem and for their participation in the consensus-building process.

Author Contributions: J.G. was the principal contributor for methodology, algorithm implementation, validation and computational experiments. G.H. was the principal contributor to the original draft and editing of the manuscript. Both authors contributed in equal measure to all other aspects of the study. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bairagi, B. (2022). A fuzzy interval based multi-criteria homogeneous group decision making technique: An application to airports ranking problem. *Decision Making: Applications in Management and Engineering*, *6*(2), 1–15. https://doi.org/10.31181/dmame622023410

Boix-Cots, D., Pardo-Bosch, F., & Alvarez, P. P. (2023). A systematic review on multi-criteria group decision-making methods based on weights: Analysis and classification scheme. *Information Fusion*, *96*(2), 16–36. https://doi.org/10.1016/j.inffus.2023.03.004

Bortolini, R. F., Cortimiglia, M. N., de Moura Ferreira Danilevicz, A., & Ghezzi, A. (2021). Lean Startup: a comprehensive historical review. *Management Decision*, *59*(8), 1765–1783. https://doi.org/10.1108/MD-07-2017-0663

Cai, Y., Jin, F., Liu, J., Zhou, L., & Tao, Z. (2023). A survey of collaborative decision-making: Bibliometrics, preliminaries, methodologies, applications and future directions. *Engineering Applications of Artificial Intelligence*, *122*, 106064. https://doi.org/10.1016/j.engappai.2023.106064

Chao, X., Kou, G., Peng, Y., & Viedma, E. H. (2021). Large-scale group decision-making with non-cooperative behaviors and heterogeneous preferences: An application in financial inclusion. *European Journal of Operational Research*, *288*(1), 271–293. https://doi.org/10.1016/j.ejor.2020.05.047

DeSanctis, G., & Gallupe, R. B. (1987). A foundation for the study of group decision support systems. *Management Science*, *33*(5), 589–609. https://doi.org/10.1287/mnsc.33.5.589

Dong, Y., Zhang, H., & 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. https://doi.org/10.1016/j.dss.2016.01.002

Dreu, C. K. W. De, & West, M. A. (2001). Minority dissent and team innovation: The importance of participation in decision making. *Journal of Applied Psychology*, *86*(6), 1191. https://doi.org/10.1037/0021-9010.86.6.1191

Goers, J., & Horton, G. (2023). Combinatorial Multi-Criteria Acceptability Analysis: A Decision Analysis and Consensus-Building Approach for Cooperative Groups. *European Journal of Operational Research*, *308*(1), 243–254. https://doi.org/10.1016/j.ejor.2022.12.002

Guo, W., Gong, Z., Zhang, W.-G., & Xu, Y. (2023). Minimum cost consensus modeling under dynamic feedback regulation mechanism considering consensus principle and tolerance level. *European Journal of Operational Research*, *306*(3), 1279–1295. https://doi.org/10.1016/j.ejor.2022.08.033

Herrera-Viedma, E., Cabrerizo, F. J., Kacprzyk, J., & Pedrycz, W. (2014). A Review of Soft Consensus Models in a Fuzzy Environment. *Information Fusion*, *17*, 4–13. https://doi.org/10.1016/j.inffus.2013.04.002

Horton, G., & 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. https://doi.org/10.1142/S0219622020500431

Janković, A., & Popović, M. (2019). Methods for assigning weights to decision makers in group AHP decision-making. *Decision Making: Applications in Management and Engineering*, *2*(1), 147–165. https://doi.org/10.31181/dmame1901147j

Kacprzyk, J., & Fedrizzi, M. (1988). A 'soft' measure of consensus in the setting of partial (fuzzy) preferences. *European Journal of Operational Research*, *34*(3), 316–325. https://doi.org/10.1016/0377-2217(88)90152-X

Kline, D. A. (2005). Intuitive team decision making. In H. Montgomery, R. Lipshitz, & B. Brehmer (Eds.), *How professionals make decisions* (pp. 171–182). Lawrence Erlbaum Associates.

Koksalmis, E., & Kabak, Ö. (2019). Deriving decision makers' weights in group decision making: An overview of objective methods. *Information Fusion*, *49*, 146–160. https://doi.org/10.1016/j.inffus.2018.11.009

Lahdelma, R., & Salminen, P. (2001). SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making. *Operations Research*, *49*(3), 444–454. https://doi.org/10.1287/opre.49.3.444.11220

Li, Y., Chen, X., Dong, Y., & Herrera, F. (2020). Linguistic group decision making: Axiomatic distance and minimum cost consensus. *Information Sciences*, *541*, 242–258. https://doi.org/10.1016/j.ins.2020.06.033

Lu, L., Yuan, Y. C., & McLeod, P. L. (2012). Twenty-Five Years of Hidden Profiles in Group Decision Making: A Meta-Analysis. *Personality and Social Psychology Review*, *16*(1), 54–75. https://doi.org/10.1177/1088868311417243

Moogk, R. (2012). Minimum Viable Product and the Importance of Experimentation in Technology Startups. *Technology Innovation Management Review*, *2*(3), 23–26. https://doi.org/10.22215/timreview/535

Moral, M. J. Del, Chiclana, F., Tapia, J. M., & Herrera-Viedma, E. (2018). A comparative study on consensus measures in group decision making. *International Journal of Intelligent Systems*, *33*(8), 1624–1638. https://doi.org/10.1002/int.21954

Nijstad, B. A., Berger-Selman, F., & Dreu, C. K. W. De. (2014). Innovation in top management teams: Minority dissent, transformational leadership, and radical innovations. *European Journal of Work and Organizational Psychology*, *23*(2), 310–322. https://doi.org/10.1080/1359432X.2012.734038

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Business.

Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R., & 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. https://doi.org/10.1037/0022-3514.91.6.1080

Schulz-Hardt, S., & 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. https://doi.org/10.1080/10463283.2012.744440

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(4), 623–666. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Stasser, G., & 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. https://doi.org/10.1037/0022-3514.48.6.1467

Tapia, J. M., Chiclana, F., del Moral, M. J., & Herrera–Viedma, E. (2023). Measuring Consensus in Group Decision-Making Problems Through an Inequality Measure. *International Conference on Computers Communications and Control*, 313–319. https://doi.org/10.13039/501100011033

Tech, R. P. G. (2018). Introduction: High-Tech Startup Financing. In *Financing High-Tech Startups: Using Productive Signaling to Efficiently Overcome the Liability of Complexity* (pp. 1–28). Springer International Publishing.

De Vreede, T. De, Reiter-Palmon, R., & Vreede, G.-J. De. (2013). The Effect of Shared Mental Models on Consensus. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 263–272. https://doi.org/10.1109/HICSS.2013.517

Xu, Z. (2009). An automatic approach to reaching consensus in multiple attribute group decision making. *Computers & Industrial Engineering*, *56*(4), 1369–1374.

Yahaya, S.-Y., & Abu-Bakar, N. (2007). New product development management issues and decision-making approaches. *Management Decision*, *45*(7), 1123–1142. https://doi.org/10.1108/00251740710773943

Yue, C. (2017). Entropy-based weights on decision makers in group decision-making setting with hybrid preference representations. *Applied Soft Computing*, *60*, 737–749. https://doi.org/10.1016/j.asoc.2017.07.033

Zhang, H., Dong, Y., Chiclana, F., & Yu, S. (2019). Consensus efficiency in group decision making: A comprehensive comparative study and its optimal design. *European Journal of Operational Research*, *275*(2), 580–598. https://doi.org/10.1016/j.ejor.2018.11.052