Title of dissertation:

Combinatorial Multicriteria Acceptability Analysis For IT-Supported Group Decisions: Detailed Analysis, Enhanced Metric and Efficient Consensus-Building

A Dissertation
accepted by the Faculty of Computer Science at the Otto-von-Guericke University Magdeburg

Jana Görs
Dipl. Ing., born 20.05.1982 in Magdeburg

Reviewers:
Prof. Dr. Graham Horton,
Prof. Dr. Gert-Jan de Vreede and
Prof. Dr. Lennart Nacke

Magdeburg, 26. September 2024

Thema der Dissertation:

Combinatorial Multicriteria Acceptability Analysis For IT-Supported Group Decisions: Detailed Analysis, Enhanced Metric and Efficient Consensus-Building

Dissertation
angenommen durch die Fakultät für Informatik der
Otto-von-Guericke-Universität Magdeburg

Jana Görs
Dipl. Ing., geb. am 20.05.1982 in Magdeburg

Gutachter:
Prof. Dr. Graham Horton,
Prof. Dr. Gert-Jan de Vreede und
Prof. Dr. Lennart Nacke

Magdeburg, 26. September 2024

Zusammenfassung

Diese Dissertation stellt ein neues algorithmisches Framework für kooperative, mutlikriterielle Gruppenentscheidungen vor. Bei dieser Art von Entscheidung haben alle Beteiligten ein gemeinsames Ziel, und sie arbeiten zusammen, indem sie ihr Wissen und ihre Expertise einbringen, um die bestmögliche Entscheidung zu erreichen. Beispiele für kooperative Gruppenentscheidungen sind ein interdisziplinäres Ärzteteam bei einem Tumorboard, das die wirksamste Behandlung für einen Krebspatienten beschließt, oder ein Ingenieurteam, das eine neue Maschine für eine Produktionsstätte aussucht. Bewertungskonflikte werden hier durch Informationsaustausch gelöst. Dieser Austausch führt dazu, dass die Entscheidung in der Gruppe besser ist, als die, die jedes einzelne Mitglied der Gruppe allein hätte treffen können.

Im Gegensatz zu kooperativen Gruppenentscheidungen haben die Beteiligten an einer konkurrierenden Entscheidung unterschiedliche (und oft gegensätzliche) Ziele. In diesem Fall sind Verhandlungen erforderlich, um Bewertungskonflikte zu lösen. Konfliktauflösungen sind nur durch Kompromisse möglich, mit denen aber keiner der Beteiligten wirklich zufrieden ist. Beide Arten von Gruppenentscheidungen sind in Organisationen verbreitet und erzeugen zwangsläufig Bewertungskonflikte. Der Unterschied zwischen den beiden Gruppenentscheidungen liegt in der Art der Auflösung von Bewertungskonflikten.

Herkömmliche Ansätze zur multikriteriellen Gruppenentscheidung eignen sich für kooperative Gruppen aus mehreren Gründen nicht. Erstens neigen die Konsensmetriken zu übertriebenem Optimistismus, sodass sie einen Konsens anzeigen, bevor er überhaupt erreicht wurde. Zweitens schlagen viele Methoden Kompromisse vor, anstatt Einigkeit zu fördern. Drittens können diese Ansätze sogar daran scheitern, einen Konsens überhaupt herbeizuführen. Darum benötigen kooperative Gruppen einen geeigneteren Ansatz zur Konsensbildung, der diese Mängel vermeidet.

Diese Forschungsarbeit führt ein neues algorithmisches Framework, die Combinatorial Multicriteria Acceptability Analysis (CMAA), ein, das die Bedürfnisse kooperativer Gruppen bedient. Statt der üblichen Reduktion von Bewertungskonflikten auf einzelne Kompromisswerte begreift CMAA einen Bewertungskonflikt als Indikator für entscheidungskritische, aber verborgene Informationen. Das ist eine wesentliche Veränderung gegenüber bisherigen multikriteriellen Gruppenentscheidungen. Werden nun alle verschiedenen Kombinationen aus allen Bewertungseingaben betrachtet, entsteht ein kombinatorischer Suchraum, für den neue Analyse- und Konsensbildungsalgorithmen entwickelt werden können.

CMAA führt eine innovative, auf Entropie basierende Konsensmetrik ein. Diese Metrik kann sowohl den aktuellen Konsens ermitteln als auch den, der erreicht wird, wenn die Gruppe eine bestimmte Bewertung zur Konfliktauflösung wählen würde. Dies ermöglicht eine 'greedy' Heuristik, die den Bewertungskonflikt identifizieren kann, der das größte Potenzial zur Verbesserung des Konsens ermöglichen würde. Diese neue Metrik ist einzigartig, weil sie erkennt, ob noch ein Bewertungskonflikt existiert, dessen Auflösung das Entscheidungsergebnis ändern würde.

Simulationsexperimente bestätigen die Fähigkeit von CMAA, einen Konsens effizienter und vollständiger als bisherige Metriken zu erreichen. Dies trifft sogar für Entscheidungen zu, bei denen Bewertungskonflikte mit Kompromissen aufgelöst wurden. In vielen Fällen kann ein großer Teil des kombinatorischen Zustandsraums für die Entwicklung eines Konsens sogar ignoriert werden. In diesen Fällen reicht die Auflösung nur eines Bruchteils der ursprünglichen Bewertungskonflikte aus, um einen Konsens zu erreichen.

Der kombinatorische Zustandsraum wächst exponentiell mit der Größe der Entscheidung und wird schnell unbeherrschbar. Allerdings kann die Monte-Carlo-Simulation verwendet werden. Sie liefert eine ausreichende Genauigkeit in weniger als einer Sekunde auf einem Notebook-Rechner, sodass sie in Live-Situationen eingesetzt werden kann.

Zwei Fallstudien wurden durchgeführt: eine Gruppenentscheidung in einem Biotechnologie-Startup und eine andere in einem Softwareunternehmen. Bei beiden wurden CMAA als digitaler Assistent für einen menschlichen Moderator eingesetzt. Alle Annahmen, die der Methode zugrunde liegen, konnten validiert werden, insbesondere in Bezug auf die Vertrauenswürdigkeit der Methode in den Augen der Entscheidungsträger und ihre Zufriedenheit mit den Ergebnissen.

Basierend auf dem neuen kombinatorischen Modell bietet CMAA verbesserte Algorithmen für die drei Basiskomponenten von multikritieriellen Gruppenentscheidungen:

- Analysevariablen,
- Konsensmetrik und
- Konsensbildungsprozess.

Die neuen Analysevariablen bieten eine große Menge an Details. Sie zeigen, wie die Leistung jeder Alternative von den einzelnen Bewertungen abhängt und berücksichtigen dabei sämtliche mögliche Kombinationen der anderen Eingaben. Die neue Konsensmetrik setzt einen höheren Standard für Konsens: Die Entscheidenden wissen am Ende nicht nur, welche Alternative sie gemeinsam bevorzugen, sondern auch, dass keine weitere Auflösung eines Bewertungskonflikts dieses Ergebnis ändern kann. Der Konsensbildungsprozess basiert auf der Kenntnis des erreichbaren Konsensniveaus für jede Konfliktauflösung, auf die die Entscheidenden sich im nächsten Schritt einigen könnten. Zusammen ermöglichen die drei neuen Basiskomponenten des CMAA Frameworks eine effiziente Konsensbildung für kooperative Gruppen, die - bis auf kleine Anpassungen an das jeweilige Datenformat - mit jeder Entscheidungsmethode zusammenarbeiten kann.

Summary

This dissertation introduces a new algorithmic framework for cooperative multi-criteria group decision-making. In this class of decisions, group members are united by a common objective, working collaboratively to share their knowledge and expertise in order to arrive at the optimal decision. Examples include a multidisciplinary team of physicians on a tumour board deciding on the most effective treatment for a cancer patient or an engineering team deciding on the selection of a new machine for a production facility. Conflicting evaluations are resolved by sharing information, leading to a decision that is superior to the one that any decision-maker could have made on their own.

In contrast to cooperative decisions, group members in competitive decisions have different (and often conflicting) goals. Negotiation is needed to resolve conflicts, resulting in compromises that no party is completely happy with. Both types of decisions are prevalent in organisations, and both inevitably create conflicting evaluations. The distinguishing factor between the two types lies in the approaches taken to address conflicting evaluations.

Conventional multi-criteria group decision approaches prove inadequate for cooperative decisions for several reasons. First, their consensus metrics tend to be overly optimistic, and they signal consensus before it has been reached. Second, many methods propose compromises rather than encouraging agreement. Last, these approaches may sometimes fail to deliver consensus altogether. Therefore, cooperative decisions require a more suitable consensus-building approach that avoids these shortcomings.

The research described in this thesis addresses the needs of cooperative multi-criteria group decision-making by introducing a novel algorithmic framework called Combinatorial Multicriteria Acceptability Analysis (CMAA). A key attribute of CMAA is its interpretation of conflicting evaluations as potential indicators of decision-critical unshared information, rather than the traditional practice of aggregating conflicting evaluations to single compromise values. This approach results in a search space consisting of all possible combinations of decision-maker inputs, for which new analysis and consensus-building algorithms have to be developed.

CMAA introduces an innovative, entropy-based output consensus metric that measures both the current degree of consensus and also the next-step consensus that would result from each available conflict resolution. This makes a greedy heuristic possible that determines the clarification task with the greatest potential for improving consensus. The new metric has the unique property of being able to detect that there is no unshared information left in the group that could change the algorithm's recommendation.

Simulation experiments validate CMAA's ability to achieve consensus more efficiently and more completely than established input and output metrics. This turns out to be true even for decisions that contain compromise resolutions. In many cases, a vast proportion of the combinatorial space can be ignored during consensus-building. In these cases, the path to consensus only needs to visit a fraction of the evaluation conflicts contained in the initial data supplied by the decision-makers.

The combinatorial state space that is generated grows exponentially with the size of the decision and quickly becomes intractable. However, Monte Carlo simulation can be used to perform the analysis to a sufficient degree of accuracy in less than a second on a notebook computer, enabling its application in real-time decision-making scenarios.

Two case studies were performed. Group decisions in a biotechnology startup and in a software company were carried out with CMAA acting as a digital assistant for a human facilitator. All assumptions underlying the method could be validated, including the trustworthiness of the method in the eyes of the decision-makers and their satisfaction with the results.

Based on its new combinatorial model, CMAA provides improved algorithms for the three fundamental components of multi-criteria group decision-making:

- the analysis variables,
- the consensus metric, and
- the consensus-building process.

The new analysis variables offer an unmatched level of detail, revealing the performance dependencies for each alternative on individual input evaluations while considering all possible configurations of inputs. The new consensus metric establishes a higher standard for consensus: not only do the decision-makers agree on the most-preferred alternative recommended by the algorithm, but this recommendation cannot be affected by any additional conflict resolutions. The consensus-building process is based on knowledge of the attainable next-step consensus degree for each possible decision by the decision-makers. Together, the three new fundamental components enable efficient consensus-building processes for cooperative decisions. The framework works together with any decision method, only requiring adaptation to the specific data format.

Contents

List of Key Concepts

Competitive decision: A competitive decision is one in which decision-makers have different, and possibly conflicting, goals or motivations. Evaluation conflicts are a result of these differences.

Cooperative decision: A cooperative decision is one in which decision-makers aspire to a common goal, and evaluation conflicts are a result of differences in knowledge or experience.

Revelation-invariance: A group's decision is revelation-invariant, when no decision-maker holds information that would change the result if it became known to the others.

Correct decision: A correct decision is one that effectively integrates and pools all decisive information from each decision-maker, resulting in a collective choice that is deemed correct for the group as a whole.

Revelation-invariant consensus-building: The new consensus-building process finds a path to a revelation-invariant consensus, ensuring that the group's decision remains unchanged, regardless of any unshared information within the group.

Efficient consensus-building: The new consensus-building process only needs to resolve a fraction of the conflicts to reach consensus.

New consensus metric: Develop a novel consensus metric capable of quantifying the proximity of the current decision state to a revelation-invariant consensus.

Impact of conflicts: Develop a method that can distinguish conflicts with larger consensus impact from those with smaller impact.

Preserve evaluations: All conflicting evaluations must be preserved, as each may hold valuable information for building consensus.

Premature consensus: Premature consensus occurs when the consensus threshold is reached, yet resolution of further discrepancies may still alter the preferredness for one or more alternatives.

Instance sampling: Using $K_{MC} = 10,000$ in a Monte Carlo simulation is recommended for randomly sampling the combinatorial space.

Entropy threshold: If a soft, but very firm consensus is sufficient, using $\tau = 0.3$ as a stopping criterion for consensus-building is recommended.

Correct decision with CAE: Decision-maker groups that value a revelation-invariant consensus, should use the CAE consensus metric.

Fast convergence: If fast consensus-building is required, a non-compensatory decision model should be used.

Invest in agreement resolutions: The larger the number of agreements in clarification conferences, the faster the consensus-building process converges. Therefore, it is worthwhile trying to reach agreement.

Confidence in the result: When a hard consensus has been reached, no further resolution would change the consensus result. This might be hard to understand for the decision-makers and should be explained to them, if needed.

List of Symbols

General group decision

a_i	alternative with index i
c_j	criterion with index j
m	number of alternatives in a decision
n	number of criteria in a decision
(c_j)	preference task for criterion j
(c_j, a_i)	judgement task for a criterion/alternative pair
DM_k	decision-maker with index k
d	number of decision-makers in a group decision
$\mu_k(j)$	individual preference for criterion j
$\lambda_k(j, i)$	individual judgement for criterion/alternative pair c_j, a_i
$\mu(j)$	set of preferences for (c_j)
$\lambda(j, i)$	set of judgements for (c_j, a_i)
CM_I	input metric on the basis of input (dis)similarities
CM_O	output metric on the basis of output (dis)similarities

CMAA-specific

$\mathbf{P}$	set of all unified preferences
$\mathbf{A}$	set of all unified judgements
$[\mathbf{P}; \mathbf{A}]$	group decision
ϕ_j	valency for the preferences $\mu(j)$
ϕ_{ji}	valency for the judgements $\lambda(j, i)$
K	number of instances of a combinatorial space
K_{MC}	number of Monte Carlo samples
$[\mathbf{P}^*; \mathbf{A}^*]$	an instance of the group decision
$[\mathbf{P}_j(k); \mathbf{A}]$	a partial decision, where $\mu(j) = \mu_k(j)$
$[\mathbf{P}; \mathbf{A}_{ji}(k)]$	a partial decision, where $\lambda(j, i) = \lambda_k(j, i)$
b_i^r	rank acceptability for the alternative a_i and the rank r
acc_i	holistic acceptability
α_r	meta-weight for rank r for holistic acceptability
$s_i(\mu_k(j))$	current preference acceptability for $\mu_k(j)$
$\hat{s}_i(\mu_k(j))$	potential preference acceptability for $\mu_k(j)$
$\sigma_i(j)$	preference sensitivity for $\mu(j)$

$q_{i1}(\lambda_k(j, i2))$	current judgement acceptability for $\lambda_k(j, i2)$
$\hat{q}_{i1}(\lambda_k(j, i2))$	potential judgement acceptability for $\lambda_k(j, i2)$
$\sigma_{i1}(j, i2)$	judgement sensitivity for $\lambda(j, i2)$
h	entropy of the current decision
$\hat{h}(\mu_k(j))$	potential preference entropy for the preference $\mu_k(j)$
$\hat{h}(\lambda_k(j, i))$	potential judgement entropy for the judgement $\lambda_k(j, i)$
$[\mathbf{P}; \mathbf{A}]_s$	decision at consensus step s
h_{stop}	consensus threshold
τ	parameter to set consensus threshold h_{stop}
$P(EO)$	probability for choosing the entropy-optimal resolution
ρ	agreement/compromise ratio for resolutions

Other MCDM decision methods

$F(a_i)$	performance of an alternative in SAW
u_l	pairwise preference or judgement in minimal AHP
(u_l)	preference task for a pairwise comparison in minimal AHP
(c_j, u_l)	judgement task for a pairwise comparison in minimal AHP
$\mu_k(l)$	preference for pairwise comparison in minimal AHP
$\lambda_k(l)$	judgement for pairwise comparison in minimal AHP
g_z	satisfaction level
$M_{n\times m}$	objective measurements for all judgement tasks
$G_{n\times m}$	subjective satisfaction levels for all judgement tasks
$Z + 1$	number of satisfaction levels
$t_{j,z}$	subjective threshold judgement

1
Introduction

This thesis introduces Combinatorial Multicriteria Acceptability Analysis (CMAA), a new approach to facilitating consensus in cooperative multi-criteria group decision-making processes. CMAA replaces standard compromise-driven consensus-building by an information asymmetry approach that enables the resolution of conflicts to unanimous evaluations. As later chapters will show, this paradigm shift enables improvements in decision analysis, consensus metrics and consensus processes.

In this chapter, the current state of multi-criteria group decision-making is discussed, highlighting its shortcomings with respect to the goals of the thesis. The new approach integrates insights from two research areas: (1) Multi-Criteria Group Decision-Making, and (2) Shared Mental Models. It outlines how the insights from these research areas can be used to create a new, efficient approach to consensus-building.

Finally, the goals, requirements, research questions and contributions of this work are presented.

1.1 Multi-criteria group decisions

This thesis uses multi-criteria decision-making (MCDM) to help a group select the best alternative. In practice, MCDM methods are favoured because they reduce the cognitive effort required for evaluating alternatives (Olson et al., 1998). A multi-criteria model breaks down a complex decision into smaller, more manageable decisions (Cinelli et al., 2020). The overall decision is then composed of these partial decisions based on the specific decision model used. The resulting ranking reflects the *preferredness* of each alternative, with the *most-preferred* alternative being the one that attains the first rank. Decisions conducted with MCDM methods are:

- *easier to evaluate*, because smaller decisions require less cognitive effort,
- *objective*, because evaluations for all criteria are less susceptible to subjective errors, and
- *trustworthy*, because the resulting recommendation can be justified.

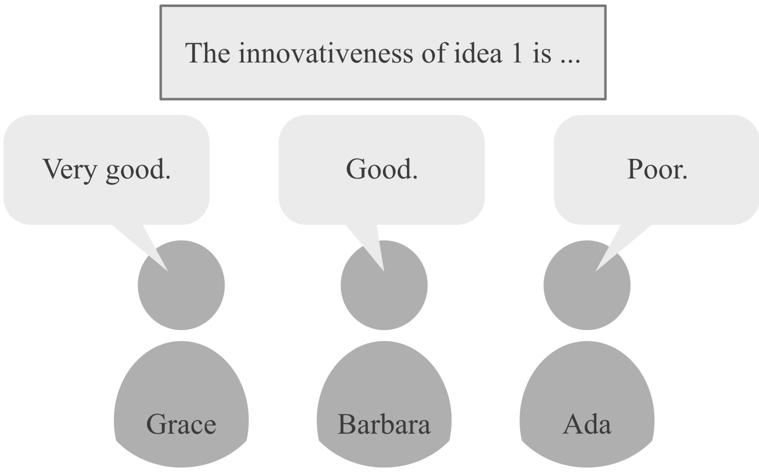


Figure 1.1: Evaluation conflict between three decision-makers

From the mid-2000s, interest in multi-criteria group decision-making began to increase. This rise was due to the growing complexity of decisions. For these groups, it is crucial to pool expertise, consider various perspectives on knowns and unknowns, and agree on the best alternative to pursue collectively.

Examples of group decisions include determining a strategy for environmental security for a region (Tervonen et al., 2007), selecting environmentally friendly options for aquaculture (Ozer, 2007), choosing a new supplier (Mi et al., 2006), deciding which talent to promote (Lai and Ishizaka, 2020), selecting the best cancer therapy for a patient (Specchia et al., 2020), or deciding which product to develop next (Youssef and Webster, 2022).

Multi-criteria decision methods are designed for single-user inputs. To accommodate groups, differing inputs are aggregated to single values. This can have detrimental effects, such as devaluing minority opinions (Tran et al., 2023). Encouraging minority dissent, as suggested by Nijstad et al. (2014),fosters divergent thinking and enables groups to process information at a broader level. Sharing perspectives from minority voices can ultimately enhance decision-making quality (De Dreu and West, 2001; Nijstad et al., 2014).

1.2 Competitive and cooperative decisions

In groups, evaluation conflicts occur when two or more evaluations of the same issue differ. Figure 1.1 illustrates an example of an evaluation conflict, where three decision-makers Grace, Barbara, and Ada evaluate the *innovativeness* of *Idea 1* differently, leading to an evaluation conflict. The more diverse a group is in terms of expertise and knowledge, the higher the likelihood of evaluation conflicts arising. For this thesis, it is important to distinguish between two types of group decision, as they have different causes and interpretations.

1.2.1 Competitive decisions

The first type of group decision will be called a *competitive decision*. These are defined as follows:

Definition 1. *(Competitive decision) A competitive decision is one in which decision-makers have different, and possibly conflicting, goals or motivations. Evaluation conflicts are a result* *of these differences.*

Competitive decisions occur when decision-makers represent different stakeholders. They have conflicting interests, which might be socially, economically or politically driven. For example, in a city council meeting, a decision has to be made about where to locate a new waste management facility. One perspective for a particular location might be the negative effect on the quality of life for local residents. At the same time, the location might be easier and cheaper to access from a logistical standpoint. Elected council members who represent different stakeholders will evaluate the location differently.

Conflicts in competitive scenarios such as these can only be overcome by negotiation and compromise. The parties move towards each other reluctantly and eventually meet 'in the middle'. For these groups, a decision support system should help to find the best compromise.

Perfect unanimity of all evaluations will be referred as *input consensus*. In a scenario such as this, input consensus is considered to be impossible to achieve. Most consensus-building approaches in multi-criteria group decision-making follow this assumption. Since perfect input consensus is not possible, Kacprzyk and Fedrizzi (1988) proposed that the next-best result is a partial, or *soft* input consensus. Different approaches for reaching consensus will be explained in the next chapter.

1.2.2 Cooperative decisions

The second type of group decision will be called a *cooperative decision*:

Definition 2. *(Cooperative decision) A cooperative decision is one in which decision-makers aspire to a common goal, and evaluation conflicts are a result of differences in knowledge or experience.*

Cooperative decisions are the opposite of competitive decisions. Decision-makers are not driven by conflicting goals or political agendas. Instead, they are driven by a common goal and the desire to find the best possible decision to achieve it. Different group members are involved, because they can contribute various perspectives, ensuring that the best possible decision is reached. This means that all group members' evaluations are equally valuable.

An example of a cooperative decision is a tumor board in a hospital, where physicians decide the best course of action for treating a cancer patient (Specchia et al., 2020). A medical oncologist can estimate the chances of success of chemotherapy, a surgical oncologist might suggest a surgical intervention to remove the tumor, and a radiation oncologist can assess the risks and potential benefits of radiotherapy.

In cooperative decisions, conflicts between decision-makers exist because of information asymmetry (Schulz-Hardt et al., 2006). Information asymmetry describes a situation where one decision-maker holds information that is unknown to the others. Evaluation conflicts disappear when the decision-makers share the information and expertise that led to their evaluations, thereby removing the information asymmetry.

1.2.3 Resolving conflicts by compromise or by agreement

Table 1.1 shows the situation for the three decision-makers Grace, Barbara, and Ada regarding the *innovativeness* of *Idea 1* that is depicted in Figure 1.1. The table also shows two possible resolutions: one for competitive decisions (in the column compromise), and one for cooperative decisions (in the column agreement). The competitive resolution is the

Table 1.1: Possible conflict resolutions for both group decision types

decision-makers	initial	compromise	agreement
Grace	very good	fair	poor
Barbara	good	fair	poor
Ada	poor	fair	poor

value ‘fair’, which is a compromise between the extremes of ‘poor’ and ‘very good’. The cooperative resolution might be ‘poor’, because Ada reveals decisive information to Grace and Barbara that was unknown to them, prompting their re-evaluation.

When a cooperative decision process uses an approach that drives consensus-building solely by encouraging compromise, it creates dissatisfaction (Tran et al., 2023). Instead of sharing information and gaining a joint perspective, decision-makers are pushed towards an evaluation that none of them had proposed, leaving the impression that no one’s perspective has been respected.

In contrast, when cooperative groups share the information that led to their initial evaluations, decision-makers (Grace and Barbara) learn new information (from Ada). This pooling of information then leads to full agreement. In a consensus-building process, agreement is highly desirable, as it fosters commitment and increases the likelihood of a successful outcome (Moscovici and Doise, 1994; Briggs et al., 2005), as highlighted by Susskind et al. (1999):

> *When decision-makers’ preferences and concerns are considered, they are much more likely to actively participate in the implementation of the obtained solution.*

Keeney (2009) supports this view:

> *The collaborative decision analysis process allows each group member to incorporate his or her knowledge, information, and judgement into the model. This makes explicit any differences about knowledge, information, and judgements held by the group members. Once clarified, many of these differences may be eliminated in productive discussions.*

In cooperative decisions, all group members’ evaluations hold value and are essential because each evaluation may reflect decisive information. Therefore, all decision-maker evaluations must be treated equally, because any of them might alter the group decision result. Cooperative decisions require that no resolution should change the decision result, leading to the concept of revelation-invariant consensus:

Definition 3. *(Revelation-invariance) A group’s decision is revelation-invariant, when no decision-maker holds information that would change the result if it became known to the others.*

Most multi-criteria group decision-making and consensus-building research focuses on competitive decisions. Their consensus-building approaches aim to minimise the effort needed to reach a compromise and are not designed to create revelation-invariant consensus. Thus, these approaches do not meet the needs of cooperative decisions. Therefore, the primary objective of this thesis is to develop algorithms for multi-criteria group decision-making and consensus-building that can provide a revelation-invariant consensus.

1.3 Motivation for a new consensus-building framework

Cooperative decisions require pooling of decisive information regarding conflicting evaluations to achieve a revelation-invariant consensus. However, current approaches do not meet the requirements of cooperative decisions. The arguments are briefly outlined in this section. Previous research on these topics will be reviewed in Chapter 2.

1.3.1 Correct group decisions

Brodbeck et al. (2007) point out that although groups are expected to make better decisions thanks to their collective knowledge, this expectation might not always prove to be true:

> *[...] more and more groups of employees make decisions under conditions of distributed knowledge, but the common expectation that they effectively use their superior knowledge base when necessary seems empirically not justified.*

In this context, a *correct* decision is defined as the one that the group agrees on when all relevant information has been shared (Schulz-Hardt and Mojzisch, 2012):

Definition 4. *(Correct decision) A correct decision is one that effectively integrates and pools all decisive information from each decision-maker, resulting in a collective choice that is deemed correct for the group as a whole.*

Research in Social Psychology suggests that groups can reach such a correct decision if they successfully share information (Stasser and Titus, 1985; Bittner and Leimeister, 2014). However, studies have found that groups are often unsuccessful at pooling decisive information and fail to outperform individual decision-makers (Nijstad et al., 2014; Tindale and Winget, 2019).

Studies observing this phenomenon often conduct group decisions by discussing alternatives as a whole. These will be referred to as *holistic* decisions. However, holistic group decisions have a tendency to miss important information, because the discussions are unstructured, and they are often outperformed by individual decision-makers (Lu et al., 2011).

Disadvantage 1. *Holistic group decisions tend to overlook important information and are often outperformed by individual decision-makers.*

It is worth considering whether groups are more likely to identify important information when applying a multi-criteria decision model. Group members may be more likely to engage in discussion about the performance of an alternative with respect to a specific criterion, because the scope of the discussion is much narrower and takes less cognitive effort. Therefore, if the new approach is able to provide a revelation-invariant consensus, it may indeed reflect a correct decision.

1.3.2 Aggregation of conflicting evaluations

Often in practice, consensus is not actively pursued. These methods, referred to as *single-step* methods in this thesis, do not facilitate information exchange or conflict resolution.

Instead, conflicting evaluations are combined into average values, as if they had originated from a single decision-maker. Tran et al. (2023) note that this averaging can be problematic for groups with polarised evaluations, as it may force group members to compromise on their preferences, leading to an outcome that no one truly favours. In cooperative decisions, reducing evaluation conflicts to average values can be detrimental. This aggregation obscures the unique information underlying each evaluation, potentially resulting in a loss of valuable insights.

For instance, consider Table 1.1, where a single-step multi-criteria group decision-making approach might simplify the conflict to 'fair', which is an incorrect evaluation of *Idea 1*. However, had the decision-makers shared their information, they would have learned of Ada's knowledge of an industry development that undermines the *innovativeness* of *Idea 1* and collectively (and correctly) revised their evaluation to 'poor'. Hence, averaging conflicting evaluations removes the opportunity to uncover important information known to only a subset of decision-makers.

Disadvantage 2. *Fusing conflicting evaluations to a single value eliminates the possibility of discovering important information.*

1.3.3 Biased consensus-building

Consensus-building represents an improvement over single-step decision-making by aiming to enhance agreement among decision-makers. This iterative process enables evaluations to be refined across multiple iterations, progressively enhancing a consensus metric. The underlying premise of most approaches is the resolution of conflicts through compromise. Decision-makers are either urged or forced to align their evaluations with the group average.

Disadvantage 3. *Any automatic, forced or suggested adjustment regarding an evaluation conflict introduces a bias into the group decision.*

1.3.4 Consensus metrics

A consensus metric measures the level of agreement among decision-makers. There are two main types of metric: input-based and output-based. The first and most commonly used type are input-based consensus metrics. These measure the (dis)similarities between input evaluations. Input metrics are usually used to drive a consensus-building approach, because (dis)similarities between inputs are easy to compute, and the reduction of the largest dissimilarity to a single value would improve an input consensus metric the most.

Output metrics, on the other hand, measure the differences in individual decision results and are therefore a more accurate model of consensus. However, no existing output metric can detect a revelation-invariant consensus (Definition 3 on page 4).

Both types of metric yield conflicting feedback when applied to the same decision, and achieving one type of consensus does not guarantee achieving the other. In cooperative decision-making, both input and output metrics face challenges:

Disadvantage 4. *Input metrics do not reflect agreement upon the decision outcome.*

Disadvantage 5. *Existing output metrics cannot drive a consensus-building process or recognise a revelation-invariant consensus.*

Developing a new metric that combines the strengths of both input and output metrics, while ensuring revelation-invariant consensus, would mark a significant advancement in consensus-building methodologies.

1.4 Goals for consensus-building in cooperative decisions

The previous section highlighted several drawbacks of current approaches for cooperative decisions. In response to these challenges, this section outlines the specific attributes that cooperative decision-making requires and formulates the goals of the thesis based on these requirements.

1.4.1 Building and recognising consensus in cooperative decisions

In cooperative decisions, achieving consensus is crucial. One approach for a cooperative group to reach a revelation-invariant consensus is by thoroughly discussing and resolving all conflicts. If all conflicts were resolved in this brute-force approach, it would lead to a revelation-invariant consensus, where the group's decision remains unchanged even in the face of new arguments or information.

However, this brute-force approach is impractical, requiring excessive time and cognitive effort from decision-makers. A more efficient approach is needed to achieve a revelation-invariant consensus.

Goal 1. *(Revelation-invariant consensus-building) The new consensus-building process finds a path to a revelation-invariant consensus, ensuring that the group's decision remains unchanged, regardless of any unshared information within the group.*

Goal 2. *(Efficient consensus-building) The new consensus-building process only needs to resolve a fraction of the conflicts to reach consensus.*

To achieve these goals, a new type of output consensus metric is necessary. It should detect a revelation-invariant consensus and measure the distance between the current decision state and this ideal consensus.

At first glance, current output consensus metrics appeared to hold promise for cooperative groups. However, upon closer examination, they exhibit two significant drawbacks. First, they struggle to pinpoint conflicting evaluations that could enhance consensus. Second, their hard consensus determination can be premature and may not align with a revelation-invariant outcome. Therefore, the development of a new consensus metric that accurately measures the distance to achieving a revelation-invariant consensus is crucial. This will be further explored in the upcoming section.

Goal 3. *(New consensus metric) Develop a novel consensus metric capable of quantifying the proximity of the current decision state to a revelation-invariant consensus.*

1.4.2 Impact of conflicting evaluations

To efficiently guide a group towards consensus, it is important to understand the impact of each conflict on the consensus-building process. The new analysis should be able to distinguish conflicts with larger and smaller impacts:

Goal 4. *(Impact of conflicts) Develop a method that can distinguish conflicts with larger consensus impact from those with smaller impact.*

Conflicts are valuable indicators of unshared information in groups and potentially contain consensus-relevant asymmetries. However, it is unknown *a priori* which resolution will be chosen. Therefore, it is essential to consider the contributions of all differing evaluations in a conflict to the consensus-building process. Each evaluation must be preserved in the computational model to accurately measure its impact on consensus. This preservation of evaluations is not currently followed by any multi-criteria group decision-making approaches, making it a crucial requirement for the new approach.

Requirement 1. *(Preserve evaluations) All conflicting evaluations must be preserved, as each may hold valuable information for building consensus.*

1.4.3 Unbiased resolution of conflicting evaluations

In current consensus-building algorithms, it is common to suggest adjustment directions for resolving conflicts to prevent stalemates and motivate decision-makers to move their evaluations for creating input-based consensus. However, in cooperative decisions, this influence can bias the group and inhibit open discussion about the conflict at hand. Therefore, the new framework should ensure:

Requirement 2. *(Unbiased resolutions) The algorithm should not bias the group by suggesting a direction for resolving a conflict.*

The new approach should be able to compute the impact each conflict would have on consensus, as outlined in Goal 4 on page 7. Then, it can suggest the conflict with the highest consensus-relevant impact, allowing decision-makers to share information and resolve both the information asymmetry and the conflict. This process facilitates unbiased resolutions and fosters open discussion, ultimately leading to consensus.

1.4.4 Method-independent framework

There are many MCDM methods available, each with its own set of features and suitability for different decision-making needs. Consequently, the new approach should offer a framework that decouples the decision analysis and consensus-building process from the specific MCDM method. By providing this flexibility, the framework enables the application of different methods while adhering to their respective rules and requirements. This allows a decision-making group to choose the most appropriate MCDM method based on their application and preferences.

Goal 5. *(Method-independent framework) Create a framework for analysing a group decision and building consensus that is independent of the multi-criteria decision-making method to be used.*

1.4.5 Digital facilitation and assistance for consensus-building

Ideally, the new approach would reduce the interaction to a single face-to-face meeting, which is used to resolve important conflicts and reach consensus. Apart from that, it should be possible to perform all other interactions independently and without location restrictions. Thus, the interactions during and after the process should be clarified and any challenges or rules should be illustrated.

Apart from pursuing the previously introduced goals, the new approach should allow for a transparent process, which the decision-makers can understand and an audit trail that allows the consensus path to be reconstructed.

To be of practical use, the algorithms must be implementable as a digital assistant to a human facilitator or even as a stand-alone digital facilitation tool.

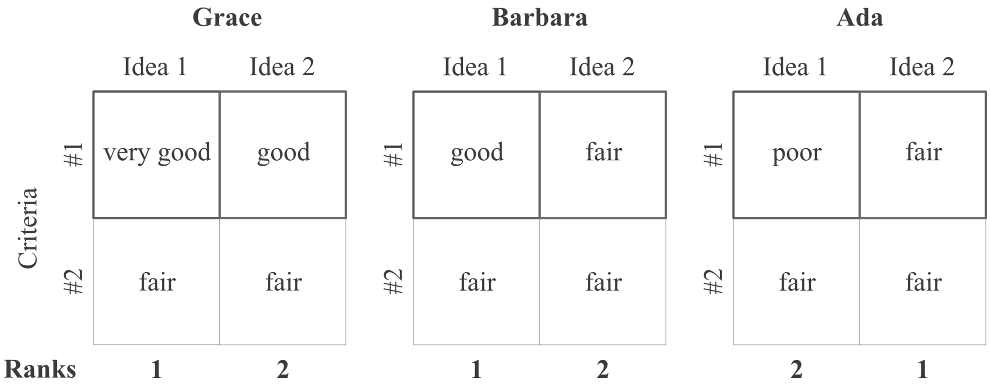


Figure 1.2: Individual evaluations for a decision with two ideas

Goal 6. *(Automation) Provide ideas for a digital facilitation process assisting a human facilitator and an automatic facilitation of clarification conferences.*

1.5 Outline of the new approach

This thesis aims to provide an efficient consensus-building process that caters to the needs of cooperative decisions (Definition 2 on page 3). The recommendation produced by the consensus-building process should be revelation-invariant (Definition 3 on page 4). This requires that decisive information gets shared during consensus-building. Decisive information is present whenever two or more decision-makers provide conflicting evaluations, indicating an inter-decision-maker information asymmetry. To uncover these, all decision-maker inputs need to be treated as equally important.

The new approach treats conflicts as indicators of potentially consensus-relevant information asymmetry. When the group shares all important information, conflicting evaluations are resolved by agreement. The new approach needs to preserve all conflicting evaluations and measure their impact on the group's decision outcome. This is achieved by combining all conflicting evaluations into a combinatorial space. The combinatorial space unifies all possible agreements a group could potentially achieve during a consensus-building process. Hence, the term 'combinatorial' in the name Combinatorial Multicriteria Acceptability Analysis.

This combinatorial space can then be analysed. For each possible combination of evaluations a group could potentially agree on, a MCDM decision method can be applied to compute a ranking of the alternatives. When this has been done for all combinations, the analysis can derive which alternative performed best for the whole combinatorial space or which alternative's performance strongly depended on specific conflict resolutions. Using a new consensus metric, the closeness to a revelation-invariant consensus can be derived. The analysis of the combinatorial space can also be used to determine which conflict resolution has the potential to progress consensus the most.

Because the analysis of the combinatorial space can distinguish conflicts by their consensus relevance, it can be used in a consensus-building algorithm to guide a group efficiently to consensus. The general idea of the combinatorial space is that only some conflicts need resolving to achieve a revelation-invariant consensus.

Figure 1.2 shows a decision consisting of two *criteria #1* and *#2*, two alternatives *Idea 1* and *Idea 2* and the three decision-makers Barbara, Grace, and Ada. The situation shown in

Figure 1.1 forms part of this decision. Using a simple MCDM method, each decision-maker generates the individual rankings shown. Since there exist three different judgements for *Idea 1* with respect to *Criterion #1* and two different judgements for *Idea 2* with respect to *Criterion #1*, there is a total of six different judgement combinations overall. If the group were convinced by Grace's or Barbara's argument at *Criterion #1* ('very good' or 'good'), *Idea 1* would be most-preferred, regardless of the fact that there remain two different, unresolved judgements concerning *Idea 2*. Likewise, if Ada were to convince the other two decision-makers (to 'poor'), *Idea 2* would be most-preferred, even though three unresolved evaluations for *Idea 1* remain.

As the number of conflicts increases, the complexity grows rapidly. Nevertheless, the fundamental concept remains valid and adaptable to any decision size.

This very simple example outlined the general idea of CMAA. The ranks for all combinations are used to compute a consensus measure. Thus, even when conflicting evaluations remain, as long as the first rank is the same for all combinations, the group would have achieved a revelation-invariant consensus.

1.6 Hypotheses and research questions

After outlining the general concept for the new approach to building consensus for cooperative groups, this section formulates two hypotheses and presents some key research questions. The hypotheses will be tested in Chapter 7 and Chapter 9, and the research questions will be addressed in their respective sections.

1.6.1 Hypotheses

The previous section introduced the combinatorial space as a key concept for analysing the conflicting evaluations and deriving a path towards consensus. The example provided in Figure 1.2 was very simple, containing only two sets of conflicting evaluations. In practice, group decisions contain many more conflicts, resulting in extremely large combinatorial spaces. However, many of these combinations may result in identical choices or rankings.

Definition 5. *(Slack) A group decision contains 'Slack', when many combinations of inputs result in the same most-preferred alternative or ranking.*

If the consensus-building process is designed efficiently, Slack means that only a fraction of the conflicts must be resolved to reach a revelation-invariant consensus. This leads to the following hypothesis:

Hypothesis 1. *(Slack in group decisions) The majority of conflicting evaluations in groups do not need to be resolved in order to reach consensus.*

The combinatorial space can become so large that analysis becomes intractable. Thus, the question arises: will a small set of samples from this combinatorial space provide sufficient accuracy?

Hypothesis 2. *(Sampling accuracy) Only a small fraction of the combinatorial space needs to be sampled to provide sufficient accuracy for the decision analysis.*

1.6.2 Research questions

Following the outline for a new group decision analysis and consensus-building approach, the following research questions must be answered to validate its applicability.

Achieving a hard, revelation-invariant consensus can be accomplished through a brute-force approach. Other consensus-building approaches based on input metrics are more efficient, but unsuitable for cooperative decisions. It is therefore important to determine the consensus-building performance of CMAA.

Question 1. *(Efficient consensus-building) How efficient is consensus-building with CMAA compared to a brute-force approach or to comparable input metric-based consensus-building approaches?*

While CMAA is designed for cooperative decisions, real-life decisions can be wholly or partially competitive. Therefore, it is important to investigate whether the algorithm can still efficiently reach consensus when some or all resolutions are compromises.

Question 2. *(Competitive decisions) Will consensus-building with CMAA still be effective with compromise resolutions?*

The CMAA approach generates a large amount of combinatorial information, and it is to be hoped that this additional information can provide an additional level of usefulness with its decision analysis.

Question 3. *(Detailed analysis) What new insights into the decision structure will the combinatorial analysis provide?*

In principle, CMAA can be combined with any MCDM method. However, adaptations may be needed for certain methods.

Question 4. *(Compatibility) How complex is the adaptation of different decision methods to the CMAA framework?*

CMAA can achieve a revelation-invariant consensus without resolving every evaluation conflict. However, unresolved evaluation conflicts may reduce the decision-makers' trust in the algorithm, as they might feel that the result could change, if additional conflicts were resolved.

Question 5. *(Acceptance) Will a group of decision-makers accept the recommendation by CMAA, despite the unresolved evaluation conflicts?*

1.7 Contributions

This thesis introduces a novel framework for group decision analysis and consensus-building known as Combinatorial Multicriteria Acceptability Analysis (CMAA). Its approach to handling conflicting evaluations is unique within the field of multi-criteria decision analysis, offering new research opportunities in group decision-making and consensus-building. These opportunities include developing more efficient consensus-building methods, applying CMAA to competitive decisions, and enhancing the efficiency and reliability of group decision analysis.

One of CMAA's key contributions is its novel consensus metric, which integrates the benefits of both input and output metrics. This metric meets the requirement for a revelation-invariant consensus in cooperative decisions, representing the highest standard in consensus metrics. Additionally, it can identify conflicts with significant impacts on consensus, thus facilitating a more efficient consensus-building process.

The innovative treatment of conflicting evaluations by constructing a combinatorial space allows for the derivation of new analysis variables. These variables enable a comprehensive group decision analysis, providing an unparalleled understanding of the decision and its interdependencies.

The choice of decision model is critical in decision-making, as it depends on the specific application and the capabilities of the decision-maker. Popular models have been adapted to suit various applications, extending their use to group decision-making contexts. CMAA enables the use of any decision model in consensus-building.

CMAA incorporates insights from shared mental model research into multi-criteria group decision-making. This integration opens new avenues for studies in group decision-making in Social Psychology, aiding in the identification of unshared information through multi-criteria decision analysis.

The CMAA framework is examined in detail in Chapters 7, 8, and 9. These chapters provide practical insights into using the CMAA framework, integrating various decision methods with CMAA, understanding the number of iterations required for consensus-building, and exploring ways to improve efficiency. Two case studies validate the acceptance of CMAA's consensus-building process.

1.8 Structure of the thesis

The thesis is divided into ten chapters:

- In Chapter 1, the research topic was introduced and motivated. Then Chapter 2 will present a comprehensive literature review, detailing the research gap that this thesis aims to address.
- Chapter 3 will introduce the CMAA analysis algorithm, including a group decision model, an analysis algorithm and new group decision analysis variables. An illustrative example will demonstrate the new framework.
- In Chapter 4, the new consensus metric will be described, illustrating its effects using the continued example from Chapter 3.
- Chapter 5 will present the CMAA consensus-building model and the algorithm based on a greedy heuristic.
- Chapter 6 will outline some ideas for a digital facilitation using CMAA, including an example of an automated facilitation script and other mock-ups.
- Chapter 7 will study the performance of the new framework for differently sized problems and types of conflict resolution. From these studies, considerations for the practical application of the method will be drawn.
- Chapter 8 will describe the application of CMAA to eight case studies taken from the MCGDM literature. These studies will showcase the framework's effectiveness.
- Chapter 9 will contain two new case studies in which CMAA consensus-building was applied to real-life group decisions.
- Chapter 10 concludes the dissertation with a summary, a discussion of CMAA and its experimental results and an outlook for future research.

2
Literature Review and Research Gap

This chapter discusses the pertinent scientific literature for this thesis, covering MCDM methods, group decision analysis, group decision-making, and consensus-building. Additionally, it examines related work for its relevance to cooperative decisions, incorporating studies from group decision-making in Social Psychology. From this review, the research gap that the thesis aims to address is identified. This chapter builds partially on published work (Horton and Goers, 2021; Goers and Horton, 2023a; Goers et al., 2024).

2.1 Notation

This section introduces terms that will be used throughout the thesis.

In multi-criteria decision-making, two types of results are considered: selecting the most-preferred alternative or set of alternatives, or ranking the entire set of alternatives. Decision methods compute a preferredness or performance score for each alternative, from which a ranking can be derived. This thesis focuses on selecting the most-preferred alternative.

A group decision consists of d decision-makers, denoted by DM_k, m alternatives a_i, and n criteria c_j. Each decision-maker evaluates the importance of each criterion as well as the performance of each alternative concerning each criterion, also known as a criterion/alternative pair. Throughout this dissertation, an evaluation of criterion importance will be referred to as a *preference*, and the evaluation of an alternative's performance will be referred to as a *judgement*. The data type of each varies according to the chosen decision method.

There exist two types of evaluation task: a preference task (c_j) and a judgement task (c_j, a_i). For most methods, each decision-maker is required to submit evaluations for n preference tasks and $n \times m$ judgement tasks. There are exceptions, which will be explicitly noted.

Preferences are denoted by $\mu_k(j)$, and judgements are denoted by $\lambda_k(j, i)$, where the index k is local to each judgement or preference task.

Conflicts between decision-makers' evaluations are termed *discrepancies*. There are two types of discrepancy:

(i) A *preference discrepancy* occurs when two or more decision-makers submit different preferences $\mu_{k1}(j) \neq \mu_{k2}(j)$ for the same preference task (c_j).

(ii) A *judgement discrepancy* occurs when two or more decision-makers submit different judgements $\lambda_{k1}(j,i) \neq \lambda_{k2}(j,i)$ for the same judgement task (c_j, a_i).

A preference discrepancy is *resolved* when all preference differences are eliminated: $\forall k1, k2 : \mu_{k1}(j) = \mu_{k2}(j)$. Similarly, a judgement discrepancy is resolved when all judgement differences are removed: $\forall k1, k2 : \lambda_{k1}(j,i) = \lambda_{k2}(j,i)$. Each resolution eliminates one discrepancy from the decision-making process.

2.2 Multi-criteria decision-making methods

This Section examines various characteristics of multi-criteria decision models and methods. It outlines two distinct types of models (compensatory and non-compensatory) and their input variants. Subsequently, three multi-criteria decision methods are elaborated upon in detail, as they will be employed in Chapters 7 and 8. Other methods used in a comparison study in Section 7.4.2 are briefly discussed. The details regarding these methods are primarily sourced from Cinelli et al. (2020).

2.2.1 Classification of models and methods

When developing and comparing multi-criteria decision models and methods, various attributes need to be taken into consideration.

Types of preferences and judgements

Judgements in MCDM methods can be objective or subjective, while preferences are always subjective. Some studies, such as Kangas et al. (2006), have used a combination of objective and subjective judgements.

At times, even when objective measurements are accessible, subjective judgements may be used (Banaeian et al., 2018; Thakkar, 2021). The studies covered in Section 8.1 embody this choice. It arises when specific objective measurements are less suitable than the value they provide (Noori et al., 2018). For instance, when two job opportunities offer annual incomes of \$95,000 and \$100,000 respectively, it might be more appropriate for the decision-maker to subjectively assess them both as ‘very attractive’ rather than using their numerical values.

Objective measurements can be either beneficial or non-beneficial (S. Dhiman et al., 2020). In the case of a non-beneficial criterion, a lower numerical value indicates a higher subjective value to the decision-makers. A common example of a non-beneficial criterion is purchase price.

Subjective evaluations can take various forms. Scores can provide subjective evaluations within a certain range, such as integers in the range of $[1, 9]$. Most decision methods can process numerical evaluations directly. Fasolo and Bana e Costa (2014) discovered that decision-makers who are not well-versed in numerical representations find it easier to use linguistic variables. In such instances, a Likert scale (Likert, 1932) can be used. The linguistic responses can then be mapped to numerical values.

To incorporate uncertainty in subjective evaluations, fuzzy values can be used (Zadeh, 1975). An example of mapping linguistic values to a triangular fuzzy number (TFN)

Table 2.1: Triangular fuzzy number (TFN) evaluations

Preferences		Judgements	
Linguistic value	TFN	Linguistic value	TFN
very low (VL)	(0.0, 0.1, 0.2)	very poor (VP)	(0, 1, 2)
low (L)	(0.2, 0.3, 0.4)	poor (P)	(2, 3, 4)
medium (M)	(0.4, 0.5, 0.6)	fair (F)	(4, 5, 6)
high (H)	(0.6, 0.7, 0.8)	good (G)	(6, 7, 8)
very high (VH)	(0.8, 0.9, 1.0)	very good (VG)	(8, 9, 10)

representation is depicted in Table 2.1 (Banaeian et al., 2018). Linguistic variables for preferences and judgements are specified along with the corresponding triangular fuzzy number representations in the ranges [0,1] and [0,10], respectively.

Another input type for subjective evaluations are ordinal values. These can be ranking positions for preferences or ordinal equivalence classes for judgements. An equivalence class represents all judgements that are assigned the same subjective value, for example 'fully satisfies the criterion' (Yee et al., 2007).

Subjective evaluations can be relative rather than absolute. Pairwise comparisons conducted with linguistic variables, such as 'much more important than' for preferences or 'much better than' for judgements, serve this purpose. Pairwise comparisons, which are integral to the Analytical Hierarchy Process (AHP) introduced by Saaty (1972), are preferred, because they are more instinctively understood by decision-makers (Saaty, 2004).

In addressing uncertainty or ambiguity, evaluations can also be provided as probability density functions (Lahdelma et al., 1998). For instance, the unknown cost of a project might be represented by a uniform distribution ranging from \$100,000 to \$120,000.

To meet the requirements of Goal 5 on page 8, the new framework must accommodate all input formats. The implementation of CMAA presented in Chapter 3 will demonstrate this versatility.

Compensatory and non-compensatory decision models

A decision model can be classified as compensatory or non-compensatory. In a compensatory model, an alternative can offset poor performance in one criterion by delivering strong performance in another criterion. This rate of compensation is determined by the criteria weights.

In cases involving objective measurements, compensatory decision models face the challenge of incommensurability. When two criteria are measured using different units, a specific performance value in one criterion may disproportionately influence the decision compared to another. For instance, if one criterion represents the size of cloud storage space and another relates to the price of a cloud storage service, decision-makers must provide preferences that make the criteria comparable. This ensures that, for instance, 100 TB of cloud storage is considered equivalent in value to the cost of 50 cents per month. This complexity underscores the importance of employing subjective value interpretations of objective measurements, where decision-makers convert both storage capacities and prices into subjective value judgements like 'poor' or 'very good'.

However, compensation may not always be the preferred approach. Horton et al. (2016) argue that this is especially relevant in scenarios in which one goal is paramount, such as pursuing an innovative product idea to capture a new market segment. The leading selection criterion is derived from the innovation goal. It would be inappropriate to select an innovation project which is not the best in this leading criterion, even if it performs well with respect to other criteria. In such cases, non-compensatory models may be more suitable. These models do not permit any compensatory capacity (Thakkar, 2021). Criteria are ranked based on their importance on an ordinal scale, where the most important criterion is known as a 'dictator'. Alternatives that excel concerning the dictator criterion cannot be surpassed by other alternatives on the basis of their performance in other criteria.

To illustrate, a non-compensatory model might rank a job offering a $70,001 annual income with two days of vacation higher than a job offering a $70,000 annual income with three weeks of vacation, if the income criterion is the dictator (Evans, 2016). This scenario is clearly unrealistic, as three vacation weeks hold more value than a one-dollar salary difference between the two jobs. As a result, Horton and Goers (2021) advocate the use of equivalence classes with non-compensatory decision methods. Moreover, non-compensatory methods offer the advantage of requiring fewer evaluations to determine the top-ranked alternative compared to compensatory methods.

2.2.2 Selected decision models

Simple additive weighting (SAW)

Simple Additive Weighting (SAW) is a compensatory decision model (Kaliszewski and Podkopaev, 2016). This model is versatile and can accommodate both objective and subjective evaluations, including linguistic variables (Chou et al., 2008).

In the context of SAW, preferences are denoted as weights, serving to establish the compensation strategy among criteria. On the other hand, judgements represent the performance evaluations of an alternative concerning a criterion and are commonly termed scores or points. These judgements are typically integer values within the range of $[1, 9]$. The performance $F(a_i)$ of alternative a_i is the weighted sum of its judgements:

$$F(a_i) = \sum_{j=1}^{n} \sum_{i=1}^{m} \lambda_1(j, i) \cdot \mu_1(j) \ . \tag{2.1}$$

To calculate a decision recommendation using the SAW model, a total of $n+(n \cdot m)$ evaluations are required.

Table 2.2: An example decision matrix for an SAW decision

	weight	a_1	a_2	a_3	a_4	a_5	a_6
c_1	1	5	8	3	2	7	4
c_2	1	2	1	7	5	1	3
c_3	1	7	2	1	5	6	9
c_4	1	4	7	3	5	5	2
c_5	1	1	3	4	3	7	9
c_6	1	9	4	5	6	1	2
$F(a_i)$		28	25	23	26	27	29
Ranks		2	5	6	4	3	1

The performance metric F of each alternative plays a crucial role in determining its rank and ultimately selecting the most-preferred alternative. An example is presented in Table 2.2, where judgements range from 1 to 9, and the criteria weights are all set to 1. The computed performance F for each alternative is shown in the second-to-last row, resulting in the ranking positions in the final row. Alternative a_6 emerges as the most-preferred one, albeit by a very slight margin.

The SAW model will serve as a key component in Chapters 3, 4, 5, and 9 to exemplify the CMAA framework. Moreover, it will be used in the experimental analysis presented in Chapters 7 and 8 to explore and assess CMAA's performance characteristics.

Analytical Hierarchy Process (AHP)

AHP was initially proposed by Saaty (1990) and has emerged as one of the frequently employed compensatory decision methods in MCDM research (Munier et al., 2021). AHP distinguishes itself from other decision methods in two significant ways. First, it arranges criteria in a hierarchical structure and calculates compensatory influences accordingly. Second, AHP uses relative evaluations by comparing pairs of criteria or alternatives.

Table 2.3: Saaty's scale of relative importance

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgement slightly favour one activity over another
5	Essential or strong importance	Experience and judgement strongly favour one activity over another
7	Demonstrated importance	An activity is strongly favored and its dominance demonstrated in practice
9	Absolute importance	The evidence favouring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values between two adjacent judgements	When compromise is needed

The pairwise comparisons in the AHP are selected from a 17-point scale of the integers 1 to 9 and their reciprocals. The scale's interpretations are outlined in Table 2.3. These comparisons are used to construct transitive, reciprocal matrices, whose first eigenvectors determine the preference and judgement priorities. Subsequently, these priorities are treated akin to weights and judgements in SAW, using Equation 2.1.

Although pairwise comparisons are generally deemed easier for decision-makers to provide, AHP demands a larger number of evaluation tasks compared to other methods. The number of required judgement comparisons for each criterion grows quadratically as $m \cdot (m-1)/2$ with the number of alternatives. Additionally, considering the criteria comparisons, the workload grows quadratically with n. For example, in a decision featuring six first-level criteria and six alternatives, a total of 96 pairwise comparisons would be needed. In contrast, for a similar-sized decision, SAW would need only 42 evaluations.

An example of a preference matrix comparing five criteria on the same level is depicted

$$\begin{pmatrix} 1 & \mathbf{3} & \mathbf{1/2} & \mathbf{8} & \mathbf{4} \\ 1/3 & 1 & \mathbf{1/3} & \mathbf{7} & \mathbf{3} \\ 2 & 3 & 1 & \mathbf{9} & \mathbf{6} \\ 1/8 & 1/7 & 1/9 & 1 & \mathbf{1/4} \\ 1/4 & 1/3 & 1/6 & 4 & 1 \end{pmatrix}$$

Figure 2.1: Example AHP preference matrix

in Figure 2.1. In this matrix, the bold values in the upper triangular part represent input comparisons provided by decision-makers. The reciprocals of these values are displayed in the lower triangular part of the matrix.

Pairwise comparisons, a key feature of AHP, have been adapted for integration into other decision methods such as AHP-TOPSIS (Majumdar et al., 2005), AHP-PROMETHEE (Turcksin et al., 2011), and AHP-ELECTRE (Uddin et al., 2019).

Criticisms of AHP by Munier et al. (2021) have highlighted challenges that are pertinent to its integration with CMAA; two of them are:

- For even medium-sized decisions, a large number of pairwise comparisons are necessary.
- The cognitive effort saved by conducting pairwise comparisons is diminished by the need to maintain consistency across the 17-point scale.

Pairwise comparisons introduce a distinct input structure compared to scores, and the requirement for consistency in AHP impacts the construction of the combinatorial space in CMAA. The adaptation of AHP for use within the CMAA framework will be elaborated in Section 8.2, demonstrating how it can be harnessed effectively. Additionally, in Section 7.4.2, the consensus-building performance of AHP will be compared to that of other decision methods.

ABX-Lex - a variant of the Lexicographic method

The ABX-Lex method, introduced in Horton and Goers (2021), is a variant of the lexicographic decision approach (Fishburn, 1974) that employs three equivalence classes for judgements. Criteria in ABX-Lex are ordinal and non-compensatory, ordered from the highest priority c_1 to the lowest priority c_n.

The decision to use a small number of equivalence classes in judgements is based on considerations such as the limitations of human short-term memory (Mandler, 2020) and the need for quick decision recommendations (Mandler et al., 2012). The three equivalence classes are shown in Table 2.4, where $A \succ B \succ X$.

Table 2.4: The three equivalence classes of ABX-Lex

Evaluation	Meaning
A	The alternative fulfils the criterion excellently.
B	The alternative fulfils the criterion satisfactorily.
X	The alternative fulfils the criterion hardly or not at all.

An example of an ABX-Lex decision is presented in Table 2.5. The data is structured as a matrix, where criteria are listed in rows, organised in descending order of importance, and alternatives are displayed in columns. The ranking process involves arranging the resulting vertical 'words' in lexicographic ascending order. The ranking position of each alternative corresponds to its position based on alphabetical order.

In the context of this example, alternative a_2 emerges as the most-preferred choice, as indicated by the evaluation vector 'ABAABB'. Alternatives a_3 and a_5 lack an 'A' judgement for c_1, which excludes them from consideration as most-preferred, irrespective of their performance on criteria c_2 to c_6.

ABX-Lex's notable feature is that not all judgements may be required. These are shaded in grey in the table. This method provides the determination of all ranks by only $n \cdot \log_3(m)$ judgements on average, as outlined in Horton and Goers (2021).

Table 2.5: An example decision matrix for an ABX-Lex decision

	a_1	a_2	a_3	a_4	a_5	a_6
c_1	A	A	X	A	B	A
c_2	B	B	A	B	X	B
c_3	X	A	B	A	A	A
c_4	A	A	A	B	B	X
c_5	B	B	A	A	A	A
c_6	A	B	A	X	A	B
Ranks	4	1	6	2	5	3

In Section 7.4.2, the performance of consensus-building using CMAA with ABX-Lex will be compared to that of other methods. Additionally, ABX-Lex will be used in the case study described in Section 9.1.

Other multi-criteria decision methods

In addition to the methods discussed earlier, there are several other decision methods to be explored in later chapters to showcase their compatibility with CMAA.

Fuzzy SAW (FSAW) is a weighted sum version of SAW that incorporates fuzzy triangular numbers to represent linguistic variables (Hwang et al., 1981). FSAW is particularly useful when decision-makers prefer expressing their preferences and judgements using linguistic terms. SAW can also be adapted to a non-compensatory variant, denoted here as SAW-NC, where criteria are weighted to emulate an ordinal scale.

TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) is a compensatory method that evaluates alternatives based on their proximity to ideal and worst-case scenarios (Hwang et al., 1981). There is also a fuzzy version of TOPSIS called FTOPSIS (Salih et al., 2019).

The Weighted Product Method (WPM) has the same structure as SAW, but replaces multiplication by exponentiation and addition by multiplication (Triantaphyllou and Triantaphyllou, 2000).

PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation) is a compensatory outranking method that evaluates alternative performance by comparing judgements across all alternatives (Brans and Vincke, 1985).

The compatibility and consensus-building performance of CMAA in conjunction with these methods will be examined in Section 7.4.2.

2.2.3 Stochastic Multi-Criteria Acceptability Analysis

Lahdelma et al. (1998) introduced Stochastic Multi-Criteria Acceptability Analysis (SMAA) as a method to address cases where preferences or judgements are missing or ambiguous. The concept of SMAA was derived from ideas proposed by Bana e Costa (1988). SMAA can be used to analyse and understand the decision structure by providing insights about the strengths and dependencies for each alternative.

SMAA represents uncertain or incomplete evaluations as probability density functions. For example, if the investment cost of an alternative falls within the range of 100,000 to 120,000 Euros, a uniform probability distribution can be applied over this range. This process involves substituting each missing or ambiguous evaluation with a continuous, real-valued input in the form of a probability density function. The analysis examines the resulting high-dimensional continuous space by assuming a utility function that leads to a specific ranking position for each alternative. As this space cannot be evaluated analytically, Monte Carlo simulation is employed. The simulation samples the probability space and computes the ranking of alternatives based on it.

Throughout the simulation, various variables are monitored. The primary variable in SMAA is the rank acceptability index b_i^r, representing the fraction of the probability space resulting in rank r for alternative a_i.

Several additional analytical variables focus on the strengthening or weakening effects of various preferences. One of these variables are the confidence factors, which identify the favorable preference vector for an alternative's maximum strength. Its counterpart is the cross-confidence vector, which measures an alternative's performance with a non-favorable preference vector. Another variable is the central weight vector, representing the preference vector positioned at the center of the probability space.

These preference-related variables allow for an examination of the sensitivity of alternatives across various preference vectors, providing insights into their performance characteristics.

2.2.4 Conclusions

There exist a large number of multi-criteria decision-making methods which vary considerably in their structure and input data formats. Stochastic Multi-criteria Acceptability Analysis is unique in that it generates a large state space which has to be analysed with Monte Carlo simulation. This led to the rank acceptability as a new type of metric for preferredness, which is adopted by CMAA (see Chapter 3).

2.3 Shared Mental Model paradigm

In this section, group decision research from a Social Psychology perspective is examined. According to Stasser and Titus (1985), groups often fail to uncover important information, despite the claim that they should be able to surpass individual decision-making capabilities. This is attributed to the failure to share information that is known only to individuals within the group. Had the group accessed all pertinent information, it could have made the correct decision. Further exploration of this phenomenon is conducted in Shared Mental Model research studies (Brodbeck et al., 2007; Kim and Kim, 2008; Lu et al., 2011; Schulz-Hardt and Mojzisch, 2012).

2.3.1 Hidden profiles cause incorrect decisions

A *mental model* represents the view of a specific issue held by a decision-maker. When multiple decision-makers in a cooperative setting disagree on an evaluation, they possess an *unshared mental model* (UMM). By sharing important information, a group can establish a common understanding of the issue, known as a *shared mental model* (SMM). With a shared mental model, decision-makers can collectively assess a judgement or preference task. In their review of 25 years of SMM research Lu et al. (2011), claimed

> *[...] that groups do not exchange information efficiently and that decision quality suffers as a result.*

Successful sharing of critical information within a group, would lead to a correct decision, which no individual decision-maker is capable of on their own.

The unavailability of important information inhibits the formation of a shared mental model, causing a *Hidden Profile* situation (Schulz-Hardt et al., 2006). Hidden Profiles often result in premature consensus (Definition 6 on page 29), because already-shared information can impede the integration of unshared information (Lu et al., 2011; Mukherjee et al., 2018).

Schulz-Hardt et al. (2006) demonstrated that dissent plays a vital role in eliciting Hidden Profiles. Presenting minority opinions during discussions can enhance decision quality, when the group is cooperative. While research on Hidden Profiles confirms that sharing UMM improves decision quality, identifying Hidden Profiles remains a challenge in practice due to the holistic nature of the discussion.

2.3.2 Causes of conflicting evaluations

In group settings, each judgement and preference task can lead to conflicts, a common occurrence in practice. Achieving consensus hinges on resolving these conflicts, making it crucial to grasp their origins. According to Brodbeck et al. (2007), conflicts can arise from information asymmetries among decision-makers. Briggs et al. (2005) further categorise conflict causes into five factors, with two not stemming from information asymmetry:

1. *Differences of meaning*: The decision-makers associate different concepts with the same words or symbols, or they use different words or symbols for the same concepts.
2. *Differences of mental models*: The decision-makers base their evaluations on different assumptions.
3. *Conflicting information*: There are asymmetries in the information held by the decision-makers.
4. *Mutually exclusive individual goals*: The decision-makers have mutually exclusive individual goals.
5. *Differences of taste*: The decision-makers have different tastes. [This thesis suggests adding "... or fundamental beliefs" to this option.]

Causes 4 and 5 specify factors in competitive decisions (Definition 1 on page 2). Even if the group shares their perspectives and information, the optimal outcome it can achieve is a *compromise*. In a compromise, a decision-maker sacrifices something for the sake of reaching a solution with the rest of the group. In some scenarios, the resolution can even be the result of appeasement by one or more decision-makers (Liu et al., 2018).

On the other hand, in cooperative decisions (Definition 2 on page 3), evaluation conflicts stem from causes 1, 2, and 3, which can be resolved through the sharing of mental models by the group.

In practice, mixed cooperative/competitive decisions scenarios may arise which contain both classes of conflict causes.

2.3.3 Conclusions

In cooperative groups, correct decisions are attainable when groups successfully share their mental models. This necessitates addressing the root causes of evaluation conflicts: the unshared mental models, in order to uncover hidden profiles. Ultimately, the sharing of crucial information enhances the quality of cooperative decision-making.

Shared mental model research has not yet produced an efficient method of uncovering sharing mental models, owing to the holistic nature of the discussions. Conversely, multi-criteria decision-making holds promise in this regard, because the scope of the discussions is reduced to individual criteria/alternative pairs.

2.4 Single-step multi-criteria group decision-making

Traditional MCDM methods are designed for individual decision-maker inputs and lack the capability to handle multiple evaluations directly (Aruldoss et al., 2013; Thakkar, 2021). Consequently, these methods cannot be directly applied to group decision-making scenarios.

In contrast, Multi-Criteria Group Decision-Making (MCGDM) studies how groups can reach decisions despite encountering conflicting evaluations within the group. Often in practical applications, consensus-building efforts are omitted (Bagočius et al., 2014; Memari et al., 2019; Gao et al., 2020). Instead, each conflicting evaluation is consolidated into a single value in a straightforward manner. This methodology will be denoted as single-step group decision-making.

Aggregating conflicting evaluations

In group settings, evaluation conflicts arise when subjective inputs on a preference or judgement task differ. Single-step MCGDM methods address these conflicts by aggregating evaluations as if from a single-user (Ossadnik et al., 2016; Sodenkamp et al., 2018; Gao et al., 2020). Depending on the input type, different aggregation methods have been proposed.

For crisp numerical values, the arithmetic mean is most commonly used. For example, in a judgement task where seven decision-makers rate on a scale from 1 to 9, producing a judgement discrepancy $\lambda(j,i) = \{9,9,9,9,9,9,1\}$, the average (7.86) is typically calculated for decision computation. On the other hand, Horton et al. (2016) advocate using the median over the arithmetic mean, especially for selecting innovative projects. When dealing with fuzzy numbers, alternative aggregations like the Heronian mean have been suggested (Kumar and Chen, 2022). Pairwise comparisons, as seen in the AHP, often use the geometric mean for aggregation (Huang et al., 2009; Srdjevic et al., 2011).

In the group version of SMAA (Lahdelma and Salminen, 2001), probability density functions are used in aggregation. Conflicting input evaluations are fused into a probability density function. For example, for the judgements $\lambda(j,i) = \{1,1,2,2,6,6,7\}$ a continuous uniform distribution for the evaluation range from $[1,7]$ might be used. During the analysis, SMAA samples from this continuous distribution, which includes judgements, such as 4, that were not supplied by the decision-makers.

A drawback of aggregating conflicting evaluations is the concealment of discrepancies. Extreme discrepancies, such as a polarising $\lambda(j,i) = \{1,1,1,1,9,9,9\}$, get averaged to about 4, potentially masking important minority opinions.

Assigning different importance to decision-makers

In MCGDM, one approach to modifying aggregated evaluations involves promoting or demoting decision-makers' inputs based on their expertise level (Ozer, 2007; Sorooshian, 2018), conformity (Asuquo et al., 2019), or other attributes (Srdjevic et al., 2011; Koksalmis and Kabak, 2019). Some strategies explicitly favour majority opinions to achieve decisions (Kacprzyk and Fedrizzi, 1988; Koksalmis and Kabak, 2019).

In certain methods, decision-makers' importance is determined by the consistency of their evaluations (Blagojević et al., 2023). For example, in AHP, the influence of each decision-maker on the outcome can be calculated based on the consistency of their pairwise comparisons (Aguarón et al., 2019).

Conclusions

The primary limitation of aggregation lies in its tendency to obscure important information. In MCGDM methods, aggregating evaluations into a single value diminishes the effects of individual inputs, disregards important information and neglects minority perspectives.

In cooperative decision-making contexts, merging evaluation differences and suppressing minority viewpoints can hinder the exploration of unshared mental models. Failure to address information disparities may limit the group's ability to outperform individual decision quality (Brodbeck et al., 2007). Tran et al. (2023) highlight the dissatisfaction that can arise from single-step multi-criteria group decision approaches.

In specific group decisions, many methods assign more weight to certain decision-makers, such as senior engineers selecting new equipment. Conversely, in this thesis, in cooperative decisions (Definition 2 on page 3), all participants are assumed to be equally important, despite varying expertise levels, promoting the integration of diverse knowledge.

Single-step group decision processes that lack knowledge-pooling mechanisms may not effectively support cooperative decision-making or align with the goals of this thesis.

2.5 Consensus-building in multi-criteria group decisions

Single-step MCGDM approaches present two disadvantages in group decisions. First, the decision outcome derived from aggregating conflicting inputs may lack specificity. Second, such approaches can lead to group dissatisfaction with the resulting decision. In response to these drawbacks, consensus-building approaches are designed to enhance satisfaction levels within the group decision-making process.

2.5.1 Consensus-building strategies

In contrast to single-step MCGDM methods, consensus-building is a dynamic process aimed at enhancing agreement among decision-makers iteratively, as illustrated in Figure 2.2. Once

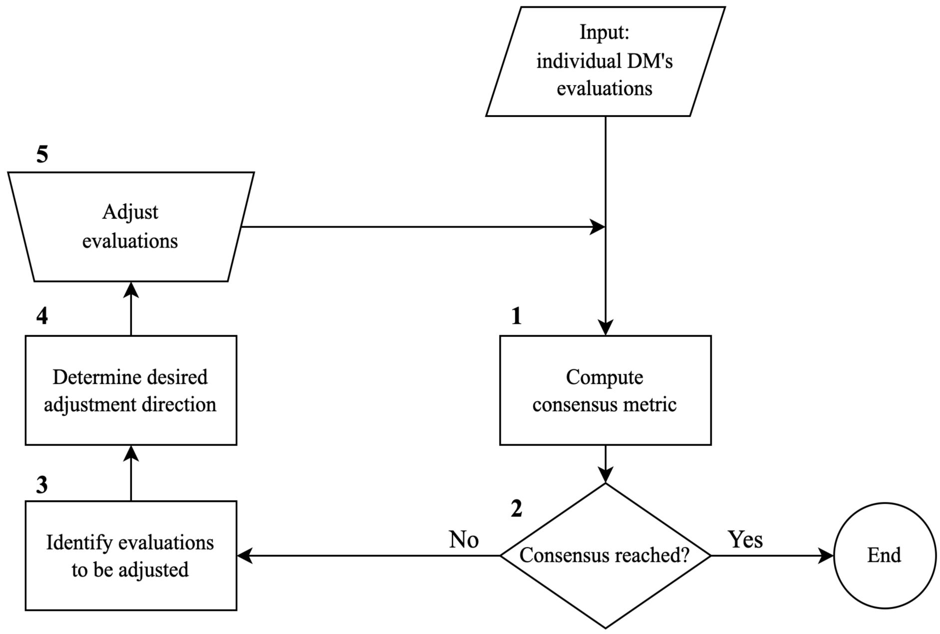


Figure 2.2: Consensus-building process

the decision has been defined and all evaluations have been provided by decision-makers, the consensus-building process unfolds through the following steps:

Step 1: A consensus metric measures the level of agreement.

Step 2: If the metric has reached the predetermined threshold, the process terminates.

Step 3: Using an *identification rule*, the evaluations to be adjusted in order to improve the consensus metric are selected.

Step 4: An adjustment suggestion is computed according to a *direction rule* and delivered to the decision-makers.

Step 5: The adjustment is carried out. The cycle repeats in Step 1 with the updated values.

Consensus-building methods vary in which consensus metric they use (Step 1), which threshold level is considered sufficient (Step 2), how they identify evaluations that need adjusting (Step 3), and the suggestion for an adjustment direction (Step 4). These are closely associated with the decision method that is being used.

In their review, Tran et al. (2023) show two ways to terminate consensus-building, if the process fails to converge. The first is to persuade towards or enforce adjustments and the second is to promote or demote decision-makers.

Persuade towards or enforce adjustments

The first approach is to identify individual evaluations which needs adjustments by the decision-makers (Herrera-Viedma et al., 2002; Dong and Saaty, 2014; Pérez et al., 2018; Zhang et al., 2019). In this approach, adjustment suggestions can be rolled out to the whole group or to individuals, depending on the identification strategy.

Other consensus-building methods prescribe a desired evaluation (Xu, 2009; Palomares et al., 2013; Dong et al., 2016; Pang et al., 2017; Lin et al., 2022). Consensus-building can fail when decision-makers refuse to adjust their evaluations. In such cases, some approaches suggest penalising these decision-makers by devaluing their preferences or judgements (Dong et al., 2016; Chao et al., 2021).

Minimal adjustment cost approaches view each adjustment as an expense that should be optimised (Dong and Xu, 2016; Zhang et al., 2019). Cost can be measured in the number of affected decision-makers, the number of adjustments made, or the number of affected alternatives. The general idea is to identify discrepancies whose resolution to a compromise promises to minimise the overall cost of consensus.

Persuading, enforcing, or automatically adjusting evaluations can only be used for competitive decisions (Definition 1 on page 2), as each measure aims to find a 'middle ground' compromise. Hence, these approaches do not fulfill the unbiased resolution requirement (Requirement 2 on page 8).

Promote and demote decision-makers

Another approach is to exclude the decision-makers that are furthest from consensus and thereby improve the consensus in the remaining group (Kacprzyk and Fedrizzi, 1988).

Some researchers assign more weight to judgements by decision-makers who are more trusted or who are considered to have more expertise for a particular criterion (Kacprzyk and Fedrizzi, 1988; Herrera-Viedma et al., 2014; Cabrerizo et al., 2017; Tundjungsari et al., 2012).

Koksalmis and Kabak (2019) have summarised approaches to deriving weights for decision-makers based on their expertise, attitudes, or knowledge. Decision-makers' judgements and preferences are promoted or demoted according to their level of knowledge and expertise (Xue et al., 2020). Some decision-makers were considered more knowledgeable than others based on the similarity in their evaluations or decision results, the consistency of their evaluations, or the entropy of their judgements (Koksalmis and Kabak, 2019).

In other approaches, a decision-maker's weight is linked to their willingness to make adjustments. Decision-makers who refuse to adjust their judgements are labeled 'non-conformists', and their evaluations are demoted. This practice is common in large-scale group decisions (Dong et al., 2016; Zhang et al., 2018; Chao et al., 2021). It raises the question of why a decision-maker who is deemed to be a non-conformist was invited to the decision. In a cooperative decision (Definition 2 on page 3), demoting a decision-maker's opinion can be detrimental as it suppresses unique information that could benefit the group, failing to meet the evaluation-preservation requirement (Requirement 1 on page 8).

Other approaches that promote or demote decision-makers take into account the evaluation consistency of decision-makers. The weights assigned to decision-makers are computed from their consistency level. The more consistent a decision-maker is, the higher their weight. These approaches have been used in fuzzy preference (Zhang and Xu, 2014) and pairwise comparison decision models (Blagojevic et al., 2016). Although inconsistencies are considered an objective way to adjust a decision-maker's influence based on their evaluation abilities, this approach raises questions.

Decision methods using consistency ratios view the redundancy of evaluations as important to allow some flexibility for decision-makers. However, maintaining consistency in evaluations requires significant cognitive effort, especially for pairwise comparisons, where mistakes are common. Instead of penalising decision-makers for inconsistencies, decision methods should aim to reduce the effort to provide consistent evaluations. However, demoting decision-makers is not appropriate in cooperative decisions, as it fails to meet the requirement that all input evaluations should be treated as equally important.

Concluding comment

Tran et al. (2023) reviewed more than 80 articles that examine multi-criteria decision-making methods and consensus-building approaches. Although it was not stated explicitly, the study only considered competitive decisions. They summarised their findings as follows:

> *Moreover, creating a group recommendation based on aggregated strategies could not guarantee a satisfactory outcome for individual group members. [...] Average can be a bad strategy for polarity groups where the preferences of group members are distributed on both sides of a spectrum. In this context, this strategy forces group members to give up their preferences to achieve an outcome that no one really likes.*

Tran et al. (2023) conclude that multi-criteria group decision-making approaches cannot guarantee a high level of agreement among all decision-makers regarding the decision recommendation.

2.5.2 Consensus metrics

The linchpin of every consensus-building process is the consensus metric. Consensus metrics can be categorised into two types:

(i) input metrics, which measure the similarity of individual input evaluations (also referred to as *coincidence between preferences*) (Palomares et al., 2018), and

(ii) output metrics, which measure the similarity of individual output choices or rankings (also known as *coincidence between solutions*) (Del Moral et al., 2018).

In Collaboration Engineering, Briggs et al. (2005) also differentiate between input and output consensus:

> *For instance, the term 'agreement' could signify 'identical mental models' or it could signify 'mutually acceptable commitments', two distinct concepts.*

Del Moral et al. (2018) emphasise that output-based consensus metrics provide a better reflection of consensus:

> *We should point out that this coincidence [between solutions] approach provides a more realistic measure of consensus among experts.*

Nonetheless, output metrics have scarcely been explored in consensus-building (Tran et al., 2023).

In competitive decisions, attaining full unanimity for all input evaluations is often deemed infeasible or excessively time-consuming. Thus, Kacprzyk and Fedrizzi (1988) introduced the concept of soft consensus. The soft consensus establishes a satisfactory level of agreement, particularly in competitive scenarios. In contrast, *hard* or *full* consensus implies achieving complete unanimity.

Input-based consensus metrics

An input metric calculates the (dis)similarities among decision-maker evaluations and is also known as a proximity measure. One such input metric involves comparing, for each evaluation, the difference between the group average and individual evaluations (Zhang et al., 2019).

The comprehensive input metric assesses the group average concerning individual evaluations through the following equation:

$$CM_I = \frac{1}{n \cdot d} \sum_{j=1}^{n} \sum_{k=1}^{d} |\mu_k(j) - \overline{\mu(j)}| + \frac{1}{m \cdot n \cdot d} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{d} |\lambda_k(j,i) - \overline{\lambda(j,i)}| \,, \tag{2.2}$$

where $\overline{\mu(j)}$ or $\overline{\lambda(j,i)}$ represents the mean preference and judgement discrepancy, respectively. A hard consensus is achieved when $CM_I = 0$ as it indicates the absence of dissimilarities.

Tran et al. (2023) outline various input metrics, with the arithmetic mean being the most prevalent, termed as *ordered weighted averaging*. Other methods include Manhattan distance, Euclidean distance, Dice distance, and Jaccard distance, each potentially yielding different consensus levels (Del Moral et al., 2018; Chiclana et al., 2015). Some metrics take into account the decision-makers' reference domain (Kacprzyk and Fedrizzi, 1988), their proximity or similarity (Cabrerizo et al., 2009), their conformity (Dong et al., 2016; Zhang et al., 2018), or their evaluation consistency (Zhang et al., 2014).

Table 2.6: Input and output consensus metrics for three judgement matrices

	a_1 {DM$_1$, DM$_2$}	a_2 {DM$_1$, DM$_2$}	a_1 {DM$_1$, DM$_2$}	a_2 {DM$_1$, DM$_2$}
c_1	{0.9, 0.2}	{0.8, 0.1}	{0.2, 0.1}	{0.1, 0.2}
c_2	{0.9, 0.2}	{0.8, 0.1}	{0.2, 0.1}	{0.1, 0.2}
$F(a_i)$	{0.9, 0.2}	{0.8, 0.1}	{0.2, 0.1}	{0.1, 0.2}
Ranks	{1, 1}	{2, 2}	{1, 2}	{2, 1}
CM_I	0.35		0.05	
CM_O	0.00		0.50	

	a_1 {DM$_1$, DM$_2$}	a_2 {DM$_1$, DM$_2$}
c_1	{0.2, 0.2}	{0.8, 0.1}
c_2	{0.9, 0.2}	{0.8, 0.1}
$F(a_i)$	{0.55, 0.2}	{0.8, 0.1}
Ranks	{2, 1}	{1, 2}
CM_I	0.26	
CM_O	0.50	

Input consensus metrics are preferred because they promptly highlight conflicts with the highest dissimilarities, and resolving these conflicts can enhance the consensus metric

(Chiclana et al., 2013). In Table 2.6, a minimal example with $m = n = d = 2$ is presented to demonstrate the impact of input consensus metrics. The decision model used is SAW, with preferences set at 0.5 for both criteria, without any conflicts. The permissible range of judgements is $[0, 1]$.

In the top-left quadrant, all judgements by DM_1 are at the high end, while those of DM_2 are at the low end. Both decision-makers prefer a_1 over a_2. The input metric CM_I according to Equation 2.2 would indicate a poor consensus level, with $CM_I = 0.35$ within a range from 0.0 to 0.5. Thus, despite both decision-makers favoring a_1 over a_2, the input metric suggests poor consensus.

In the top-right quadrant of Table 2.6, where all judgements are relatively close, the input consensus metric signifies a high degree of consensus with $CM_I = 0.05$ according to Zhang et al. (2019), who terminate consensus-building at $CM_I \leq 0.1$. However, in this scenario, each decision-maker prefers a different alternative: DM_1 favors a_1 over a_2, while DM_2 prefers a_2 over a_1.

Thus, the soft input consensus does not detect the dissent in ranking between the decision-makers. Lastly, the bottom quadrant of Table 2.6 shows high input ranges, except for the judgement task (c_1, a_1) where the decision-makers are in agreement. The input metric is high, and the rankings of each decision-maker differ. In all three instances, the input metric can identify disparities in inputs, although these dissimilarities do not inherently imply unity or discord in individual rankings.

Input metrics may give the impression of consensus progress, as eliminating discrepancies within the group leads to an improvement in the input metric, showing a continuous decline in the metric value. However, these metrics, lacking consideration for the output implications, fall short in identifying a revelation-invariant consensus (Definition 3 on page 4). Consequently, they are not suitable for cooperative decisions (Definition 2 on page 3), as they do not align with the dissertation's Goal 1 on page 7.

Output-based consensus metrics

Only a small number of output metrics have been proposed, including Kendall's coefficient of concordance (Kendall and Smith, 1939) and Spearman's rank correlation coefficient (Spearman, 1987). These metrics measure the ranking discrepancies across all decision-makers. An adapted output metric close to one by Herrera-Viedma et al. (2002) can be expressed as:

$$CM_O = \frac{1}{m \cdot d \cdot (m-1)} \sum_{i=1}^{m} \sum_{k=1}^{d} \left| r_k(i) - \overline{r(i)} \right| , \tag{2.3}$$

where $r_k(i)$ denotes the ranking for alternative a_i by decision-maker k, and $\overline{r(i)}$ represents the mean ranking for this alternative. A hard output-based consensus would be attained when $CM_O = 0.0$, while the maximum dissent would be observed with $CM_O = 0.5$.

Output-based consensus metrics are used to assess whether individual decision-makers' first choices or rankings align. However, deriving precise recommendations for adjusting evaluations based on these metrics remains challenging. For example, if a decision-maker ranks alternative a_i higher than the group average, they are simply advised to adjust their evaluations to lower its ranking position (Del Moral et al., 2018; Herrera-Viedma et al., 2002). This only provides a very vague guidance.

In reference to the illustrative example of a consensus metric in Table 2.6, an output metric is achieved solely for the table in the top-left quadrant, while the other two exhibit maximum dissent. Notably, these metrics yield conflicting results for all displayed decisions. For the example on the top-right quadrant, the input metric signals a good soft consensus, whereas the output metric is in full dissent. Conversely, in the top-left quadrant, the output metric implies hard consensus, whereas the input metric falls short of achieving even the soft consensus threshold. While a hard input consensus inherently implies output consensus, a soft input consensus may not accurately detect whether output consensus is reached or imminent.

In cooperative decisions, achieving a revelation-invariant consensus as per Definition 3 on page 4 demands an even stricter consensus level. The top-left example in Table 2.6 illustrates the insufficiency of both metrics in detecting a revelation-invariant consensus. Adjusting the evaluation of input judgement $\lambda(1,1)$ to 0.2, where both decision-makers align, changes the group decision to that of the bottom-left table. This change leads to a shift from initial hard consensus for the output metric to full dissent, as DM_1 now favors a_2 over a_1, while DM_2 maintains the opposite view. Meanwhile, the input metric only slightly adjusts to $CM_I = 0.26$, still distant from achieving even the soft consensus threshold. Consequently, the hard consensus of the output metric in the top-left subtable is premature, although showing hard consensus, fails to capture the dynamic nature of consensus under evolving evaluations.

Definition 6. *(Premature consensus) Premature consensus occurs when the consensus threshold is reached, yet resolution of further discrepancies may still alter the preferredness for one or more alternatives.*

2.5.3 Conclusion

Although output metrics are known, the majority of consensus-building approaches in MCGDM are predominantly based on input metrics. These metrics gauge the similarity or disparity in input evaluations across decision-makers, aiming to align the group towards more similar evaluations. These approaches are not well-suited for cooperative decisions for several reasons:

- Methods that prioritise certain experts' opinions over others contradict the assumption of cooperative decisions, which necessitate equal consideration of each decision-maker's expertise and knowledge (see Section 1.2.2).
- Approaches that advocate or enforce evaluation adjustments overlook the fact that each evaluation might contribute valuable information towards achieving agreement, thereby failing to meet the requirement that each input holds important information (Requirement 1 on page 8).
- Consensus metrics may prematurely signal the conclusion of the consensus-building process, failing to account for the possibility of further resolutions altering the preferredness of specific alternatives. Consequently, they do not adhere to the principle of revelation-invariant consensus (Definition 3 on page 4).

Chapter 4 will explore potential resolutions to address these challenges and enhance the effectiveness of consensus-building approaches in MCGDM by providing a new consensus metric.

2.6 Information entropy in MCGDM

In this thesis, Shannon Information Entropy will be used as the basis for an improved consensus metric. It has been employed in the past in MCDM and MCGDM contexts for other purposes.

Shannon Information Entropy, originating from communication theory (Shannon, 1948), quantifies the amount of information within a dataset. It treats data as if it was generated by a discrete random variable, with the probabilities of outcomes represented by a probability vector. A low entropy corresponds to a probability vector that is close to a standard unit vector, signifying a more ordered set of values, whereas high entropy implies low information content in the dataset and greater similarity between the probabilities.

In MCDM, entropy has been used to determine criteria weights, particularly in compensatory decision methods, where deriving weights manually can be intricate and time-consuming (Pagone et al., 2020). This led to the development of automatic weight generation models that eliminate the need for explicit preference inputs (Ciomek et al., 2017; Van Valkenhoef and Tervonen, 2016).

The Entropy Weighted Method pioneered by Hwang and Yoon (1981) and Zeleny (2012) was one of the initial applications of Shannon Information Entropy to automatically generate weights. This method is a so-called objective weighting approach, as it computes weights for criteria independently of decision-makers' input. It assesses the separation ability of each criterion based on the normalised judgements across alternatives, assigning higher weights to criteria with lower entropy, reflecting their greater importance.

In efforts to limit the feasible space of criteria weights when explicit preferences are unavailable, Ciomek et al. (2017) and Van Valkenhoef and Tervonen (2016) developed methods using Shannon Information Entropy to streamline the determination process. These methods involve decision-makers conducting pairwise comparisons between alternatives, and entropy is used to reduce the number of prompts needed to establish preferences.

Meanwhile, Yue (2017) leveraged evaluation entropy to assess the significance of decision-makers. By computing the entropy of evaluations from each decision-maker, those with contributions exhibiting low entropy (indicating high separability) are assigned higher weights, and vice versa. This approach contrasts with the common practice of downplaying extreme evaluations to minimise their impact.

While there have been diverse applications of Shannon Information Entropy in MCDM and MCGDM, a thorough review of the literature did not reveal any instances of its use as a consensus metric.

2.7 Research Gap

This dissertation aims to support cooperative decisions in achieving a revelation-invariant consensus (Definition 3 on page 4) through an efficient consensus-building process (Goal 2 on page 7).

The most important requirements for cooperative decisions are a) treat all evaluations equally, b) preserve conflicting evaluations in order to enable the resolution of important information asymmetries in conflicts and c) provide a reliable metric to detect a revelation-invariant consensus.

The key requirements for algorithmic support for cooperative decisions include:

- treating all evaluations equally (see Section 1.2.2),

- retaining conflicting evaluations to address important information asymmetries in conflicts (Requirement 1 on page 8), and
- employing a reliable metric to identify a revelation-invariant consensus (Goal 3 on page 7).

The literature review has highlighted the following key points:

1. **Shared Mental Models**.
   Successful decision-making involves sharing mental models within a group. Yet, existing research on Shared Mental Models has faced challenges in effectively sharing mental models due to the holistic nature of the discussions. MCDM approaches could enhance the exploration of unshared mental models, as the scope of each issue is much reduced.
2. **MCDM**.
   MCDM methods break down decisions into components, which simplifies the process for decision-makers to provide well-founded evaluations. However, these methods are typically designed for single decision-maker use only.
3. **Decision methods**.
   There is a wide array of decision methods available, allowing decision-makers to choose input formats and compensation strategies based on the specific decision context. Any group approach should be flexible enough to accommodate various decision methods to meet the decision-makers' diverse needs.
4. **MCGDM**.
   While MCGDM can integrate multiple decision-makers' evaluations by aggregating conflicting inputs towards a mean, this approach conflicts with the principles of treating all evaluations equally (see Section 1.2.2) and measuring the impact of each input on the group decision outcome (Goal 4 on page 7), which are crucial aspects of cooperative decisions.
5. **Consensus-building**.
   Consensus-building approaches may reach impasses and encourage decision-makers to cut corners to achieve consensus. Encouraging evaluation adjustments to specific values could bias decision-makers, undermining the openness and fairness required for cooperative decisions (Requirement 2 on page 8).
6. **Consensus-metrics**.
   Existing consensus metrics, whether based on input or output, are not able to detect a revelation-invariant consensus. Hence, there is a need for a new metric capable of fulfilling this requirement.

Consequently, the goal of this dissertation is to propose a consensus-building framework tailored to these specific requirements of cooperative decision-making.

3

Combinatorial Multicriteria Acceptability Analysis

In this chapter, the new CMAA framework will be presented. It facilitates a detailed analysis of group decisions (Research Question 3 on page 11) using several new analytical variables and an algorithm to search the combinatorial space.

CMAA is designed for two primary goals of this work. First, it should be able to integrate with any multi-criteria decision-making method, and should provide a comprehensive framework for group decisions (Goal 5 on page 8). Second, the analysis should provide variables capable of predicting the consensus impact of conflict resolutions, which would align with Goal 4 on page 4.

Additionally, the combinatorial space should meet Requirement 1 on page 8 by preserving all input data. To demonstrate the new analytical capabilities of CMAA, an illustrative example will be presented.

This chapter builds partially on published work by Goers and Horton (2023a).

3.1 The combinatorial model

3.1.1 Design considerations and principles

The new framework is specifically designed for cooperative decisions and their need to reach a revelation-invariant consensus (Definition 3 on page 4). All groups initially create preference or judgement discrepancies. The new framework is built upon the observation that a discrepancy is caused by an unshared mental model:

Observation 1. *In a cooperative decision, all decision-makers will submit the same judgement or preference, when their mental models for it have been shared.*

From this observation, three corollaries can be derived.

Corollary 1. *In a cooperative decision, all decision-makers will agree on the preferredness of each alternative, once they have shared their mental models for each judgement and preference. In this sense, a cooperative decision has a unique correct solution (Definition 4 on page 5).*

This Corollary leads to the first design principle of CMAA, that a group decision and consensus-building process must provide the opportunity to share mental models among all decision-makers.

Corollary 2. *In a cooperative decision, any unshared mental model might turn out to be decisive when it is shared. From this follows that all judgements and preferences are potentially equally important and should be treated as such.*

From Corollary 1 follows that a consensus-building approach that prescribes adjustment corrections may introduce a bias into the decision, which leads to the third Corollary:

Corollary 3. *In a cooperative decision, any consensus-building process that encourages or requires decision-makers to modify their inputs towards a particular value (such as a group mean) introduces a bias.*

So, for the design principle that a consensus-building process needs to provide an opportunity for sharing mental models, Corollary 3 adds the condition that the process of sharing mental models must not be biased by suggesting a specific adjustment for one or more decision-makers.

Catering for these corollaries implements Requirement 1 on page 8 and Requirement 2 on page 8. These goals require the preservation of all input evaluations and an unbiased opportunity to clarify the information that lead to all evaluations.

If a consensus-building method fulfils the former requirements, it should be able to pass the Unique Argument Test (Horton and Goers, 2021). This test demands of a consensus-building process:

> *If there is a clinching argument for a judgement or preference that is only known to one member of the group, then the group decision method must – in principle – be able to produce that judgement or preference as a unanimous result.*

3.1.2 The combinatorial space generated by all inputs

In a cooperative decision, a discrepancy reveals that decision-makers have an unshared mental model. In order to keep this indication, all inputs need to be preserved for the group decision analysis. Unfortunately, no standard decision analysis or decision method can be applied to ambiguous inputs without aggregating them. So, a new decision analysis approach is required.

The premise is that each preference and judgement discrepancy must be kept as-is, because it is unknown which resolution the group will choose. The consensus-building algorithm assumes the resolution to each of the inputs of a discrepancy is equally probable.

The discrepancies create a space of combinations. It is assumed that large numbers of combinations will lead to the same most-preferred alternative, making it unnecessary to resolve all discrepancies to build a revelation-invariant consensus (Hypothesis 1 on page 10). In other words decision-makers will need to share their mental models only for a subset of the discrepancies in order to arrive at the correct decision.

The analysis can be conducted in the following way. The combinatorial space consists of all possible combinations of preferences and judgements. An *instance* of the combinatorial space is a combination of preferences and judgements that corresponds to a standard (i.e., single-user) input for a decision method. For each instance, a ranking can be computed using the chosen multi-criteria decision method. From all instances and their resulting rankings, two types of variables can be derived:

i. **Overall**: The *rank acceptability index* b_i^r is the proportion of instances in which alternative a_i attained rank r, in an analogous manner to SMAA (Lahdelma and

Salminen, 2001). The rank acceptabilities can be used to compute each alternative's preferredness.

ii. **Evaluation-related**: Variables that measure the effect of each preference and judgement on the strength of each alternative – both in the current state and also as a prediction for the next consensus-building step.

Although SMAA introduced rank acceptabilities into decision analysis, its aggregation of ambiguous preferences and judgements to probability density functions destroys the identifiability of individual evaluations. It is no longer possible to trace which input affects the rank acceptability of an alternative, which is essential for a detailed understanding and subsequent consensus-building.

Thus, CMAA is the first approach currently able to compute the effects each input has on the ranking performance of the alternatives and locate the most or least influencing discrepancies. By introducing these abilities, CMAA enables a detailed understanding of the effects of individual preferences and judgements on the strength of each alternative.

3.2 The combinatorial analysis algorithm

3.2.1 Notation

This section defines the notation for the combinatorial analysis. The set of inputs for either preferences or judgements will be noted in curly brackets, for example for linguistic variables {very good, poor, poor, fair}, or for numerical values $\{1, 2, 5, 3\}$.

Table 3.1: Example for an individual (left) and a unified preference vector (right)

	DM_1	$\mathbf{P}_j$
c_1	{6}	{5,6,7}
c_2	{5}	{5,7,8}
c_3	{1}	{1,8,9}
c_4	{3}	{3,7,9}
c_5	{4}	{2,4}
c_6	{2}	{1,2}

Individual preferences can be arranged as an individual preference vector for each decision-maker DM_k. An example for DM_1 is shown in Table 3.1 (left). It is assumed that identical preference or judgement evaluations by different decision-makers are a reflection of a shared mental model and can be unified. The set of individual preference vectors of all decision-makers creates the *unified preference vector* $\mathbf{P}$. Table 3.1 (right) illustrates an example for a unified preference vector. Each element $\mathbf{P}_j$ consists of the union of all individual preferences for criterion c_j:

$$\mathbf{P}_j = \bigcup_k \mu_k(j). \tag{3.1}$$

The *preference valency* ϕ_j is the cardinality of the set of inputs for the corresponding preference task (c_j).

$\mathbf{P}_j(k)$ is a version of the unified preference vector $\mathbf{P}$, where for one preference discrepancy $\mu(j)$, only the preference $\mu_k(j)$ is considered. k is a local variable to each discrepancy.

Table 3.2: Example of an individual judgement matrix

	DM_1					
	a_1	a_2	a_3	a_4	a_5	a_6
c_1	{8}	{2}	{9}	{7}	{6}	{6}
c_2	{4}	{5}	{5}	{5}	{3}	{7}
c_3	{5}	{1}	{5}	{2}	{3}	{8}
c_4	{5}	{4}	{4}	{3}	{8}	{3}
c_5	{7}	{2}	{7}	{5}	{3}	{6}
c_6	{4}	{6}	{6}	{2}	{5}	{5}

Individual judgement tasks can be arranged as an individual judgement matrix for each DM_k. An example is illustrated in Table 3.2. The *unified judgement matrix* $\mathbf{A}_{n\times m}$ unifies all individual judgement matrices. Each element $\mathbf{A}_{ji}$ of the unified judgement matrix $\mathbf{A}$ is the set formed by the union of the corresponding individual judgements:

$$\mathbf{A}_{ji} = \bigcup_k \lambda_k(j,i). \tag{3.2}$$

Table 3.3: Example of a unified judgement matrix

	$\mathbf{A}_{ji}$					
	a_1	a_2	a_3	a_4	a_5	a_6
c_1	{7,8,9}	{2,5,7}	{1,8,9}	{2,7}	{3,6}	{4,6,7}
c_2	{3,4,8}	{1,3,5}	{5,7}	{5,9}	{3,4}	{6,7}
c_3	{2,3,5}	{1,6}	{2,5,8}	{2,6,7}	{1,3}	{5,8}
c_4	{2,5,8}	{1,4,8}	{4,6,9}	{1,3,8}	{2,6,8}	{3,5,6}
c_5	{3,4,7}	{2,5,9}	{6,7}	{2,5,6}	{3,5}	{2,5,6}
c_6	{4,6}	{6,8}	{4,5,6}	{1,2}	{5,8,9}	{5}

Table 3.3 shows an example of a unified judgement matrix. $\mathbf{A}_{ji}(k)$ is a version of the unified judgement matrix $\mathbf{A}$, where for one judgement discrepancy $\lambda(j,i)$, only the judgement $\lambda_k(j,i)$ is considered. k is a local variable at each discrepancy.

The unified preference vector $\mathbf{P}$ combined with the unified judgement matrix $\mathbf{A}$ creates the combinatorial space of CMAA. The complete group decision will be denoted by $[\mathbf{P};\mathbf{A}]$.

An instance of the unified preference vector is denoted by $\mathbf{P}^*$. Similarly, an instance of the unified judgement matrix is denoted by $\mathbf{A}^*$. Thus, one instance of the combinatorial space that the decision $[\mathbf{P};\mathbf{A}]$ spans corresponds to a single-user representation $[\mathbf{P}^*;\mathbf{A}^*]$. The instance $[\mathbf{P}^*;\mathbf{A}^*]$ can be processed by a standard decision method.

The number of preference instances $||\mathbf{P}||$ is given by

$$||\mathbf{P}|| = \prod_j \phi_j.$$

The number of judgement instances $||\mathbf{A}||$ is given by

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

The number of possible unique combinations (instances) K in $[\mathbf{P}; \mathbf{A}]$ is

$$K = ||\mathbf{P}|| \cdot ||\mathbf{A}||. \tag{3.3}$$

The combinatorial space grows exponentially with the number of discrepancies. Thus, even a very small decision ($d = m = n = 3$) with twelve discrepancies can generate more than half a million instances.

With this notation, CMAA is independent of the various input types. There are two exceptions: the decision method Analytic Hierarchy Process (AHP) and a new version of a decision method that maps objective measurements to subjective judgements. Because these methods use non-standard input formats, CMAA must be adapted to accommodate them. These adaptations will be explained in Chapter 8.

3.2.2 Combinatorial acceptability analysis algorithm

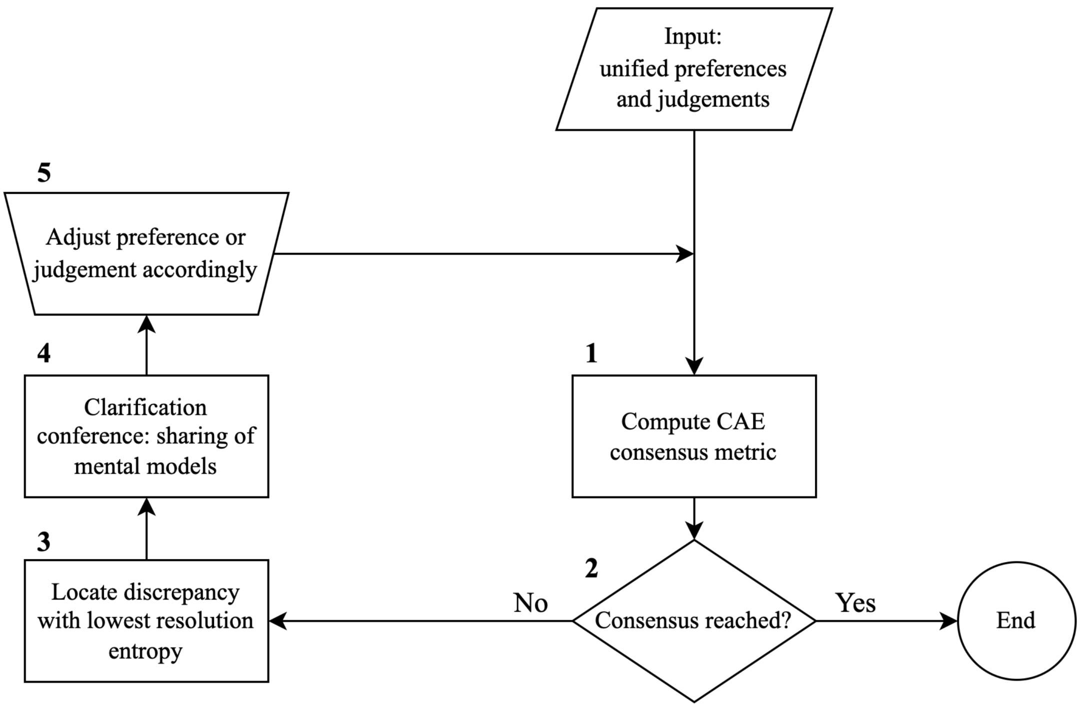


Figure 3.1: Flowchart of the CMAA analysis algorithm

The structure of the combinatorial decision analysis algorithm is depicted in Figure 3.1. The decision-makers provide evaluations in the format required by the selected decision method. Then, all instances of the combinatorial space of all inputs are generated. The decision method is applied to each instance, and the corresponding rankings are computed. For each instance, the ranking of each alternative is noted as well as the preferences and judgements it contains. Furthermore, it is counted how often each alternative reached each rank for the entire set of instances. From these, the rank-acceptabilities can be computed. Additionally, the contribution of each preference or judgement to the rank-1 acceptability of each alternative and the next-step acceptability that would result from each resolution are computed. In practice, only a small fraction of the total number of instances is generated, as will be described in Section 3.4.

Algorithm 1: Combinatorial acceptability analysis

Given: Decision model and decision algorithm
Input: Decision $[\mathbf{P}; \mathbf{A}]$
1 **for** *each instance* $[\boldsymbol{P}^{*}; \boldsymbol{A}^{*}]$ **do**
2   Apply the decision algorithm to this instance;
3   Update counter variables;
4 **end**
5 Compute statistics from the counter variables;
Output: Acceptability indices and other statistics

In Algorithm 1, the analytical procedure is formalised. Given are the decision model and its corresponding decision method. The input to the framework is the decision $[\mathbf{P}; \mathbf{A}]$, which is defined in Equation 3.1 and Equation 3.2.

Lines 1-4 The following computations are applied to each instance $[\mathbf{P}^{*}, \mathbf{A}^{*}] \in [\mathbf{P}, \mathbf{A}]$.

Line 2 The decision algorithm is applied to the instance, which returns a ranking of the alternatives.

Line 3 The counters are updated. These counters will be explained in the next Section.

Line 5 When all instances have been processed, the counters are used to compute the output statistics.

The outputs of the CMAA analysis algorithm are the rank acceptability indices and other statistics derived from the two types of variables.

The complexity $\mathcal{C}$ of Algorithm 1 is

$$\mathcal{C} = \mathcal{O}(\mathcal{D} \cdot K) \ , \tag{3.4}$$

where $\mathcal{D}$ is the complexity of the decision algorithm for one instance of the decision (line 2 in Algorithm 1). For many compensatory decision methods, $\mathcal{D} = \mathcal{O}(n + (m \cdot n))$. Since $\mathcal{D}$ is usually relatively small, the complexity of the CMAA analysis $\mathcal{C}$ is dominated by the number of instances K.

3.2.3 Measuring the performance of each alternative

In Algorithm 1, various variables are computed that provide insights for the group regarding their decision structure. All variables are illustrated in Section 3.5.

The variable B_i^r counts the number of times an alternative a_i achieves rank r for all accounted instances.

One of the most important statistics is the *rank acceptability index* b_i^r. It computes the proportion of the K instances for which alternative a_i achieves rank r:

$$b_i^r = B_i^r[\mathbf{P}; \mathbf{A}]/K \ . \tag{3.5}$$

This variable is used to determine the overall performance for each alternative. Under the condition that no instance generates tied ranks, the following equality holds:

$$\sum_{i=1}^{m} b_i^r = 1 \ . \tag{3.6}$$

If the decision involves instances with ties, Equation 3.6 will not hold, because B_i^r will be incremented by each tied alternative at rank r, and the value of B_i^{r+1} will be depleted correspondingly. But even when ties are present in the decision, the following equation holds

$$\sum_{r=1}^{m} b_i^r = 1 \ . \tag{3.7}$$

This identity can be used in unit tests to verify the counter results.

The rank acceptability index can be used to compute the *holistic acceptability* acc_i of an alternative a_i as a measure of its preferredness. For decisions where consensus-building cannot be carried out, the preferredness can be used to determine the choice, if the holistic acceptability indicates a strongly preferred alternative. The holistic acceptability is a linear combination of the rank acceptabilities:

$$acc_i = \sum_{r=1}^{m} \alpha_r \cdot b_i^r \ , \tag{3.8}$$

where the coefficients α_r must satisfy $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_m$. One possibility is to use the inverse weights $\alpha_r = 1/r$ suggested by Lahdelma and Salminen (2001). Another is to use 'Olympic' meta-weights, for example $\alpha_1 = 0.6, \alpha_2 = 0.3$ and $\alpha_3 = 0.1$, and all other $\alpha_r = 0.0$. This choice of meta-weights is justified by Goers et al. (2024). However the α-weights are chosen, their sum should always be 1. Note that this requirement is violated by SMAA's inverse weights with $\alpha_r = 1/r$.

In this thesis, the rank-1 approach will be used to identify the most-preferred alternative: if not otherwise stated,

$$\alpha_r = \begin{cases} 1, & \text{if } r = 1 \\ 0 & \text{otherwise} \end{cases} \ . \tag{3.9}$$

If an alternative has a rank-1 acceptability of 1, it means that the alternative ranks first for every instance of the combinatorial space; if it is 0, there is no instance for which it achieves rank 1. Both insights can be used to include or exclude the alternative from the decision.

The shape of the rank-1 acceptability vectors provides insight about the need for consensus-building. For example, when the rank-1 acceptabilities of three alternatives are $(0, 1, 0)$, no consensus-building is required, because the second alternative is most-preferred in all instances, and consensus is revelation-invariant. However, for other shapes where the rank-1 acceptabilities are shared between the alternatives such as $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ or $(0, \frac{1}{5}, \frac{4}{5})$ no recommendation can be made. The reason is that significant numbers of instances return different most-preferred alternatives. This means that there are active discrepancies with unshared mental models that favour more than one alternative, whose resolution could result in large swings in the proportion of instances that return different alternatives as most-preferred.

The rank acceptability index and the holistic acceptability are novel analysis measures, because they take all possible combinations of inputs into account when measuring the rank proportions for each alternative.

3.2.4 Dependency of performance on decision-maker evaluations

The next set of variables measure the effect of each individual preference and judgement on the current and next-step performance of each alternative. These variables provide the opportunity to analyse the performance dependencies of an alternative and thereby facilitate consensus-building.

Contribution of preferences to current performance

The first of these variables measures the contribution each preference makes to the current rank-1 performance of each alternative. This helps in understanding whether the performance of an alternative is more or less dependent on a specific preference for a discrepancy.

The variable $S_i(\mu_k(j))$ counts the number of rank-1 positions for alternative a_i across all instances of the decision $[\mathbf{P}_j(k); \mathbf{A}]$ when the criterion value of c_j is equal to the preference $\mu_k(j)$:

$$S_i(\mu_k(j)) = B_i^1[\mathbf{P}_j(k); \mathbf{A}] \ . \tag{3.10}$$

The *current preference acceptability* $s_i(\mu_k(j))$ is the proportion of instances in the decision $[\mathbf{P}_j(k); \mathbf{A}]$ in which the criterion value c_j is equal to the preference $\mu_k(j)$ and alternative a_i achieves rank 1:

$$s_i(\mu_k(j)) = \frac{S_i(\mu_k(j))}{K} \ . \tag{3.11}$$

The current preference acceptability for an alternative shows to what extent its rank-1 performance depends on each preference $\mu_k(j)$ in the preference discrepancy $\mu(j)$. When the s_i values for a particular preference discrepancy are approximately equal, the alternative's rank-1 performance is only weakly dependent on a specific preference. However, if the values are unevenly distributed, the decision-makers' preferences make very different contributions to that alternative's rank-1 performance.

For an alternative a_i, the sum of all current preference acceptabilities $s_i(\mu_k(j))$ for the preference discrepancy $\mu(j)$ should be equal to its rank-1 acceptability:

$$\sum_k s_i(\mu_k(j)) = b_i^1 \ .$$

This equality can be used also for unit testing the analysis results.

Effect of preference resolutions on performance

In consensus-building, understanding which preference discrepancy resolution moves the decision towards an improved consensus is crucial. Consensus is improved when the rank-1 acceptabilities of a decision become more similar to a standard unit vector. Consequently, the analysis must provide a measure to assess the impact of a preference resolution on the rank-1 acceptability of the alternatives. The *potential preference acceptability* $\hat{s}_i(\mu_k(j))$ offers this insight. It is the proportion of rank-1 acceptabilities for a_i if the group resolved a discrepancy to the preference $\mu_k(j)$. This index is formed from the potential unified preference vector $\mathbf{P}_j(k)$ and the post-resolution number of instances K_j:

$$\hat{s}_i(\mu_k(j)) = \frac{S_i(\mu_k(j))}{K_j} \ . \tag{3.12}$$

K_j is the number of instances for the unified preference vector $\mathbf{P}_j(k)$, if the group resolved a preference discrepancy to one of its preferences $\mu_k(j)$, which is

$$K_j = \frac{K}{\phi_j} \ . \tag{3.13}$$

In other words, if the group chose the preference resolution $\mu_k(j)$, then $\hat{s}_i(\mu_k(j))$ expresses the new acceptability that the alternative a_i would achieve. Essentially, the potential preference acceptability can predict the acceptability of a_i one consensus-step into the future.

Potential preference acceptabilities can be computed for all alternatives and preference resolutions. This allows an analysis of by how much each preference resolution would boost or diminish the rank-1 acceptability of each alternative. It can be used for additional guidance in consensus-building.

Preference sensitivities

The *preference sensitivity* measures the largest positive or negative change a preference resolution for the discrepancy $\mu(j)$ would cause to alternative a_i's rank-1 acceptability b_i^1. The preference sensitivity $\sigma_i(j)$ is computed as follows:

$$\begin{aligned} \sigma_i(j) &= \begin{cases} \delta_+ & \text{if } \ \delta_+ > |\delta_-| \\ \delta_- & \text{otherwise} \end{cases} , \quad \text{where} \\ \delta_+ &= \max_k(\hat{s}_i(\mu_k(j)) - b_i^1) \\ \delta_- &= \min_k(\hat{s}_i(\mu_k(j)) - b_i^1) \ . \end{aligned}$$

The ranges of these two variables are $\delta_+ \in [0, 1 - b_i^1]$ and $\delta_- \in [-b_i^1, 0]$.

Contribution of judgements to performance

Each of the three preference-based analysis variables has a judgement-based counterpart.

The counter $Q_i(\lambda_k(j, i2))$ counts how often alternative a_i achieved rank 1 for all instances of the decision $[\mathbf{P}; \mathbf{A}_{ji2}(k)]$ in which the judgement at c_j, a_{i2} has the value of $\lambda_k(j, i2)$:

$$Q_i(\lambda_k(j, i2)) = B_i^1[\mathbf{P}; \mathbf{A}_{ji2}(k)] \ . \tag{3.14}$$

The *current judgement acceptability* $q_i(\lambda_k(j, i2))$ is the corresponding proportion of K instances:

$$q_i(\lambda_k(j, i2)) = \frac{Q_i(\lambda_k(j, i2))}{K} \ . \tag{3.15}$$

For the judgement discrepancy $\lambda(j, i2)$, the sum of all current judgement acceptabilities $q_i(\lambda_k(j, i2))$ for an alternative a_i is equal to its rank-1 acceptability.

$$\sum_k q_i(\lambda_k(j, i2)) = b_i^1 \ .$$

The current judgement acceptability illustrates the amount of rank-1 acceptability an alternative receives from all instances in which the judgement at c_j, a_{i2} has the value $\lambda_k(j, i2)$. An alternative's rank-1 acceptability is completely dependent on a judgement $\lambda_{k1}(j, i2)$, when its current judgement acceptability is equal to b_i^1, and the current judgement acceptability for all other judgements is 0.

Effect of judgement resolutions on performance

The *potential judgement acceptability* $\hat{q}_i(\lambda_k(j, i2))$ measures the impact of a judgement resolution on the performance of an alternative. It is the proportion of instances in which alternative a_i would achieve rank 1, if the group resolved the judgement discrepancy $\lambda(j, i2)$ to the judgement $\lambda_k(j, i2)$:

$$\hat{q}_i(j, i) = \frac{Q_i(\lambda_k(j, i2))}{K_{ji2}} \;, \tag{3.16}$$

where K_{ji2} is the number of instances remaining after the resolution:

$$K_{ji2} = \frac{K}{\phi_{ji2}} \;. \tag{3.17}$$

Judgement sensitivity

Judgement resolutions can have significantly different impacts on an alternative a_i's rank-1 acceptability. Understanding these dependencies is an important element of decision analysis. Thus, the *judgement sensitivity* $\sigma_i(j, i2)$ measures the largest positive or negative impact on an alternative's a_i rank-1 acceptability when the group resolves judgement discrepancy $\lambda(j, i2)$ to the judgement resolution $\lambda_k(j, i2)$:

$$\sigma_i(j, i2) = \begin{cases} \delta_+ & \text{if } \delta_+ > |\delta_-| \\ \delta_- & \text{otherwise} \end{cases} , \quad \text{where} \tag{3.18}$$

$$\delta_+ = \max_k(\hat{q}_i(\lambda_k(j, i2)) - b_i^1) \tag{3.19}$$

$$\delta_- = \min_k(\hat{q}_i(\lambda_k(j, i2)) - b_i^1) \;. \tag{3.20}$$

The ranges of the two supporting variables are $\delta_+ \in [0, 1 - b_i^1]$ and $\delta_- \in [-b_i^1, 0]$.

Summary

These new variables provide detailed insights into the contributions of each preference or judgement on the alternative's performance as well as the performance effects of specific preference or judgement resolutions. Current state-of-the-art sensitivity analysis is based on single-input decision models and can only consider the effects of changes in aggregated evaluations. By contrast, the new variables can measure the effects of each individual input while taking into account all possible combinations of inputs.

3.3 Active, inactive and pivot discrepancies

Discrepancies can vary in their potential impact on the performance of the alternatives, and identifying low-impact and high-impact discrepancies is a helpful guide for the group facilitator. Two types of discrepancies, termed *active* and *inactive*, provide a quick understanding of which discrepancies are irrelevant and which impact the separation of rank-1 acceptabilities the most.

Inactive discrepancies are those whose resolution would not impact any rank-1 acceptabilities. Hence, they are irrelevant for the group decision and do not need to be resolved. A preference or judgement discrepancy is inactive, when all preference or judgement sensitivities are 0, respectively:

$$\begin{aligned} \forall i : \sigma_i(j) = 0 \ , \\ \forall i : \sigma_i(j, i2) = 0 \ . \end{aligned} \tag{3.21}$$

On the other hand, *active discrepancies* are those whose resolution would impact the rank-1 acceptabilities and must therefore be considered during consensus-building. A preference or judgement discrepancy is active when it contains a non-zero preference or judgement sensitivity:

$$\begin{aligned} \exists i : \sigma_i(j) \neq 0 \ , \\ \exists i : \sigma_i(j, i2) \neq 0 \ . \end{aligned} \tag{3.22}$$

Pivot discrepancies are active discrepancies whose resolution can lead to the elimination of at least one alternative. A discrepancy is considered a pivot when it contains a potential preference or judgement acceptability that is equal to 0:

$$\begin{aligned} \exists i : \hat{q}_i(\mu_k(j)) = 0 \ , \\ \exists i : \hat{q}_i(\lambda_k(j, i2)) = 0 \ . \end{aligned} \tag{3.23}$$

In CMAA decision analysis, these distinctions can visualise an important aspect of the decision structure for the group or their facilitator by making irrelevant, relevant and critical discrepancies identifiable at a glance. Examples of their use will be shown in Section 5.4.3 on page 65, in Section 9.1.4 on page 127, and in Section 9.2.3 on page 133.

3.4 Monte Carlo simulation for very large spaces

A decision $[\mathbf{P}; \mathbf{A}]$ can generate a combinatorial space of instances of a size K that makes it intractable. For a medium-sized decision with five criteria, ten alternatives, and three decision-makers, K can become $||\mathbf{P}|| \cdot ||\mathbf{A}|| = 3^5 \cdot 3^{50} = 1.74 \cdot 10^{26}$. When run on a MacBook Pro with a 2.3 GHz 8-core processor and using JavaScript, the average runtime to compute a ranking using SAW for 10,000 instances was 35 ms. For all runtime simulations in this thesis, the same runtime environment was used. The runtime to analyse this decision would be on the order of 193 quadrillion years. Since the CMAA analysis is intended for interactive use during a group decision meeting, it needs to produce results within seconds.

For this reason, Monte Carlo simulation could be used instead to reduce computation times to an interactive level. Monte Carlo simulations are often used to study large spaces that would not otherwise be tractable (Von Neumann and Ulam, 1951; Kroese et al., 2014). They use random samples to explore the space and compute an approximation of the desired variable.

The idea behind Monte Carlo simulation is that a sufficiently large random sample of the combinatorial space of size K_{MC}, where $K_{MC} \ll K$, will provide a sufficiently good approximation. This has the potential to make CMAA analysis feasible for use in a live consensus-building process. The key questions is therefore: what is the required number of samples to achieve a sufficiently accurate approximation for CMAA?

Stochastic Multicriteria Acceptability Analysis also uses a Monte Carlo simulation to analyse its continuous state space. Tervonen and Lahdelma (2007) stated that to achieve a 95% confidence interval with half-width 0.01 for b_i^r, 9,604 samples are needed. Because the arguments for SMAA and CMAA are the same, $K_{MC} = 10,000$ will be used here for CMAA as well.

In CMAA, rank acceptabilities fall within the interval [0,1]. In the worst-case scenario, if an alternative's rank-1 acceptability of 0.01 were erroneously computed as 0 due to insufficient sampling, its exclusion would not significantly impact the decision. However, if an alternative's rank-1 acceptability is already 0.99, the error might affect the consensus efficiency and elongate the consensus path. For this work, an error bound of 0.01 will be taken to be sufficient. If desired, the group can reduce the error by increasing the number of samples, if the computation times allow it.

To adapt CMAA to a Monte Carlo simulation, line 1 of Algorithm 1 must be

1: **for** K_{MC} random instances $[\mathbf{P}^*; \mathbf{A}^*]$ do ,

and the variable K in all equations in Section 3.2.3 must be replaced by K_{MC}.

The accuracy of the approximations of the Monte Carlo simulation will be investigated experimentally in Section 7.1, where potential effects on the analysis are discussed, and runtimes are examined.

3.5 Illustrative example

The group decision

This section contains an example group decision to illustrate the capabilities of the new variables in the initial analysis step. The decision consists of six criteria, six alternatives ($n = m = 6$), and three decision-makers ($d = 3$). Evaluations are integers between 1 and 9, and the decision method used is SAW (see Section 2.2.2). The unified preferences and judgements are depicted in Tables 3.1 and 3.3, respectively.

The group decision consists of six preference tasks, all of which are discrepancies, and 36 judgement tasks, of which 35 are discrepancies. Four of the six preference discrepancies have a valency of three, and the other two have a valency of two. 14 judgement discrepancies have a valency of two, and 21 discrepancies have a valency of three. So, the number of instances $K = ||\mathbf{P}|| \cdot ||\mathbf{A}|| = (1^0 \cdot 2^2 \cdot 3^4) \cdot (1^1 \cdot 2^{14} \cdot 3^{21}) > 5.5 \cdot 10^{16}$ according to Equation 3.3. Approximate values for the variables are computed using a Monte Carlo simulation with $K_{MC} = 10,000$ instances.

Rank acceptabilities

The rank acceptability indices for this problem are shown in Table 3.4. The sum of the rank-1 acceptabilities is greater than 1, indicating that there are some instances containing rank-1 ties. The rank-1 acceptabilities yield the ranking $a_3 \succ a_6 \succ a_1 \succ a_4 \succ a_2 \succ a_5$, with two strong alternatives (a_3 and a_6), two medium-performing alternatives (a_1 and a_4), and two weakly performing alternatives (a_2 and a_5).

In the last column of Table 3.4, the Olympic holistic acceptabilities acc_i according to Equation 3.8 are displayed. They produce the same ranking as the rank-1 acceptabilities, but

Table 3.4: Rank Acceptability Index and Olympic-acceptability acc_i

	b_i^1	b_i^2	b_i^3	b_i^4	b_i^5	b_i^6	acc_i
a_1	0.149	0.207	0.248	0.225	0.129	0.042	0.176
a_2	0.034	0.067	0.109	0.168	0.249	0.373	0.051
a_3	0.451	0.207	0.145	0.104	0.065	0.027	0.347
a_4	0.129	0.155	0.177	0.209	0.182	0.147	0.142
a_5	0.004	0.024	0.072	0.177	0.335	0.389	0.017
a_6	0.250	0.348	0.249	0.114	0.033	0.005	0.280

the advantage of a_3 is reduced. This is a consequence of the strong rank-2 performance by a_6. When the rank-1 performance of two alternatives is close, the holistic acceptability can provide a more reliable comparison by taking into consideration performance at additional ranks.

Current and potential preference acceptabilities

Table 3.5: Current preference acceptabilities for the preference discrepancy $\mu(4)$

	$\mu(4)$		
	$\mu_1 = 3$	$\mu_2 = 7$	$\mu_3 = 9$
s_1	0.035	0.075	0.040
s_2	0.006	0.018	0.011
s_3	0.108	0.229	0.115
s_4	0.032	0.065	0.032
s_5	1E-4	0.002	0.002
s_6	0.074	0.122	0.055

Current preference acceptabilities exist for all preference discrepancies. Table 3.5 shows an example from the illustrative decision: the current preference acceptabilities regarding the preference discrepancy $\mu(4)$. It shows that the alternative a_3 receives about half of its rank-1 acceptability of 0.451 (refer to Table 3.4) from the preference evaluation $\mu_2(4) = 7$.

Table 3.6: Potential preference acceptabilities for the preference discrepancy $\mu(4)$

	$\mu(4)$		
	$\mu_1 = \{3\}$	$\mu_2 = \{7\}$	$\mu_3 = \{9\}$
$\hat{s}_1$	0.14	0.15	0.16
$\hat{s}_2$	0.02	0.03	0.04
$\hat{s}_3$	0.42	0.45	0.45
$\hat{s}_4$	0.13	0.13	0.13
$\hat{s}_5$	5E-4	4E-3	8E-3
$\hat{s}_6$	0.29	0.24	0.22
$\sum$	1.00	1.00	1.00

In Table 3.6, the potential preference acceptabilities $\hat{s}_i$ regarding the preference discrepancy $\mu(4)$ are shown for the illustrative example. Analogous values exist for the other discrepancies. It shows the rank-1 acceptability the corresponding alternative a_i would reach, if the decision-makers resolved the preference discrepancy to $\mu_k(j)$. The table shows that no resolution would change the rank-1 acceptability of an alternative very much, compared to their current rank-1 acceptabilities, shown in Table 3.4. By inspecting the values of the preference discrepancies, the facilitator can quickly see if there is a resolution that would significantly affect the preferredness of any of the alternatives.

Current and potential judgement acceptabilities

Table 3.7: Current judgement acceptabilities q_i for the judgement discrepancy at $\lambda(j, 3)$

	$q_1(\lambda_k(j,3))$			$q_2(\lambda_k(j,3))$			$q_3(\lambda_k(j,3))$		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
c_1	0.08	0.04	0.03	0.02	0.01	0.01	0.04	0.19	0.22
c_2	0.08	0.07	–	0.02	0.02	–	0.19	0.26	–
c_3	0.07	0.05	0.03	0.02	0.01	0.01	0.08	0.15	0.22
c_4	0.06	0.05	0.03	0.02	0.01	0.01	0.10	0.14	0.21
c_5	0.08	0.07	–	0.02	0.02	–	0.22	0.23	–
c_6	0.05	0.05	0.05	0.01	0.01	0.01	0.15	0.15	0.16

	$q_4(\lambda_k(j,3))$			$q_5(\lambda_k(j,3))$			$q_6(\lambda_k(j,3))$		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
c_1	0.06	0.04	0.03	3E-3	5E-4	4E-4	0.14	0.06	0.05
c_2	0.07	0.06	–	3E-3	2E-3	–	0.14	0.11	–
c_3	0.06	0.04	0.03	2E-3	1E-3	1E-3	0.12	0.08	0.04
c_4	0.05	0.05	0.03	2E-3	1E-3	7E-4	0.11	0.09	0.06
c_5	0.07	0.06	–	2E-3	2E-3	–	0.13	0.12	–
c_6	0.04	0.04	0.04	2E-3	1E-3	1E-3	0.09	0.08	0.08

The current judgement acceptabilities for each alternative can be displayed as a matrix. In Table 3.7, the current judgement acceptabilities are arranged around all judgement tasks for all c_j of a_3. Consequently, this table expresses the rank-1 performance dependency of each alternative a_i on respective judgements. The current judgement acceptabilities at each $\lambda_k(j, 3)$ correspond to the judgement inputs of the unified judgement matrix in the a_3 column of Table 3.3.

Thus, in row c_1 of Table 3.7, each of the three cells for the current judgement acceptability q_i corresponds to one of the input judgements $\lambda(1, 3) = \{1, 8, 9\}$. The corresponding current judgement acceptabilities at $q_3(\lambda(1, 3))$ are $\{0.04, 0.19, 0.22\}$. As a control mechanism, the sum of all $\sum_k q_3(\lambda_k(1, 3))$ is equal to the alternative a_3's rank-1 acceptability of 0.45. Slight differences might be caused by rounding.

It can be deduced that half of the instances that return a_3 as most-preferred contain the judgement $\lambda_3(1, 3) = 9$ (see Table 3.3). On the other hand, only 9% of the instances that return a_3 as most-preferred contain the judgement $\lambda_1(1, 3) = 1$. In contrast, the current judgement acceptabilities $q_3(6,3) = \{0.15, 0.15, 0.16\}$ represent approximately equal numbers of instances.

Table 3.8: Potential judgement acceptabilities for the judgement tasks (c_j, a_3)

	$\hat{q}_1(\lambda_k(j,3))$			$\hat{q}_2(\lambda_k(j,3))$			$\hat{q}_3(\lambda_k(j,3))$		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
c_1	0.23	0.12	0.09	0.06	0.03	0.02	0.11	0.57	0.65
c_2	0.17	0.13	–	0.04	0.03	–	0.37	0.52	–
c_3	0.19	0.15	0.10	0.05	0.03	0.02	0.23	0.44	0.66
c_4	0.19	0.15	0.10	0.05	0.03	0.02	0.28	0.42	0.63
c_5	0.15	0.14	–	0.04	0.03	–	0.43	0.46	–
c_6	0.15	0.15	0.14	0.03	0.03	0.03	0.43	0.44	0.46

	$\hat{q}_4(\lambda_k(j,3))$			$\hat{q}_5(\lambda_k(j,3))$			$\hat{q}_6(\lambda_k(j,3))$		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
c_1	0.19	0.11	0.09	0.01	2E-3	1E-4	0.41	0.18	0.15
c_2	0.14	0.11	–	0.01	3E-3	–	0.28	0.21	–
c_3	0.17	0.13	0.08	0.01	4E-3	3E-3	0.36	0.24	0.13
c_4	0.16	0.13	0.09	0.01	4E-3	2E-3	0.32	0.25	0.16
c_5	0.13	0.13	–	4E-3	4E-3	–	0.26	0.24	–
c_6	0.13	0.13	0.12	0.01	4E-3	4E-3	0.25	0.25	0.24

Potential judgement acceptabilities can be computed for all judgement tasks. They can be arranged as a matrix for each alternative a_i. However, Table 3.8 shows the potential judgement acceptabilities q_i for the judgement task at (c_j, a_3). The potential judgement acceptabilities q_i correspond to the judgement inputs of the unified judgement matrix in the column a_3 of Table 3.3.

Potential judgement acceptabilities show the acceptability an alternative would receive, if the group were to resolve a discrepancy to a particular judgement $\lambda_k(j,i)$. The highest potential judgement acceptability for this problem is 0.66 for alternative a_3, if the group resolved the judgement discrepancy at $\lambda(3,3)$ to $\lambda_3 = \{8\}$.

On the other hand, alternative a_6's highest achievable rank-1 acceptability would occur if the group resolved the discrepancy at $\lambda(1,3)$ to $\lambda_3 = \{1\}$, resulting in 0.41. It is interesting to note that the judgement discrepancy that results in the highest potential judgement acceptability for a_6 would decrease the potential judgement acceptability for a_3 to 0.11. Thus, potential judgement acceptabilities can be used to understand the effects resolutions would have on all alternatives.

Judgement sensitivities

Preference and judgement sensitivities can be computed for each discrepancy. This information can be used to distinguish high-sensitivity preference or judgement resolutions from low-sensitivity ones, serving the goal of identifying consensus-relevant input conflicts (Goal 4 on page 7).

Table 3.9 shows judgement sensitivities for the judgement discrepancy at $\lambda(1,3)$, computed according to Equation 3.20. The sensitivity is extremely high for the alternative a_3, with

Table 3.9: Judgement sensitivity for the judgement discrepancy at $\lambda(1,3)$

	a_1	a_2	a_3	a_4	a_5	a_6
δ_+	0.082	0.022	0.199	0.059	0.005	0.156
δ_-	-0.057	-0.015	-0.341	-0.041	-0.003	-0.101
$\sigma_i(1,3)$	0.082	0.022	-0.341	0.059	0.005	0.156

$\sigma_3(1,3) = -0.341$. This judgement discrepancy contains at least one resolution that has the potential to reduce a_3's acceptability to $0.451 - 0.341 = 0.11$: one quarter of its original rank-1 acceptability. And the sensitivity $\sigma_i(1,3)$ reveals that the alternative a_3 is most sensitive to this judgement discrepancy.

On the other hand, alternative a_3 may gain substantially from a different resolution of this discrepancy with its $\delta_+ = 0.199$. Only one other alternative, a_6, would experience significant changes from potential resolutions of this judgement discrepancy as well. For a_6, the potential gains are higher than the potential losses, with at least one resolution improving a_6's rank-1 acceptability by about two-thirds of its initial value. For all other alternatives, the sensitivity is low. Thus, a_3 and a_6 are highly sensitive to resolutions of $\lambda(1,3)$, while the rank-1 acceptabilities of other alternatives would not change much.

Table 3.10: Judgement sensitivity for the judgement discrepancy at $\lambda(4,5)$

	a_1	a_2	a_3	a_4	a_5	a_6
δ_+	–	0.000	–	–	0.007	0.001
δ_-	-0.005	-0.002	-0.010	-0.003	-0.004	-0.008
$\sigma_i(4,5)$	-0.005	-0.002	-0.010	-0.003	0.007	-0.008

An example of a judgement discrepancy with low judgement sensitivities is shown in Table 3.10 for the judgement discrepancy $\lambda(4,5)$. Here, the highest potential acceptability impact of this discrepancy is only in the second or third decimal place, which has almost no impact on the alternatives' rank-1 acceptabilities.

Active, inactive and pivot discrepancies

CMAA can help the facilitator to recognise the consensus-relevant impact of a discrepancy quickly, by categorising them into inactive, active and pivot discrepancies (see Section 3.3). The analysis for the illustrative example yields 41 active discrepancies, one of which is a pivot. No discrepancies are inactive.

Table 3.11 shows that the discrepancy $\lambda(4,5)$ is a pivot. It is a pivot, because alternative a_5 would be eliminated, if the group resolved the discrepancy to $\lambda_1(4,5) = 2$.

The distinction between active and inactive discrepancies provides a quick understanding of which discrepancies have an impact on consensus. Active discrepancies that contain a pivot show which preference or judgement resolution would exclude an alternative from reaching a rank-1 acceptability.

Table 3.11: The active discrepancy at $\lambda(4, 5)$, including a pivot for a_5

	$\lambda_1 = 2$	$\lambda_2 = 6$	$\lambda_3 = 8$
$\hat{q}_1$	0.148	0.147	0.144
$\hat{q}_2$	0.034	0.032	0.034
$\hat{q}_3$	0.446	0.441	0.442
$\hat{q}_4$	0.127	0.127	0.127
$\hat{q}_5$	**0.000**	0.001	0.011
$\hat{q}_6$	0.244	0.251	0.243

Concluding analysis

For the illustrative example decision, the analysis returns a_3 as the currently most-preferred alternative and a_6 as the second-most-preferred alternative. However, this result is not yet final; all potential discrepancy resolutions are active and may therefore cause a change in rank-1 acceptabilities of the alternatives. One active judgement discrepancy contains a pivot, whose resolution has the potential to exclude the alternative a_5 from the decision.

Discrepancies whose resolution would impact the rank-1 acceptability of the most-preferred alternative a_3 the most are $\lambda(1, 3)$, $\lambda(3, 3)$, and $\lambda(4, 3)$. Resolving these judgements to the smallest evaluation would be detrimental to the performance of a_3 and result in a change of the most-preferred alternative to a_6. No preference discrepancy holds much potential for changing the rank-1 acceptabilities.

CMAA's decision analysis can highlight preference and judgement discrepancies that have a strong impact on the preferredness of alternatives, along with the consequences of each potential resolution. This connection between input evaluations and output performance is unparalleled in decision analysis, offering insights for cooperative decisions and their decision structure, which answers research Question 3 on page 11. It is possible to deduce the effects certain input preferences or judgements have on the performance of alternatives, which contributes to Goal 4 on page 7.

3.6 Conclusion

The innovative paradigm on which CMAA is based is to consider all combinations of decision-maker evaluations. The CMAA analysis Algorithm 1 considers the complete combinatorial space and is able to measure the effect each input preference or judgement has on the performance of each alternative. This fulfils Requirement 1 on page 8 to preserve all input evaluations.

CMAA is independent of the decision method, thus providing a framework enabling each decision method's use in a group decision environment. The combinatorial space can be created for any type of input, and generating instances from this space is also universal, so instances can be generated in the format required by any decision method. CMAA only requires the ranking produced by the decision-method, which is independent of the input format. There are only two exceptions, as they use a different structure for the input data: AHP and an objective measurement and subjective evaluation approach. These require minor modifications to CMAA, which will be presented in Chapter 8. Thus, CMAA serves Goal 5 on page 8.

The variables help to track the overall performance of each alternative for all input combinations. They predict the performance of each alternative for any preference or judgement resolution the decision-makers might choose. These variables allow cause-and-effect analysis for each individual evaluation while considering every possible combination of the other evaluations. From this, the potential impact of each discrepancy on the performance of each alternative is known, serving Goal 4 on page 7. This level of analysis is unparalleled in multi-criteria decision-making and provides an answer to research Question 3 on page 11.

Because the combinatorial space quickly becomes intractable, Monte Carlo simulation can be used. It is assumed that a randomly independent sample of the combinatorial space will suffice to compute a sufficiently good approximation for all CMAA variables. This assumption will be studied in Chapter 7. If the necessary number of samples is sufficiently small, the CMAA analysis will require only a few seconds, and it can be used interactively during live decision-making.

4

Combinatorial Consensus Metric

This chapter introduces the new combinatorial consensus metric and its application, continuing the illustrative example from the previous chapter. It addresses Goal 3 on page 7 for providing a consensus metric that can measure the distance of the current state of the decision to a revelation-invariant consensus. This approach offers a unique method for measuring consensus, standing apart from existing input and output metrics. At the end of the chapter, the illustrative example from Chapter 3 will be continued. Part of the findings presented in this chapter were published in the articles Goers and Horton (2023a, 2024a).

4.1 Combinatorial Acceptability Entropy

In Section 2.5.2, it was shown that current input-based and output-based consensus metrics cannot detect a revelation-invariant consensus. But cooperative decisions require a revelation-invariant consensus (Definition 3 on page 4), so as to avoid announcing consensus prematurely (Definition 6 on page 29). Consequently, cooperative decisions require a new output-based consensus metric that fulfils these requirements.

The basic idea for the new metric is to use the CMAA analysis to assess the degree of consensus. For example, a hard consensus is reached, when for all instances, the same alternative achieves the first rank and no resolution of any discrepancy would change this result, in other words the consensus is revelation-invariant.

The new consensus metric is called the *Combinatorial Acceptability Entropy* (CAE) metric. The CAE metric measures two types of consensus. First, it delivers the consensus degree for the current decision $[\mathbf{P}; \mathbf{A}]$. This will be referred to as the *current entropy.* Second, it predicts the consensus degree that would be obtained from each available resolution of a preference discrepancy or a judgement discrepancy. This will be referred to as a *potential preference entropy* or a *potential judgement entropy*, respectively. This second consensus metric is computed from the reduced combinatorial space that is induced by a preference resolution $\mu_k(j)$ or by a judgement resolution $\lambda_k(j, i)$.

4.1.1 Consensus degree of the current decision

The current consensus degree depends on the rank-1 acceptabilities of the decision $[\mathbf{P}; \mathbf{A}]$. The closer the rank-1 acceptabilities are to a standard unit vector, the closer the decision is to consensus. In contrast, if all rank-1 acceptabilities have the same value $b_i^1 = \frac{1}{m}$, the further away the group is from reaching consensus. The Shannon information entropy of the rank-1 acceptabilities b_i^1 is given by

$$h = -\sum_{i=1}^{m} b_i^1 \cdot \log_2(b_i^1) \ . \tag{4.1}$$

It measures the similarity of the vector b^1 to a standard unit vector, thereby fulfulling the conditions for a consensus metric.

If there are instances that yield tied ranks, $\sum b_i^1 > 1$. In this case, the b_i^1 must be normalised before computing the entropy. The current entropy forms part of the output of the CMAA analysis in line 5 of Algorithm 1.

The entropy has its minimum value of 0 for a standard unit vector, i.e., when there is one alternative a_{i*} whose rank-1 acceptability is $b_{i*}^1 = 1$, and the rank-1 acceptabilities for all other alternatives is 0. This is a hard consensus, which is also revelation-invariant (Definition 3 on page 4). When all rank-1 acceptabilities have the same value $\frac{1}{m}$, the entropy has its maximum of $h_{max} = \log_2(\frac{1}{m})$. In Section 7.2.3, the relationship between entropy and rank-1 acceptabilities is studied.

In Section 2.5.2, an input metric CM_I and an output metric CM_O were shown and applied to the sample decisions in Table 2.6 on page 27. When the CAE metric is applied to the same decisions, it returns the current entropies 0.90, 1.00 and 1.00, reading from left to right and top to bottom. The maximum entropy for two alternatives is already 1.00. Thus, the decisions on the top-right and on the bottom-left are in full dissent, and the one on the top-left is close to a full dissent. The input metric CM_I signals a soft consensus for the decision in the top-right, and CM_O indicates a hard consensus for the decision on the top-left. Both indications are premature and clearly not revelation-invariant. This shows that the CAE metric is much more demanding than standard input and output metrics, because it can be used to recognise the different contributions of individual preferences and judgements on the ranking of alternatives. This prevents premature indication of consensus (Definition 6 on page 29).

4.1.2 Consensus degree for potential discrepancy resolutions

In order to guide a group efficiently to consensus, the discrepancy which would improve the consensus degree in their next consensus iteration is needed. For this, the potential preference and the potential judgement acceptabilities are used, which were introduced in Section 3.2.4. These are able to predict the rank-1 acceptability that an alternative would achieve for any preference or judgement resolution the group may agree on. Each potential acceptability is computed from a partial combinatorial space – either for the preference resolution $\mu_k(j)$ for the combinatorial space $[\mathbf{P}_j(k); \mathbf{A}]$, or for the judgement resolution $\lambda_k(j, i)$ for the combinatorial space $[\mathbf{P}; \mathbf{A}_{ji}(k)]$.

Instead of the current rank-1 acceptabilities b_i^1, the potential preference acceptabilities $\hat{y}_i(\mu_k(j))$ or the potential judgement acceptabilities $\hat{s}_i(\lambda(j, i2))$ are used in Equation 4.1. If their respective sums are greater than 1, their values must be normalised. The resulting *potential preference entropy* $\hat{h}(\mu_k(j))$ or *potential judgement entropy* $\hat{h}(\lambda_k(j,i2))$ reveal how close the respective rank-1 acceptabilities would be to a standard unit vector, if the corresponding resolution were to be chosen:

$$\hat{h}(\mu_k(j)) = -\sum_{i=1}^{m} \hat{y}_i\left(\mu_k\left(j\right)\right) \cdot \log_2\left(\hat{y}_i\left(\mu_k\left(j\right)\right)\right). \tag{4.2}$$

$$\hat{h}\left(\lambda_k\left(j,i2\right)\right) = -\sum_{i=1}^{m} \hat{q}_i\left(\lambda_k\left(j,i2\right)\right) \cdot \log_2\left(\hat{q}_i\left(\lambda_k\left(j,i2\right)\right)\right). \tag{4.3}$$

The potential preference entropies and the potential judgement entropies form part of the output of the CMAA analysis in line 5 of Algorithm 1.

Thus, the potential entropies reflect the impact each resolution would have on the consensus metric. In a consensus-building process, this information can be used to select the discrepancy whose resolution has the potential to improve consensus the most. An optimistic approach would select the discrepancy containing the resolution with the greatest improvement in the consensus metric, in the hope that this is the one that will be selected by the decision-makers. A more conservative approach might be to select a discrepancy that would definitely improve the consensus degree for all available resolutions, although the consensus progress might be smaller. A third approach would be to select the discrepancy for which the expected value of the potential entropy is smallest.

4.2 Interpretation of entropy and a soft consensus threshold

Entropy measures the similarity of a probability vector to a standard unit vector. In CMAA, this probability vector consists either of all alternatives' current rank-1 acceptabilities or of their potential rank-1 acceptabilities with respect to a particular discrepancy resolution. The interpretation of entropy as a consensus measure is not intuitive. In the following, entropy will interpreted visually, in order to motivate the choice of a threshold for terminating the consensus-building iteration.

Figure 4.1 shows six different rank-1 acceptability vectors and their corresponding entropies for five alternatives a_1 to a_5. The vertical axis shows the rank-1 acceptability index from 0.0 to 1.0. The horizontal axis shows the entropy, which descends from the maximum possible entropy of 2.32 for a vector of length 5 down to 0.1. The depth axis contains the five alternatives. The six rank-1 acceptability vectors have been chosen so that their entropies are equally spaced from 2.32 down to 0.10.

The vector corresponding to the maximum possible entropy of $h = 2.32$ contains five elements with the same value $b_i^1 = \frac{1}{5}$. On the opposite side of the diagram, the vector with entropy $h = 0.1$ is almost indistinguishable from the standard unit vector $(0, 0, 0, 1, 0)$, which would be achieved when $h = 0$.

For vectors with larger entropies, a small change in entropy corresponds to a significant qualitative change in the vector. At entropy $h = 2.32$, the elements of the vector are indistinguishable, but at $h = 1.87$, there is already a clear largest rank-1 acceptability for a_4. By contrast, the same reduction in entropy from $h = 0.54$ to $h = 0.10$ only corresponds to a small visual change in the acceptability vector. At $h = 0.10$, the rank-1 acceptability of b_4^1 has reached 0.99. This means that only 1% of the instances have a different most-preferred alternative.

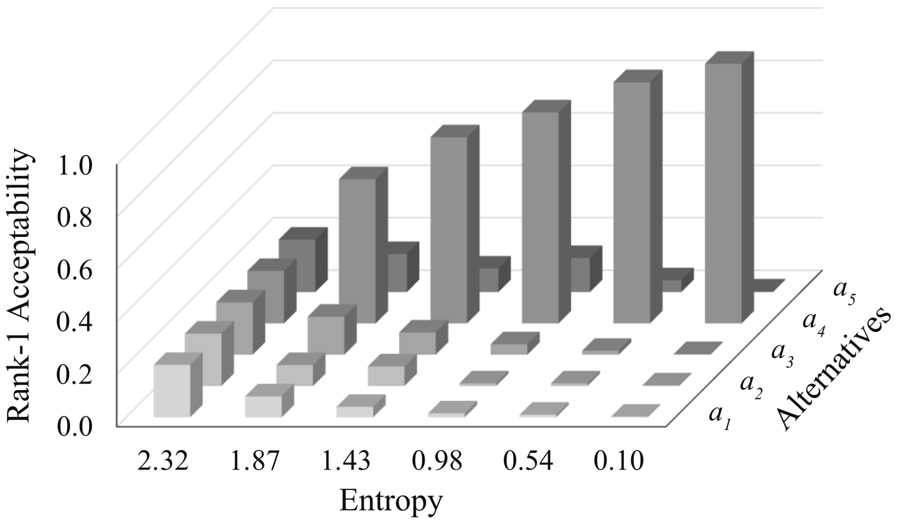


Figure 4.1: Entropies of six different acceptability vectors

At $h = 0.98$, the rank-1 acceptability of alternative a_4 is 0.81, which is a very strong recommendation for an alternative. We may conclude that an entropy threshold of $h < 1.0$ is sufficient to declare a very firm soft consensus on the most-preferred alternative. However, since the maximum entropy h_{max} grows slowly with the number of alternatives, a dynamic termination threshold h_{stop} can be defined:

$$h_{stop} \leq \tau \cdot h_{max} \ , \tag{4.4}$$

where τ is set to 0.3. In Section 7.2, the appropriateness of this threshold will be studied.

The early termination of consensus-building before $h = 0$ means the group has not yet achieved a revelation-invariant consensus. For a small fraction of combinations, one or more alternatives can still become the most-preferred. However, if the group is still in doubt about the result, they can continue consensus-building.

Note that, in exceptional circumstances, a decision may lead to two or more alternatives each receiving a rank-1 acceptability of $b^1_{i*} = 1$. In such cases, these alternatives cannot be excluded from the decision by any discrepancy resolution. If, at the same time, all other alternatives attain a rank-1 acceptability of $b^1_i = 0$, further consensus improvement is not possible. Even though the current entropy h did not reach 0, the group has achieved a hard consensus. In this case, all alternatives receiving $b^1_i = 1$ receive a recommendation.

4.3 Illustrative example

4.3.1 Applying the new CAE metric

In this section, the CAE metric is illustrated continuing the example decision from Section 3.5. The current entropy for the rank-1 acceptability vector (0.149, 0.034, 0.451, 0.129, 0.004, 0.250) in Table 3.4 is 2.00. The maximum entropy for six alternatives is 2.58. Thus, the current entropy signals only a modest degree of separation between the rank-1 acceptabilities.

Table 4.1 shows the potential preference entropies as per Equation 4.2 that correspond to each preference resolution $\mu_k(j)$ in Table 3.1. The preference entropies are computed

Table 4.1: Potential preference entropies

Criteria	μ_1	μ_2	μ_3
c_1	2.01	1.99	1.99
c_2	2.02	1.99	1.99
c_3	1.97	1.96	1.95
c_4	1.94	2.00	2.04
c_5	2.01	1.98	–
c_6	2.00	2.00	–

from the corresponding potential preference acceptabilities. The smallest entropy of 1.94 can be achieved if the group resolved the preference discrepancy at $\mu(4)$ to $\mu_1(4) = 3$. This preference resolution would result in the new rank-1 acceptabilities (0.14, 0.02, 0.42, 0.13, 5E-4, 0.29) for the six alternatives. All potential entropies are very close to 2.00, meaning that no preference resolution can significantly improve (or worsen) the current consensus degree.

Table 4.2: Potential judgement entropies

	a_1			a_2			a_3		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
c_1	1.96	2.00	2.02	1.90	1.98	2.07	2.12	1.76	1.56
c_2	1.85	1.92	1.98	1.91	1.98	2.07	2.09	1.88	–
c_3	1.95	1.97	2.02	1.91	2.06	–	2.15	2.00	1.55
c_4	1.83	1.96	1.98	1.87	1.92	2.11	2.17	2.02	1.63
c_5	1.96	1.99	2.02	1.93	1.98	2.06	2.02	1.97	–
c_6	1.99	2.00	–	1.99	2.01	–	2.02	2.00	1.97

	a_4			a_5			a_6		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
c_1	1.87	2.03	–	1.97	2.02	–	1.99	1.99	1.94
c_2	1.89	2.03	–	1.99	2.01	–	2.01	1.98	–
c_3	1.85	2.00	2.04	1.99	2.00	–	1.97	1.92	–
c_4	1,81	1.90	2.01	1.97	1.98	2.03	1.99	1.99	1.94
c_5	1.96	2.01	2.01	1.99	2.01	–	2.00	1.99	1.97
c_6	1.99	2.00	–	1.99	2.00	2.01		–	

Table 4.2 shows the potential judgement entropies as per Equation 4.3 that correspond to each judgement resolution $\lambda_k(j, i)$ in Table 3.3. The judgement $\lambda_3(3, 3) = 8$ promises the largest reduction in entropy to 1.55 in the next iteration. The potential judgement acceptability vector of this entropy is $q_i(\lambda_3(3, 3))$ =(0.10, 0.02, 0.66, 0.08, 3E-3, 0.13), which clearly positions a_3 as the most-preferred alternative. In contrast, resolving this judgement discrepancy to $\lambda_1(3, 3) = 2$ would increase the entropy to 2.15; consensus deteriorates, and the most-preferred alternative changes to a_6.

In Tables 4.1 and 4.2, two different potential entropy patterns within a discrepancy can be found:

1. The potential preference or judgement entropies of all available resolutions are smaller than the current entropy.
2. One or more resolutions hold the potential to reduce the current entropy, but others would increase it.

Most discrepancies belong to pattern 2. Additionally, three of them ($\lambda(1,3)$, $\lambda(3,3)$, $\lambda(4,3)$) contain the most significant entropy reduction and provide the chance to improve consensus the most. But they also contain resolutions with the largest entropy increase, which worsens consensus.

A resolution that increases entropy harms consensus, which may lengthen the consensus-building process and also have a negative psychological impact on the group. Thus, it might be beneficial to select a pattern 1 discrepancy for resolution, if available. This would guarantee an improvement in the consensus metric, even if it is only small. In the example, only six of the 41 preference and judgement discrepancies belong to pattern 1: $\mu(3)$, $\lambda(2,1)$, $\lambda(4,1)$, $\lambda(1,6)$, $\lambda(3,6)$ and $\lambda(4,6)$.

Table 4.3: Potential judgement acceptabilities for the judgement discrepancy $\lambda(3,6)$

	λ_1 $\{5\}$	λ_2 $\{8\}$
$\hat{q}_1$	0.18 ↑	0.12 ↓
$\hat{q}_2$	0.04 ↓	0.03 -
$\hat{q}_3$	0.50 ↑	0.39 ↓
$\hat{q}_4$	0.15 ↑	0.10 ↓
$\hat{q}_5$	0.01 ↑	3E-3 ↓
$\hat{q}_6$	0.13 ↓	0.36 ↑
$\hat{h}(\lambda_k(j,i))$	1.97 ↓	1.92 ↓

To demonstrate the effect on the potential judgement acceptabilities $\hat{q}_i$ for only a small but guaranteed improvement in entropy is shown in Table 4.3. It shows the potential judgement acceptabilities for the resolutions of the judgement discrepancy at $\lambda(3,6)$. The potential judgement acceptabilities are compared to their corresponding current rank-1 acceptabilities using arrows: an increase with ↑, a decrease with ↓, and an unchanged value with -. Both resolutions provide a small improvement in the entropy. λ_1 strengthens the leading alternative a_3, while λ_2 weakens it and simultaneously strengthens a_6, to almost the same level.

The CMAA analysis shows that the judgement discrepancy at $\lambda(3,3)$ holds the highest absolute potential for improving consensus, and that the largest guaranteed improvement is found at the judgement discrepancy $\lambda(4,1)$. Selecting the former for resolution represents an optimistic strategy, while selecting the latter is a more conservative approach. These examples show the potential that the CAE metric offers for guiding a consensus-building process.

4.3.2 Comparison with an input-based consensus metric

In this Section, the input-based consensus metric of Section 2.5.2 is applied to the illustrative example in order to determine which preference or judgement discrepancies it would recommend to the decision-makers for resolution. Then, a compromise resolution is applied, and the resulting values for both metrics are compared.

Table 4.4: Input-based preference and judgement distances for the illustrative example

	w_j	a_1	a_2	a_3	a_4	a_5	a_6
c_1	0.07	0.07	0.20	**0.37**	0.28	0.17	0.12
c_2	0.12	0.22	0.15	0.11	0.22	0.06	0.06
c_3	**0.37**	0.12	0.28	0.22	0.22	0.11	0.17
c_4	0.25	0.22	0.27	0.20	**0.30**	0.25	0.12
c_5	0.11	0.17	0.27	0.06	0.17	0.11	0.17
c_6	0.06	0.11	0.11	0.07	0.06	0.17	-

The input-based consensus according to Equation 2.2 is $CM_I = 0.19$. This value is closer to the (input-based) soft consensus threshold of 0.1 than it is to the maximum possible value of 0.5. By contrast, the CAE metric with a value of 2.00 is closer to its maximum value of 2.58 than to its termination threshold of 0.78. This indicates that the group is, in fact, further away from consensus than the CM_I metric suggests, and that it therefore might be considered to deliver too optimistic results.

Individual preference or judgement distance is the difference between an individual evaluation and the group average for that discrepancy. An overall distance measure for a discrepancy is obtained by summing distances for all individual discrepancies and dividing by the number of decision-makers. Table 4.4 shows the preference and judgement distances for the normalised preferences and judgements shown in Tables 3.1 and 3.3. The three largest preference and judgement distances are shown in bold typeface. An input metric-driven consensus-building process would recommend these discrepancies for resolution in descending order of distance.

Table 4.5: Input and CAE metric for three resolutions

Compromise resolution	CM_I	h
$\mu(3) = 6$	0.183	1.97
$\lambda(1,3) = 6$	0.183	2.04
$\lambda(4,4) = 4$	0.187	1.95

Table 4.5 shows the input metric CM_I as per Equation 2.2 and the entropy h as per Equation 4.1 that are achieved by the top-three discrepancy resolutions in Table 4.4. The entropy h was computed by applying a Monte Carlo simulation with $K_{MC} = 10,000$ and applying each resolution to the decision. The input metric CM_I improves minimally, as each resolution reduces the overall distance by a small amount. However, the CAE metric h exhibits no clear direction. This shows that the input-based consensus metric does not correlate well with the output metric, which is ultimately the one of interest.

The two judgement discrepancies in Table 4.4 also have the greatest potential to improve the CAE metric. On the other hand, the preference discrepancy $\mu(3)$, which has the highest input-based distance, ranks 21^{st} in consensus influence according to the CAE metric. In comparison, the discrepancy with the highest potential to reduce entropy is ranked 10^{th} by the input metric. This demonstrates that input metric-driven consensus-building does not necessarily drive the decision towards output consensus.

4.4 Conclusion

The CAE metric is unique in that it considers all possible combinations of evaluations and that it uses information entropy as a measure.

Two metric variables were introduced. The current entropy indicates the group's proximity to an overall consensus. The closer the rank-1 acceptabilities are to a standard unit vector, the lower the entropy. Thus, the CAE metric is capable of measuring the distance to a revelation-invariant consensus for any current state of a decision, fulfilling Goal 3. The potential preference and potential judgement entropies predict the impact each resolution of a preference or judgement discrepancy would have on the overall consensus. The latter values make it possible to understand the effects of discrepancy resolutions on output consensus and facilitate various consensus-building strategies.

An input-based consensus metric was applied to the illustrative example. Using the CAE metric as a baseline, it was shown that it is over-optimistic in its consensus estimate, and that improving input similarity does not necessarily have a positive effect on (output) consensus. The findings affirm that input-based consensus metrics are not able to support revelation-invariant consensus and are thus unsuitable for use in cooperative decisions. These results form part of the answer to Research Question 3 on page 11, which asks what insights into the decision the CMAA analysis can provide.

5

Combinatorial Consensus-Building

In this chapter, the CMAA consensus-building process is presented. It guides a group step-by-step to consensus while using the combinatorial analysis and consensus metric. This process should enable a group to achieve a revelation-invariant consensus, thereby it aims to fulfill Goal 1 on page 7 and prevent premature consensus (Definition 6).

The approach adopts an optimistic strategy, presenting discrepancies for resolution that have the potential to positively impact consensus the most. This leverages the analysis capability to distinguish discrepancies by their consensus impact. The combinatorial consensus-building process fulfills the requirement for unbiased discrepancy resolution (Requirement 2 on page 8).

At the end of the chapter, the illustrative example from Chapter 3 and Chapter 4 is continued. Some findings of this chapter were published in Goers and Horton (2023a).

5.1 Notation

The notation for a decision $[\mathbf{P};\mathbf{A}]$ is extended to $[\mathbf{P};\mathbf{A}]_s$, to denote the decision at consensus step s. The initial decision is defined to be $s = 0$.

When the decision-makers resolve a discrepancy with $\mu_{k1}(j)$ or $\lambda_{k1}(j,i)$, the other evaluations are overwritten:

$$\forall k2 : \mu_{k2}(j) \leftarrow \mu_{k1}(j) \ ,$$
$$\forall k2 : \lambda_{k2}(j,i) \leftarrow \lambda_{k1}(j,i) \ .$$

Each resolution during the consensus-building process creates an updated decision denoted by $[\mathbf{P};\mathbf{A}]_{s+1}$. The set of preferences or judgements is adapted towards the agreed resolution $\mu_k(j)$ or $\lambda_k(j,i)$:

$$[\mathbf{P};\mathbf{A}]_{s+1}, \text{ where } \mathbf{P}_j \leftarrow \mu_k(j) \text{ or}$$
$$[\mathbf{P};\mathbf{A}]_{s+1}, \text{ where } \mathbf{A}_{ji} \leftarrow \lambda_k(j,i) \ .$$

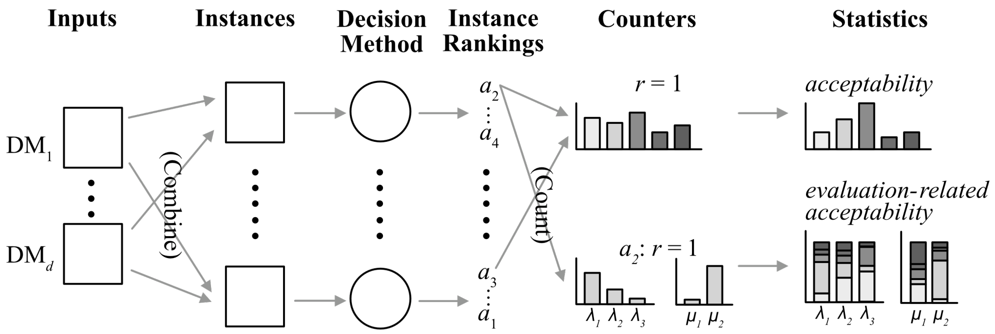


Figure 5.1: CMAA consensus-building model

The resolution of a preference or judgement discrepancy reduces the size of the combinatorial space:

$$K_{s+1} = \frac{K_s}{\phi} \ ,$$

where ϕ is the valency of the resolved preference discrepancy ϕ_j or judgement discrepancy by ϕ_{ji} (which is equal to 1 after the resolution has taken effect).

5.2 The CMAA consensus-building model

CMAA consensus-building is an iterative process of the type shown in Figure 2.2. Assuming a cooperative decision (see Section 2.3.2), at each iteration, the group creates a shared mental model of the issue, agrees on a resolution and moves one step closer to consensus. This agreement is achieved with a *clarification conference* − a discussion in which the decision-makers are asked to share the mental models that led to their evaluations.

For a cooperative decision, it is essential that the algorithm does not introduce a bias (Goal 2 on page 8). The CMAA algorithm merely presents a discrepancy to the decision-makers for discussion without any recommendation and accepts the resolved evaluation as-is. It therefore meets the requirement for being unbiased.

CMAA consensus-building begins with the output of the CMAA analysis in Algorithm 1. The group assigns a consensus threshold using Equation 4.4. Setting $\tau = 0$ corresponds to a hard consensus, while $\tau = 0.3$ corresponds to a 'very firm' consensus, as shown in Section 4.2. At the end of consensus-building, the algorithm returns the most-preferred alternative as its recommendation. In a cooperative decision, this choice is both correct, according to Definition 4 on page 5 and revelation-invariant (Definition 3 on page 4).

If the group is using a soft consensus metric with $\tau > 0$, the consensus achieved will not be revelation-invariant. However, it is very close to a revelation-invariant consensus, as long as $\tau \leq 0.3$, as was investigated in Chapter 7. The proportion of remaining instances that yield a different most-preferred alternative is very small. Therefore, a group may opt for a soft consensus, if time is limited and efficiency is prioritised over achieving a full revelation-invariant consensus.

A version of the consensus-building process containing CMAA-specific details is depicted in Figure 5.1. In each iteration, the following steps are carried out:

Step 1. *CAE consensus metric*: By applying the CMAA analysis in Algorithm 1 to the group's inputs, the current and potential entropies can be computed.

Step 2. *Consensus reached?*: If the consensus threshold $h \leq h_{stop}$ has been reached, the process terminates.

Step 3. *Locate discrepancy*: The preference or judgement discrepancy that contains the resolution providing the lowest potential entropy is identified.

Step 4. *Clarification conference*: The identified discrepancy is presented to the decision-makers for a clarification conference. In most cases, the group creates a shared mental model and a unanimous evaluation.

Step 5. *Adjust inputs*: The evaluations are modified to the agreed resolution of the discrepancy.

Repeat Repeat from Step 1.

End The decision reaches the consensus threshold and the consensus process ends. The most-preferred alternative is returned to the group, as well as the audit trail.

The identification of a discrepancy in Step 3 uses an optimistic heuristic. It selects the discrepancy that contains the resolution offering the lowest potential entropy, known as the *entropy-optimal* (EO) resolution. Usually, the other resolutions would increase entropy. Resolving a discrepancy may not lead to an improvement in consensus. Since it is unknown which resolution a group will choose, CMAA assumes that all resolutions of a discrepancy are equally probable.

It is important to note that during a clarification conference (Step 4), revealing what entropy would be attained by each resolution to the group may introduce a bias. Decision-makers who are keen to finish the decision quickly may be tempted to over-value the arguments that are made for it, leading to an incorrect resolution. This might distort the rank-1 acceptabilities or even eliminate an alternative. Therefore, revealing the entropy-optimal resolution when asking the group to resolve a discrepancy should be avoided. Instead, the clarification conference should present only the discrepancy to be resolved, asking the decision-makers to share their mental models and agree on a resolution.

If a group is unable to resolve a discrepancy, the algorithm returns to step 3, where it selects the next-best discrepancy and excludes the unresolved discrepancy from selection in the future.

5.3 The consensus-building algorithm

The CMAA consensus-building algorithm is given by Algorithm 2. It is preceded by Algorithm 1, which computes the variables and consensus metrics for the decision $[\mathbf{P}; \mathbf{A}]_0$. Then the algorithm is described as follows:

Given The group determines the consensus threshold h_{stop} they wish to use.

Input As input for consensus-building serves the output from the CMAA analysis Algorithm 1.

line 1 Initialises the consensus step counter s to 0.

lines 2-8 Iteration that repeats until the current of the decision $[\mathbf{P}; \mathbf{A}]s$ meets the consensus threshold h_{stop}.

Algorithm 2: CMAA consensus iteration

```
   Given: h_stop
   Input: Output from Algorithm 1 as [P; A]_0
1  s = 0;
2  repeat
3      select the judgement or preference discrepancy containing the largest potential
         entropy reduction;
4      propose a clarification conference for this discrepancy to the decision-makers;
5      s++;
6      create a new decision [P; A]_s by applying the resolved discrepancy;
7      recompute all variables with Algorithm 1 for the decision [P; A]_s;
8  until h ≤ h_stop or user-stop;
   Output: Recommendation for choice in consensus, statistics and audit trail
```

line 3 Identifies the lowest entropy among all potential preference and judgement entropies and selects the corresponding discrepancy.

line 4 Proposes the selected discrepancy to the group for a clarification conference, assuming they will reach an agreement on a resolution.

line 5 Marks a new consensus step and increments the step counter s by one.

line 6 Updates the decision $[\mathbf{P}; \mathbf{A}]_s$ according to the resolution agreed upon by the group in line 4.

line 7 Applies the updated decision to Algorithm 1 to compute all variables, including the new current entropy h.

line 8 Repeats the loop until the consensus threshold h_{stop} is reached, or the user terminates the consensus-building process.

Output Computes a recommendation for the group's choice based on the acceptability indices, along with other statistics and an audit trail. The audit trail records all resolutions agreed upon by the group during the consensus-building process, facilitating traceability for decision revision or learning purposes.

Even if the decision-makers opt for the EO-resolution at every step, it may not necessarily result in the fastest possible consensus path. The optimistic strategy employed here is greedy: it predicts the entropy for only one future consensus step. Consequently, it can only identify a local optimum for each step, which may not lie on the globally optimal consensus path. An example decision in which the EO consensus path is not optimal can be found in Goers and Horton (2023a). The efficiency of this optimistic strategy will be explored in Section 7.4.3.

5.4 Illustrative example

The CMAA consensus-building efficiency is demonstrated using an illustrative example. It shows both an entropy-optimal consensus path and a more realistic consensus path. Additionally, the development of different types of discrepancies (see Section 3.3) along the more realistic consensus path is illustrated. The inputs for this illustration are the same as those shown in Tables 3.1 and 3.3 of Section 3.5. The Monte Carlo simulation at each step uses $K_{MC} = 10,000$ samples. The soft consensus threshold for $\tau = 0.3$ is at $h_{stop} = 0.78$ as per Equation 4.4. The initial entropy is 2.0, which indicates that the group has not yet reached soft consensus.

5.4.1 Entropy-optimal consensus path

Table 5.1: Consensus path with EO-resolutions

Step	h	Actual Res.	b_1^1	b_2^1	b_3^1	b_4^1	b_5^1	b_6^1	Potential Res.	$\hat{h}$
0	2.00	–	0.15	0.03	0.45	0.13	4E-3	0.25	$\lambda(3,3) = 8$	1.55
1	1.55	{8}	0.10	0.02	0.67	0.08	3E-3	0.13	$\lambda(1,3) = 9$	0.66
2	0.65	{9}	0.04	0.01	0.91	0.03	4E-4	0.02	$\lambda(4,3) = 9$	0.10
3	0.09	{9}	0.01	1E-5	0.99	3E-3	0.00	0.00	$\mu(3) = 9$	0.01
4	0.01	{9}	8E-5	0.00	1.00	9E-4	0.00	0.00	$\mu(2) = 5$	0.00
5	0.00	{5}	0.00	0.00	1.00	0.00	0.00	0.00	–	–

Table 5.1 shows the consensus steps when all discrepancies are resolved in an entropy-optimal manner. Each row represents a consensus step, starting with the initial step 0. At each step, the current entropy h is presented. Each row also shows the actual resolution chosen by the decision-makers, which is always entropy-optimal for this consensus path. The rank-1 acceptabilities b_i^1 for the six alternatives are shown. The lowest next-step entropy $\hat{h}$ is given, as well as the resolution that would lead to it. A horizontal dotted line in the table marks the step at which the CMAA consensus-building stops if $\tau = 0.3$ is used.

Only two steps are needed when a soft consensus threshold at $\tau = 0.3$ is applied. The largest reduction in entropy is achieved in step 2, when $\lambda(1,3)$ is resolved to the judgement 9. At this point, alternative a_3 attains a rank-1 acceptability of 0.91, indicating a very firm soft consensus. Although none of the other alternatives have been eliminated yet, their rank-1 acceptabilities are very small, all being ≤ 0.04. It takes three more steps to definitively exclude these alternatives.

The impact on the CAE metric h varies significantly with each resolution. The first three resolutions greatly impact the consensus level, whereas the preference resolutions $\mu(3) = 9$ and $\mu(2) = 5$ have only minor impacts, mainly serving to exclude certain alternatives entirely.

The hard CAE-based consensus is revelation-invariant: no further resolution can change the rank-1 acceptabilities, because all instances that remain in the decision return alternative a_3 as most-preferred.

Table 5.2 shows the rank acceptability indices and their Olympic holistic acceptabilities acc_i as per Equation 3.8 at step 2 of the entropy-optimal consensus path. The holistic acceptabilities can help determine how close other alternatives are to the currently most-preferred alternative. The group might use this information to either approve the soft consensus or continue the consensus path to minimise any chance of the outcome changing. The holistic acceptability indicates a strong preferredness for a_3. The closest alternative is a_6 with $acc_6 = 0.182$, which is less than one-third of a_3's acceptability, supporting the firm soft consensus.

Table 5.2: Rank Acceptability Indices at step 2 (very firm soft consensus)

	b_i^1	b_i^2	b_i^3	b_i^4	b_i^5	b_i^6	acc_i
a_1	0.042	0.222	0.288	0.265	0.138	0.045	0.121
a_2	0.006	0.062	0.110	0.177	0.267	0.378	0.033
a_3	0.906	0.082	0.011	0.001	9E-5	1E-5	0.569
a_4	0.035	0.175	0.205	0.235	0.199	0.151	0.094
a_5	4E-4	0.014	0.060	0.171	0.352	0.404	0.010
a_6	0.020	0.456	0.329	0.149	0.040	0.006	0.182

Reaching consensus in just two steps, or even in five steps, is extremely efficient. However, this path is idealised, because only EO-resolutions are chosen. In the following section, the effects of a more realistic mix of entropy-optimal and non-entropy-optimal resolutions on consensus efficiency will be illustrated.

5.4.2 Realistic consensus path

Table 5.3: Consensus path with random resolutions

Step	h	Actual Res.	b_1^1	b_2^1	b_3^1	b_4^1	b_5^1	b_6^1	Potential Res.	$\hat{h}$
0	2.00	–	0.15	0.03	0.45	0.13	4E-3	0.25	$\lambda(3,3)=8$	1.55
1	2.15	$\overline{EO}:\{2\}$	0.20	0.05	0.23	0.17	0.01	0.37	$\lambda(3,6)=8$	1.92
2	2.25	$\overline{EO}:\{5\}$	0.25	0.06	0.28	0.21	0.01	0.21	$\lambda(4,3)=9$	1.97
3	2.22	$\overline{EO}:\{4\}$	0.30	0.08	0.11	0.24	0.01	0.28	$\lambda(2,1)=8$	1.79
4	1.80	$EO:\{8\}$	0.57	0.05	0.05	0.18	5E-3	0.16	$\lambda(4,1)=8$	0.76
5	0.77	$EO:\{8\}$	0.87	0.02	3E-3	0.09	0.00	0.02	$\mu(3)=1$	0.30
6	0.30	$EO:\{1\}$	0.96	3E-3	0.01	0.03	0.00	2E-4	$\lambda(5,1)=7$	0.08
7	0.08	$EO:\{7\}$	0.99	1E-4	6E-5	0.01	0.00	0.00	$\lambda(1,1)=9$	0.00
8	0.00	$EO:\{9\}$	1.00	0.00	0.00	0.00	0.00	0.00	–	–

Since it is unknown how the group will resolve a discrepancy, it is more realistic to assume that the consensus path includes both entropy-optimal and non-entropy-optimal resolutions. Table 5.3 presents an example for such a consensus path for the illustrative decision. The table is constructed similarly to Table 5.1. Each actual resolution is marked as either EO for entropy-optimal or $\overline{EO}$ for non-entropy-optimal.

In this scenario, a hard consensus is achieved in eight steps. The first three steps were resolved using a non-entropy-optimal resolution, while the other five were resolved using EO-resolutions. A soft consensus is achieved in five steps, marked by the horizontal dotted line in the Table 5.3, with a_1 as the strongest alternative, having a rank-1 acceptability of 0.87. After eight steps, a hard consensus is reached with a_1 as the most-preferred alternative. Interestingly, this consensus path is three steps longer than the entropy-optimal path, which corresponds to the number of non-entropy-optimal resolutions. It could be hypothesised that each non-entropy-optimal resolution extends the entropy-optimal path by one consensus step. The effect of non-entropy-optimal resolutions on the consensus path length will be studied in Section 7.4.3.

The initial ranking of alternatives was $a_3 \succ a_6 \succ a_1 \succ a_4 \succ a_2 \succ a_5$. This order changes in step 1 with the resolution of $\lambda(3,3)$ to the judgement 2, making a_6 the most-preferred alternative. This resolution more than halves a_3's rank-1 acceptability, while concurrently strengthening a_6's rank-1 acceptability, causing the current entropy to increase to 2.15. In step 2, the current entropy increases again to 2.25, and the order of alternatives regarding their rank-1 acceptabilities changes once more, with a_3 becoming the most-preferred alternative. The next consensus step slightly improves the current entropy to 2.22, despite the judgement resolution being non-entropy-optimal. The rank-1 acceptability order changes again, with a_1 emerging as the most-preferred alternative. Subsequently, a_1 remained the most-preferred alternative until a soft consensus was reached. Consensus steps 6 to 8 affirmed the strength of alternative a_1 until a hard consensus was achieved.

5.4.3 Development of active, inactive and pivot discrepancies

In Section 3.3, active, inactive and pivot discrepancies were introduced. These provide additional analytical information about the decision for the facilitator of the group. Active discrepancies are those whose resolution affects the rank-1 acceptabilities, while resolving an inactive discrepancy has no effect. Pivot discrepancies are active discrepancies that contain a resolution that would eliminate one or more alternatives from the decision.

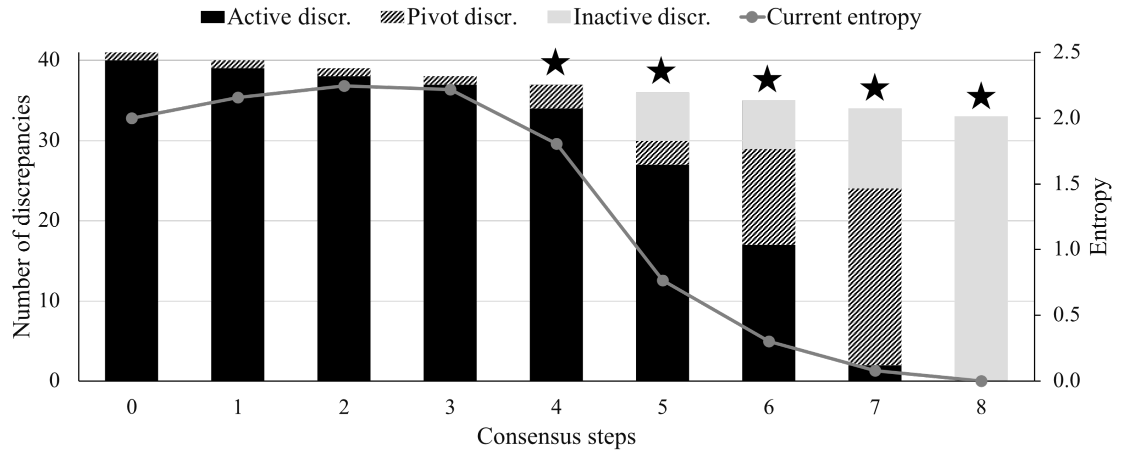


Figure 5.2: Development of discrepancy-types throughout consensus-building

Figure 5.2 illustrates the development of the three discrepancy variants for the more realistic consensus path. Each column represents the total number of remaining discrepancies and the distribution of the three discrepancy variants. Columns with a star over them mark EO-resolutions. The line graph shows the current entropy at each consensus step, with its scale displayed on the secondary vertical axis.

At each step, the number of discrepancies is reduced by one. Initially, all discrepancies are active, one of which is a pivot. Upon the elimination of alternative a_5 at step 5, all six a_5-related judgement discrepancies become inactive. Similarly, at step 7, with the removal of alternative a_6, the remaining four a_6-related judgement discrepancies become inactive. When hard consensus is achieved, the remaining 33 preference and judgement discrepancies have become inactive.

The eight discrepancy resolutions reduced the size of the combinatorial space from about $5.6 \cdot 10^{16}$ instances to $K = (1^1 \cdot 2^2 \cdot 3^3) \cdot (1^8 \cdot 2^{13} \cdot 3^{15}) = 1.3 \cdot 10^{13}$ instances, as described in Equation 3.3. The final space represents only 0.02% of the original space, but is still extremely large. Despite the considerable size of the remaining space, all instances now place alternative a_1 at the first rank. This illustrates the large amount of Slack (Definition 5 on page 10) that can be contained in a multi-criteria group decision. This is one of the key attributes of CMAA that enables it to reach consensus so efficiently.

5.4.4 Comparison to optimal distance-driven consensus path

In this section, the optimal consensus path driven by the CAE metric is compared to the optimal consensus path driven by an input metric. The distance driven consensus-building approach used here is MACRP 3 (Multi-Attribute Consensus-Reaching Process) by Zhang et al. (2019). MACRP 3 is similar to CMAA in that it selects one preference or judgement discrepancy for resolution. The difference is that MACRP 3 applies averaging to the evaluations, but CMAA lets the decision-makers resolve to one of the discrepancy's input evaluations. In the original study (Zhang et al., 2019), MACRP 3 was only tested with judgement discrepancies.

Consensus-building using MACRP 3 consists of the following steps:

1. The input metric used is given in Equation 2.2.
2. Zhang et al. (2019) use a consensus threshold of $CM_I < 0.1$.
3. The identification rule selects a judgement discrepancy $\lambda(j, i)$ (and here MACRP 3 is extended to include preference discrepancy $\mu(j)$ as well) containing the evaluation that is the most distant from the group's average:

$$\mu(j^*) = \underset{j,k}{\operatorname{argmax}}(|\mu_k(j) - \overline{\mu(j)}|),$$
$$\lambda(j^*, i^*) = \underset{i,j,k}{\operatorname{argmax}}(|\lambda_k(j, i) - \overline{\lambda(j, i)}|).$$

4. The direction rule requires that the adjusted judgement for DM_k should lie between the initial preference $\mu_k(j)$ or judgement $\lambda_k(j, i)$ and its corresponding group average $\overline{\mu(j)}$ or $\overline{\lambda(j, i)}$. The group average represent the ideal resolution in MACRP 3.

Table 5.4 shows the MACRP 3 consensus path. Each row represents a consensus step, where the optimal compromise resolution was selected. For each step, the current entropy h (Equation 4.1), the input metric CM_I (Equation 2.2), the actual compromise resolution of a discrepancy, and the CMAA rank-1 acceptabilities were computed. CMAA variables were computed using a Monte Carlo simulation with $K_{MC} = 10,000$ instances. The discrepancies with the greatest distances are resolved step-by-step to the group average until the soft consensus threshold was met.

MACRP 3 reaches a soft consensus after 13 optimal resolutions. In contrast, although the CAE metric exhibits some improvement, it still cannot achieve a soft consensus. Upon closer examination of the rank-1 acceptabilities, the current most-preferred alternative is a_3, receiving a rank-1 acceptability of 0.51. This implies that a_3 emerges as the top choice in only 51% of all instances. Conversely, in 49% of all instances within the combinatorial space, a_3 is not the preferred alternative, with other alternatives taking precedence. This fragility underscores the input metric's sensitivity to potential preference and judgement resolutions, rendering the CAE metric considerably more demanding as a consensus metric compared to input metrics.

Table 5.4: Distance-driven consensus path with optimal compromise resolutions

Step	h	CM_I	Actual Resolution	b_1^1	b_2^1	b_3^1	b_4^1	b_5^1	b_6^1
0	2.00	0.194	–	0.15	0.03	0.45	0.13	4E-3	0.25
1	2.04	0.183	$\lambda(1,3) = 6.0$	0.16	0.04	0.42	0.14	4E-3	0.26
2	2.02	0.173	$\mu(3) = 6.0$	0.16	0.03	0.43	0.14	2E-3	0.25
3	1.96	0.165	$\lambda(4,4) = 4.0$	0.17	0.03	0.45	0.09	2E-3	0.27
4	1.89	0.157	$\lambda(1,4) = 4.5$	0.18	0.04	0.47	0.05	2E-3	0.28
5	1.84	0.149	$\lambda(3,2) = 3.5$	0.18	0.02	0.47	0.06	2E-3	0.29
6	1.76	0.142	$\lambda(4,2) = 4.\bar{3}$	0.18	3E-3	0.48	0.06	3E-3	0.29
7	1.74	0.134	$\lambda(5,2) = 5.\bar{3}$	0.18	1E-3	0.48	0.06	2E-3	0.29
8	1.72	0.127	$\mu(4) = 6.\bar{3}$	0.18	1E-3	0.48	0.06	1E-3	0.29
9	1.71	0.121	$\lambda(4,5) = 4.5$	0.18	8E-4	0.48	0.06	6E-5	0.29
10	1.66	0.114	$\lambda(2,1) = 5.0$	0.14	8E-4	0.50	0.06	0.00	0.31
11	1.54	0.108	$\lambda(2,4) = 7.0$	0.14	8E-4	0.52	0.02	6E-5	0.32
12	1.52	0.102	$\lambda(3,3) = 5.0$	0.15	3E-4	0.51	0.01	0.00	0.33
13	1.52	0.096	$\lambda(3,4) = 6.\bar{3}$	0.15	2E-4	0.51	0.01	0.00	0.33

The optimally executed MACRP 3 algorithm took almost six times as long to reach a soft consensus as CMAA using EO-resolutions. CMAA achieved a very firm soft consensus for its EO consensus path, whereas MACRP 3's soft consensus is still far from reaching a revelation-invariant consensus (Definition 3). Hence, CMAA's consensus-building represents a transformative advancement for cooperative decisions, whose performance will be studied in simulations in Chapter 7.

5.5 Conclusion

The CMAA consensus-building Algorithm 2 is based on the CMAA analysis algorithm and its CAE output metric. The CAE metric determines the current consensus level and efficiently guides the consensus-building process using an optimistic strategy. Each consensus step is conducted through an unbiased clarification conference.

The CAE metric is very strict; once a hard consensus is achieved, no other resolution can change the consensus result. The expectation is that most discrepancies will not require resolution, yet the group will still reach a hard consensus. In the illustrative example, it was shown that with just eight consensus steps in a realistic scenario, a hard consensus can be achieved. The combinatorial space at hard consensus contained more than 10^{13} instances, all of which returned the same most-preferred alternative. This represents a very large amount of Slack inherent in the combinatorial space, supporting Hypothesis 1 on page 10.

Unlike the CAE metric, the standard input-based consensus metric cannot identify a revelation-invariant consensus. In the illustrative example, guidance based on the largest distance discrepancy results in a consensus path that is six times longer than that of CMAA. Therefore, CMAA represents a significant improvement in the computer support for cooperative decisions, both in terms of quality and efficiency. The extent of the differences between consensus-building using the CAE metric and input-based measures will be studied in Chapter 7.

CMAA achieves Goals 1 to 5 and serves as a suitable framework for competitive decisions.

Goal 1: It enables a revelation-invariant consensus-building (Goal 1 on page 7), while preventing to announce a premature consensus (Definition 6 on page 29).

Goal 2: The consensus-building is efficient (Goal 2 on page 7) without introducing a bias (Requirement 2 on page 8) and preserving all inputs as-is (Requirement 1 on page 8).

Goal 3: The efficient consensus-building is possible due to using a new CAE consensus metric, capable of detecting discrepancies containing the largest consensus improvement (Goal 3 on page 7).

Goal 4: By computing the potential preference or judgement entropy, CMAA consensus-building can distinguish between more and less consensus-relevant discrepancies (Goal 4 on page 7).

Goal 5: The framework can be applied to any multi-criteria decision-making method, as it is independent of the calculation of performance and does not rely on various input preferences or judgements (Goal 5 on page 8).

The CMAA framework will be verified and validated in the Chapters 7, 8 and 9

6
Ideas for Digital Facilitation

The preceding three chapters introduced the new CMAA framework. In this chapter, the focus is on outlining ideas about implementing this framework as either a digital assistant for human facilitators or as an automated facilitation tool. The ideas address Goal 6 on page 9. Some of these concepts were published in Goers et al. (2024).

6.1 Application scenario

CMAA has the potential to facilitate group decision-making processes. A software tool that implements Algorithms 1 and 2 can be extended to become a digital assistant for a human facilitator or even a fully automated digital facilitator of the group decision-making process. This would enable the group to focus on developing their shared mental models without the need for in-depth knowledge of the algorithms.

Two application scenarios are possible:

1: *Facilitator support*: In this case, the human facilitator needs an understanding of the CMAA analysis, along with procedural guidance and recommendations for carrying out their role.

2: *Automated facilitation*: The group relies on a digital facilitator to steer them through the consensus-building process. In this case, the digital guidance must be straightforward to ensure easy comprehension and implementation by the group.

The latter scenario is of interest, because facilitators can be expensive and may not be readily accessible for the group. Consequently, groups need to be able to carry out consensus-building on their own. However, without a structured facilitation process, a group of decision-makers may struggle to achieve the desired decision quality (Briggs et al., 2006).

Modern software development follows agile principles, such as Scrum (Schwaber, 1997). This approach involves delivering software iteratively and incrementally, with continuous user feedback to enhance the value it delivers. The first of the twelve principles outlined in the Agile Manifesto (Beck et al., 2001) states:

> *Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.*

In Scrum, user stories drive the development process (Jeffries, 2008). User stories encapsulate a user's expectations and needs for which a product or feature will be developed. Typically, the user story structure follows the format "As a ... I want ... so that ...". The first gap denotes the user, the second articulates the user's need, and the last conveys the benefit the user will gain from the feature.

To enhance the understanding of a user story, software development teams use a user persona tool (Blomkvist, 2002). In the upcoming sections, both a user persona and a user story will be outlined for each of the two scenarios. In scenario 1, the user is a 'facilitator'; in scenario 2, the user is a 'decision-maker'.

Excluded from the following considerations are the group's knowledge about the decision method. It is also assumed that the group starts with a well defined decision, consisting of alternatives, criteria and that all decision-makers to take part in the decision are identified. However, in any complete software scoping, these aspects need still to be considered.

6.1.1 Facilitator: user persona and user story

The facilitator's user persona is outlined by describing the role, skills, goals, and specific needs for facilitating consensus-building with CMAA. After outlining the user persona, a user story is briefly described.

A partial user persona:

- *Role*: Facilitators are typically responsible for guiding groups in problem-solving, decision-making, or other collaborative tasks. This can be achieved through face-to-face meetings or online conferences using supporting software tools such as digital whiteboards.
- *Skill Set (general)*: Facilitators employ established methods to execute collaborative tasks, such as idea generation, task list organisation, project kick-offs, strategic planning, and consensus-building efforts (Knoll et al., 2010). These tasks can include generating ideas, organising a list of tasks, kicking-off a new project, building strategies or guiding a group towards consensus (Briggs et al., 2006).
- *Skill Set (advanced)*: Proficient facilitators can adeptly navigate group dynamics, address obstacles as they arise, manage resistance, and identify opportunities for progress.
- *Goal*: The facilitator's primary objective in a cooperative group decision is to facilitate consensus and resolve conflicts hindering full agreement. This fosters the development of a shared mental model (Briggs et al., 2005) within the group, increasing the likelihood of successful decision implementation (Moscovici and Doise, 1994).
- *Specific needs*: Effective guidance in the consensus-building process requires facilitators to have insights into the current status of the decision and into possible future paths, enabling them to adjust their approach to suit the group dynamics.

An illustrative user story for a proficient facilitator could be based on: "As a proficient facilitator, I seek an overview of the group's consensus progress and of important conflicts that could affect progress." Section 6.2.5 will show a mock-up for this user story. Additional user stories can be crafted to address various aspects of the facilitation process.

6.1.2 Decision-maker: user persona and user story

In cooperative decisions, groups are formed with the expectation that their collective expertise will enhance decision quality and foster commitment.

A user persona is briefly outlined for the user 'decision-maker', encompassing their role, goals, skill set, and specific requirements in a digital consensus-building setting. Additionally, a user story for decision-makers is outlined.

A partial user persona:

- *Role*: As decision-makers, they contribute their individual knowledge and expertise to the group decision.
- *Skill Set*: Decision-makers contribute by sharing knowledge and experience that is relevant to the decision. They typically lack the experience that is required for facilitating complex collaborative tasks.
- *Goal*: The primary objective of decision-makers within the group is to transition from a set of alternatives to a collectively agreed-upon decision, leveraging their combined experience to make the right choice.
- *Specific Needs*: Due to their limited knowledge of the consensus-building algorithm and of facilitation skills, decision-makers need a systematic guide to navigate and accomplish all collaborative tasks leading to a consensus decision.

Hence, a user story for a decision-maker could be based on: "As a decision-maker, I seek a digital facilitator to guide our diverse group through consensus-building, so that we can agree on an alternative." Section 6.2.6 will show a mock-up for an implementation of this user story.

6.2 User interactions and mock-ups for digital facilitation

Before detailing the user interactions, an analysis of the consensus-building process with the CMAA framework is conducted. Subsequently, the primary user interactions for the two application scenarios are outlined, accompanied by mock-ups that depict concepts for developing the corresponding software tools; starting with the respective user story.

6.2.1 Analysis of the CMAA framework for user interactions

Figure 6.1 represents the consensus-building process of CMAA from a collaborative standpoint. The outlined steps are distinguished by different backgrounds and frames, denoting their operational modes: automated computer processing (white background with a solid frame), user input tasks (dotted frame), and facilitator instructions (gray background).

The analysis of the consensus-building steps is as follows:

Input Initially, the group provides their inputs. Decision-makers individually offer their preferences and judgements according to the decision model's guidelines.

Step 1. Upon input completion, the CMAA analysis using Algorithm 1 is executed to compute statistics and the CAE metric.

Step 2. If consensus is achieved, the process advances to the End step; otherwise, it proceeds to Step 3.

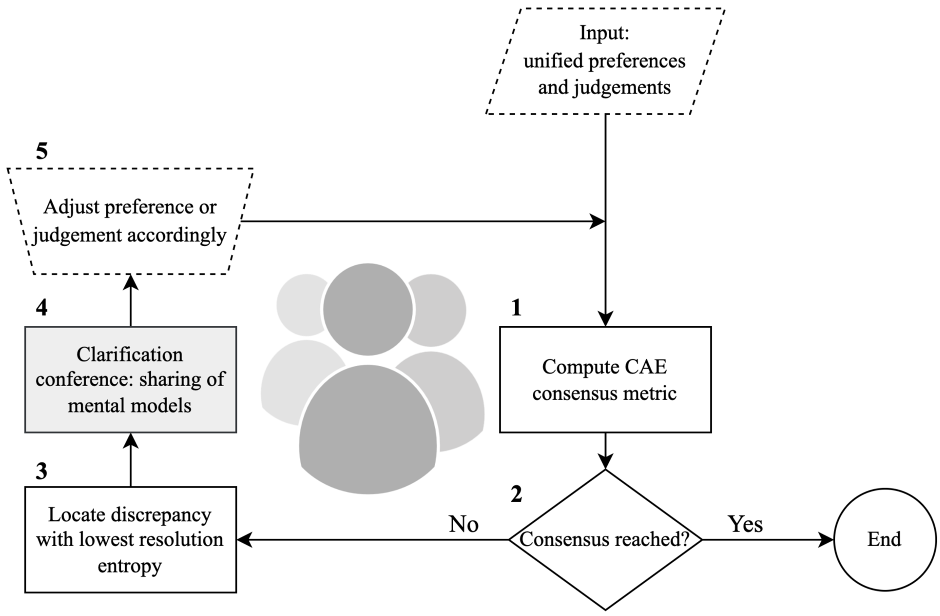


Figure 6.1: User interactions in CMAA consensus-building

Step 3. CMAA's Algorithm 2 identifies the discrepancy with the lowest entropy for resolution.

Step 4. A clarification conference is initiated, in which decision-makers share their mental models and resolve the identified discrepancy.

Step 5. The resolution is provided by the decision-makers and the decision is updated accordingly.

Repeat The consensus-building process loops back to Step 1, iterating until consensus is obtained or the group decides to halt the process.

End Upon achieving consensus in Step 2, the results are visualised in individual reports for each decision-maker, along with an audit trail.

During each clarification conference, the group shares their mental models. The formalisation of this conference is simple, requiring only a clarification and an agreement. The facilitator must take care not to introduce a bias into the clarification conference.

6.2.2 Outline of user interactions

The analysis of the CMAA consensus-building framework has identified the essential user interaction steps. In this section, these user interactions are briefly mapped to the human-computer interactions that are required for consensus-building.

Figure 6.2 illustrates the interactions for both scenarios. The flowchart visualises the interactions between the decision-makers (left column), the (digital) facilitator (centre column), and the CMAA algorithm (right column). The facilitation can be performed either by a human facilitator or by a machine.

The process commences with the decision-makers at the dashed rhomboid (top-left), titled 'Input: preferences and judgements' at step $s = 0$. Here, decision-makers provide

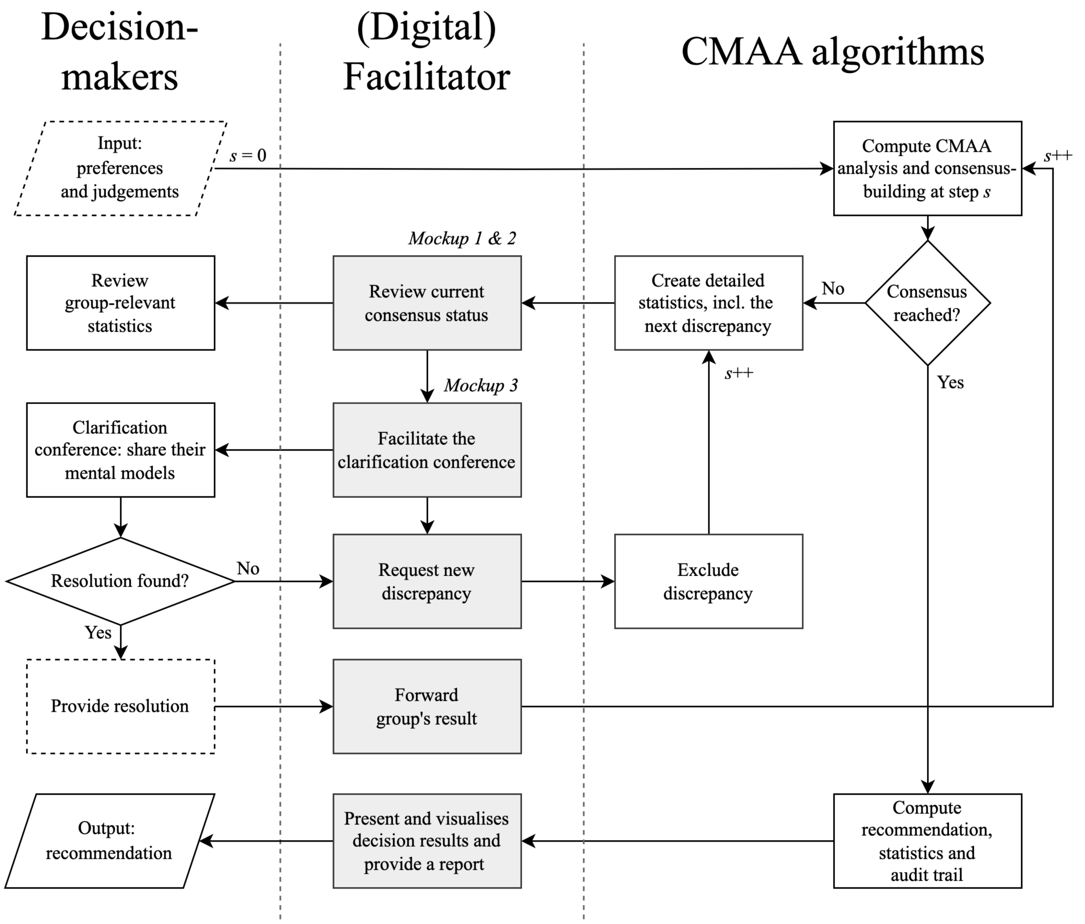


Figure 6.2: User interactions

their individual input preferences and judgements. This input serves as the basis for CMAA analysis Algorithm 1. Subsequently, the CMAA consensus-building Algorithm 2 drives the consensus-building actions.

If consensus is not achieved, the CMAA algorithm proceeds with consensus-building. Detailed analytical results and information on the discrepancy with the resolution having the lowest potential entropy are provided to the (digital) facilitator.

After the facilitator and the decision-makers were able to review the intermediate results, the clarification conference is initiated and guided by the (digital) facilitator. At the beginning, the facilitator presents the discrepancy and the individual evaluations provided by each decision-maker. Subsequently, the facilitator invites the decision-makers to articulate the rationale behind their evaluations, facilitating the exchange of their mental models. Upon sharing all relevant information and addressing any queries, the facilitator prompts the decision-makers to collaboratively agree on a resolution.

If the group is unable to reach a resolution, the facilitator initiates the next consensus step without a resolution. Subsequently, the CMAA algorithm removes the unresolved discrepancy from the set of discrepancies available for selection and identifies the discrepancy with the next lowest potential entropy for the following consensus step iteration (s++).

On the other hand, if the group successfully reaches a resolution, they communicate it to the facilitator. The facilitator then records the outcome and transmits it to the CMAA algorithm to commence the next consensus iteration (s++).

This iteration continues until a consensus is achieved. Once consensus is reached, the consensus-building results are consolidated and documented, including an audit trail and a decision report. All outcomes are then shared with the facilitator, who presents the recommendation and other decision results to the decision-makers.

6.2.3 Digitally assisted facilitation in practice

The author facilitated two case studies using a prototype of the digital facilitator assistant (see Sections 9.1 and 9.2). Decision-makers completed a Google Form to input their initial preferences and judgements. Subsequently, CMAA analysis was conducted, followed by the distribution of individual decision reports to the decision-makers. These reports provided an overview of their evaluations, the number of discrepancies, and the current level of consensus.

To facilitate the consensus-building process, an in-person meeting was organised. Each step of the consensus-building process was visualised on a digital whiteboard using Miro software (RealtimeBoard Inc., 2024), accessible to all decision-makers via a monitor. The facilitation was supported by a basic terminal interface, granting the facilitator access to the CMAA analysis results and the discrepancy yielding the fastest path to consensus. This allowed the facilitator to initiate a clarification conference. Then the facilitator updated the resolution and started the next consensus iteration.

Both the pre-consensus report and the real-time visualisation during consensus-building aimed to foster trust.

6.2.4 Potential representation as a thinkLet

According to de Vreede et al. (2021):

> *Collaboration Engineering is an approach to designing collaborative work practices for high-value recurring tasks and deploying those designs for practitioners to execute for themselves without ongoing support from expert facilitators.*

Collaboration Engineering encourages the use of thinkLets for collaborative tasks. According to Briggs et al. (2003), a thinkLet is

> *[...] the smallest unit of intellectual capital required to create one repeatable, predictable pattern of collaboration among people working toward a goal.*

The goal of a thinkLet is to eliminate the need for an expert facilitator; it provides a facilitation guide for a member of the group and a blueprint for a digital facilitator.

There are already some thinkLets which implement single-step MCDM decision methods such as the thinkLet "LogicalMulticriteriaSort" by Ducassé and Cellier (2012). Because it has a formalised description that is composed of small, standardised, self-contained tasks, CMAA fulfils the requirements for a thinkLet.

6.2.5 Mock-ups for scenario 1 (facilitator assistant)

The following two mock-ups illustrate ideas for parts of a dashboard overview for a human facilitator, providing further details for the previously introduced user story. They are an

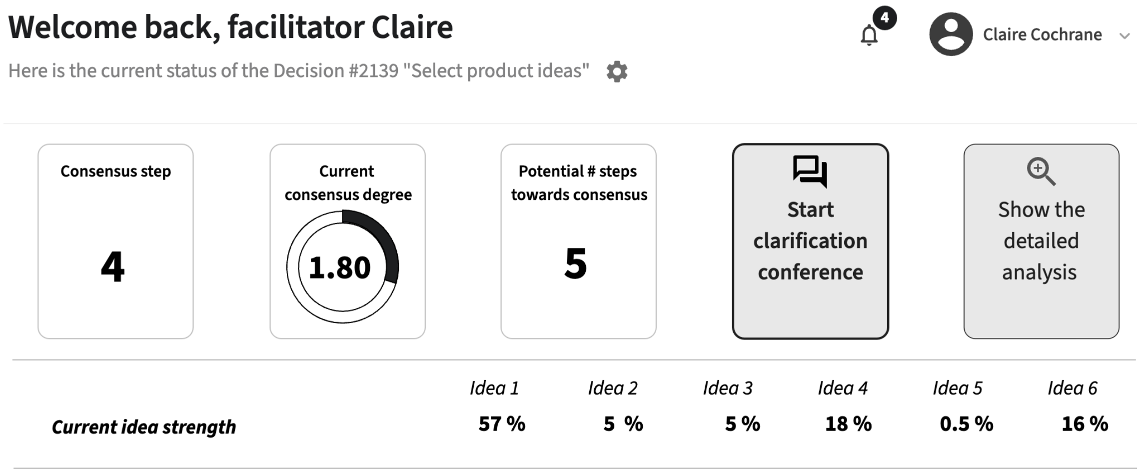


Figure 6.3: Mock-up 1: current consensus status overview for a facilitator

implementation of the first action in the column of the (digital) facilitator shown in Figure 6.1.

Figure 6.3 shows the current status of the consensus step 4 of the realistic consensus path outlined in the illustrative example in Section 5.4.2. In this mock-up it is assumed that the illustrative decision concerns a product selection task. The overview provides the information for the human facilitator, Claire. In the first row, Mock-up 1 displays the current consensus step, the current consensus degree of $h = 1.80$, and its visual representation using the donut progress bar (in black). Additionally, it indicates the projected number of consensus steps required for the group to reach their consensus threshold. Following this information are two buttons: the first for initiating the next clarification conference, and the second to access additional analytical details for Claire. At the bottom, the rank-1 acceptabilities of all six ideas, titled 'current idea strength', are displayed. It ranks the ideas, with *Idea 1* being the strongest, *Ideas 4* and *6* falling into a medium-strong category, and *Ideas 2*, *3*, and *5* considered weak. Based on this information, Claire understands that no idea can be excluded from the set of choices at this point. This overview empowers her to either initiate the next clarification conference or explore a more comprehensive decision analysis, if she requires further insights.

Discrepancy with largest consensus progress: judgement discrepancy for *Idea 1* regarding the criterion *Uniqueness*

Decision-maker	***Judg.***	***Potential consensus***	***Change on the strength of ideas***					
			Idea 1	*Idea 2*	*Idea 3*	*Idea 4*	*Idea 5*	*Idea 6*
Ada	*2*	▼ 2.25	- 40 %	61 %	126 %	42 %	148 %	91 %
Barbara	*5*	▲ 1.67	5 %	- 2 %	- 42 %	-	- 58 %	- 15 %
Grace	*8*	▲ 0.76	51 %	- 77 %	- 93 %	- 49 %	- 100 %	- 85 %

Figure 6.4: Mock-up 2: overview of the effects of each discrepancy resolution

In Figure 6.4, a second mock-up offers a detailed view of the judgement discrepancy that holds the most potential for enhancing consensus. It highlights a discrepancy among the three decision-makers Ada, Barbara and Grace in their numerical evaluations of the *Uniqueness* criterion concerning *Idea 1*. The table indicates to Claire that resolving the discrepancy to Barbara's or Grace's judgement would enhance consensus, whereas agreeing with Ada's judgement would diminish it.

Additionally, from the second mock-up, Claire can deduce that if the group resolves to Ada's judgement, *Idea 1* would lose 40% of its strength. In comparison, all other ideas would benefit by percentages ranging from 42% to 148%. On the other hand, resolving to Barbara's or Grace's judgement would result in *Idea 1* benefiting from a 5% to 51% increase in strength, while other ideas would experience a decrease.

Claire gains the insight that achieving a consensus degree of 0.76 would indicate a significant improvement. As per Requirement 2 on page 8, she should not disclose these insights at the start of the clarification conference, to avoid biasing the decision-makers. However, this prediction equips her to promptly announce the updated consensus status during a live meeting, once the group has provided their resolution.

Additional ideas for extending the dashboard for the facilitator are:

- Providing an overview of active and inactive discrepancies (see Section 3.3).
- Displaying pivot preferences or judgements for each alternative, because they are in danger of being eliminated.
- Showing the preference or judgement sensitivities for each alternative (see Section 3.2.4).
- Offering a selection of discrepancies that offer the greatest potential for consensus improvement, allowing the facilitator to choose based their observation of the group.

6.2.6 Mock-up for scenario 2 (automated facilitation)

The mock-up for the automated facilitation elaborates on the user story presented to address the task of the digital facilitator, which involves providing step-by-step directions for the clarification conference (second action of the digital facilitator in Figure 6.1). One viable implementation of the digital facilitator is through a chatbot. This setup allows the group to convene in person and follow the interactive guidance provided by the chatbot, as conducting their clarifications through this medium may facilitate smoother communication and interaction.

For designing instructions by a digital facilitator, Knoll et al. (2007) advocate using thinXels, which are defined as elementary facilitator instructions. For designing the guidance of the digital facilitator in a clarification conference, the process can be broken down into three instruction types in a thinXel-like manner:

1. *Present*: Introduce the selected discrepancy to the decision-makers, along with the initial evaluations of each decision-maker.
2. *Clarify*: Prompt each decision-maker to share the information that led to their evaluation.
3. *Re-evaluate*: Invite decision-makers to reevaluate the discrepancy after their clarification.

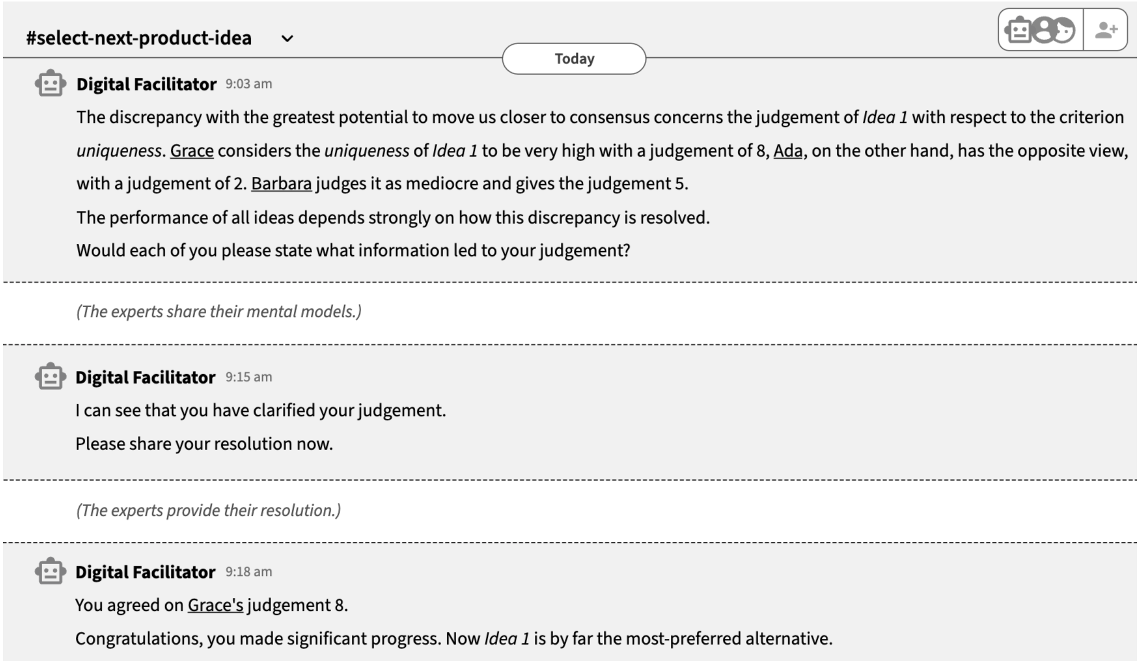


Figure 6.5: Mock-up 3: excerpt of a chat-bot-automated clarification conference

In Figure 6.5, Mock-up 3 outlines an idea for an automated chatbot conversation for a clarification conference. It involves Ada, Barbara, Grace, and the digital facilitator addressing the resolution of the same consensus step and judgement discrepancy as shown in Scenario 1. Initially, the digital facilitator provides an overview of the input judgements of the three decision-makers (instruction *present*). Next, the chatbot prompts the group to elucidate the information that influenced their respective judgements (instruction *clarify*).

After the decision-makers have shared their mental models, the digital facilitator prompts for their resolution (instruction *re-evaluate*). Once the decision-makers have provided their resolution, the digital facilitator illustrates the impact their resolution had on the strengths of the ideas. With the group opting to resolve the judgement discrepancy at *Idea 1* regarding the criterion *Uniqueness* to align with Grace's judgement of 8, significant progress is achieved in building consensus. As a result, *Idea 1* emerges as the most-preferred alternative.

By combining the facilitation instruction with a large language model, the speech patterns used by the digital facilitator could simulate a conversation between the facilitator and the decision-makers, either through a text-based tool or voice-based interaction. In a text-based tool, the language would be more natural. Using a voice-based interaction, combined with a visualisation tool to present intermediate results, written communication would no longer be necessary. This would allow decision-makers to engage in a conversational format for a more streamlined and accessible experience, simplifying their interaction.

This mock-up demonstrates how the approach extends the benefits of multi-criteria models to consensus-building: it focuses the group's attention on either a preference or a criterion/alternative pair and thereby defining a highly contained and hence more manageable discussion.

6.3 Conclusion

Throughout the case studies (see Sections 9.1 and 9.2), a terminal-based facilitation assistance was used in conjunction with the visualisation tool Miro. While this setup involved some manual effort, it remained accessible even to facilitation novices. The first two mock-ups emphasised the small amount of information needed by human facilitator, suggesting that a facilitation assistant could be implemented with minimal development efforts.

An intriguing prospect lies in developing a digital facilitator. The most intricate aspect of the interaction process is the clarification conference, which can be streamlined with just three instructions, as depicted in Mock-up 3. Such an approach could enhance the accessibility of CMAA consensus-building methods. Moreover, integrating a natural language model or a voice-based system could potentially enable the computer to conduct the entire clarification conference autonomously.

This chapter has introduced several ideas and considerations for implementing the CMAA framework, aligning with the objective of accomplishing Goal 6 on page 9.

7

Performance Studies and Practical Considerations

In this chapter, an analysis of the CMAA framework is conducted, evaluating its performance and validating it against the research questions and hypotheses outlined in Chapter 1. The size of the combinatorial space before and after consensus-building is simulated, studying Hypothesis 1 on page 10. Additionally, simulation studies show what number of samples of the combinatorial space delivers a sufficient accuracy (Hypothesis 2 on page 10) for allowing CMAA to be used in live group decision-making.

One of the goals of this work (Goal 2 on page 7) is to provide a framework that allows for efficient consensus-building. Simulation experiments studied CMAA's consensus-building efficiency for a revelation-invariant consensus.

Another study examines and compares the consensus-building of CMAA with a similarly structured consensus-building using an input metric. This comparison studies the capabilities of the CAE metric when a group is committed to reaching a revelation-invariant consensus (Goal 3 on page 3).

Additionally, a study investigates the effect of compromise resolutions on consensus-building (Research Question 2 on page 11).

Furthermore, practical insights are offered to facilitate the application of CMAA in real-world scenarios. The findings discussed in this chapter are derived from a series of studies, as documented in the articles (Goers and Horton, 2023a; Goers et al., 2024; Goers and Horton, 2024a,b,c).

7.1 Computability and accuracy using Monte Carlo simulation

In Section 3.4, it was argued that 10,000 random instances from the combinatorial space provide an error bound of 0.01. Figure 7.1 visualises results from the Monte Carlo simulation for sample sizes ranging from one million down to 10. Shown are the rank-1 acceptabilities for a randomly generated decision of size $d = n = 3$ and $m = 6$ as well as the ground truth using the entire set of 387,420,489 instances on the left. The difference to the ground truth is visible for 1,000 samples and is clearly unacceptable for 100 and 10 samples.

Even a small error can impact the CMAA analysis and the consensus-building. In the

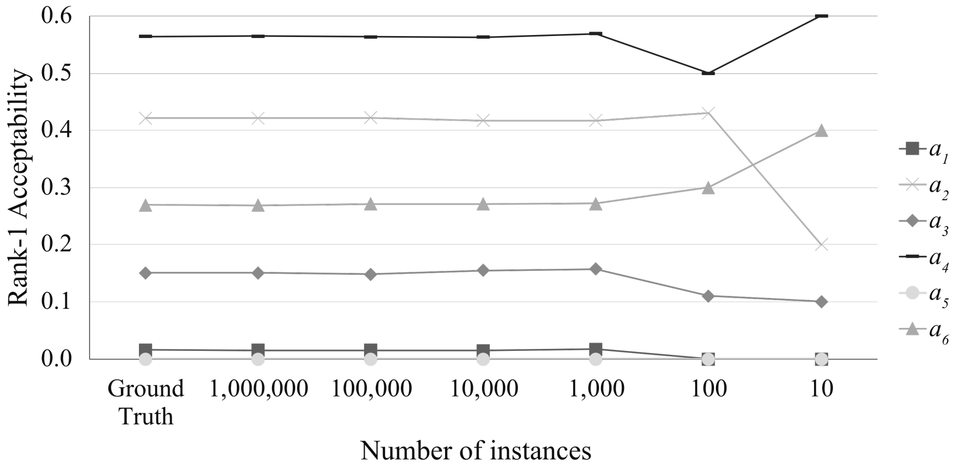


Figure 7.1: Illustration of accuracy

random instance set K_{MC} of a Monte Carlo simulation, it is possible that none of the instances that assign rank-1 to a particular alternative a_i are included in the sample. This can affect both current and potential rank-1 acceptabilities, possibly leading to the erroneous elimination of an alternative that is merely infrequent, yet still ranks first in some (uncaught) instances.

This type of error can propagate to the next iteration. It is possible for an alternative to have a rank-1 acceptability that is 0 in a given iteration, but greater than 0 in the next. This is due to the fact that a fresh set of 10,000 random samples is used in the new iteration. These new samples can contain instances which return a rank-1 position for that alternative. The consequence of this occurrence is that the entropy in the new iteration differs from the prediction in the previous iteration. The risk of it happening can be reduced by increasing K_{MC}. However, the worst that can happen is that the consensus-building process is extended by one step.

This result supports the theoretical considerations discussed in Section 3.4 and provides support for Hypothesis 2 on page 10. This allows the following recommendation to be made:

Recommendation 1. *(Instance sampling) Using* $K_{MC} = 10,000$ *in a Monte Carlo simulation is recommended for randomly sampling the combinatorial space.*

7.2 Impact of the entropy threshold on consensus efficiency

A hard CAE-based consensus is a very demanding requirement. If the group is satisfied with a very firm consensus, they can reduce the number of consensus steps by adopting a soft consensus measure. This study aims to determine the level of entropy needed to achieve a sufficiently firm consensus.

7.2.1 The long tail in consensus-building using the CAE metric

Simulations revealed that, in some cases, the presence of very weak alternatives in the decision can significantly lengthen the consensus path. This behaviour was studied for two problem dimensions: a larger-sized dimension (dimension 1) with $d = n = 6$ and $m = 20$ and

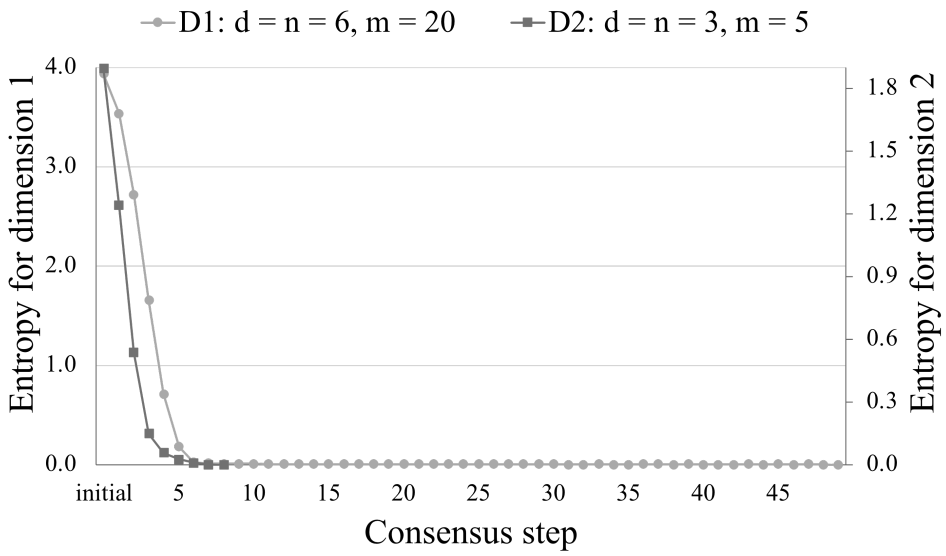


Figure 7.2: The long tail of small entropies in consensus-building

a smaller-sized dimension (dimension 2) with $d = n = 3$ and $m = 5$. The decision model used was SAW (see Section 2.2.2). Consensus paths were simulated for the two problem dimensions with $K_{MC} = 10,000$ instances and for 10,000 different decisions for each problem dimension. The consensus paths were calculated using EO-resolutions. Consensus-building continued until a hard consensus of $h = 0$ was achieved.

Figure 7.2 shows the behaviour of the entropy. The horizontal axis represents the consensus steps. The vertical axes represent the entropy. They are scaled so that both graphs start at the same height at step 0. Then the two graphs show the mean entropy at each consensus step for the two different problem dimensions. The lighter graph with round markers shows problem dimension 1. The darker graph with square markers shows problem dimension 2, which reaches hard consensus after seven steps.

For the decision with dimension 1, the entropy is reduced to a very small value within six steps on average with $h < 0.1$. The mean path length for dimension 1 problems is 4.32 steps, whereas the small-sized problem (dimension 2) has a mean path length of 2.32 steps. The mean entropy quickly drops in the first few steps quickly for both decisions. As shown in Section 4.2, entropies below 0.1 indicate that the rank-1 acceptabilities are very close to a standard unit vector. Thus, for both problem sizes, the consensus level reached at entropies below 0.1 in just six steps for dimension 1 and in four steps for dimension 2.

The ‘long tail’ of iterations for the larger decision from steps 7 to 49 is caused by rare cases in which the sample set at the new iteration ‘discovers’ instances that were not contained in the previous set, leading to a long sequence of iterations in which only acceptabilities that are very close to 0 are affected. This problem disappears when a soft consensus is used.

7.2.2 Mean consensus path length for various consensus thresholds

The next study examines the effect of different values of h_{stop} on the consensus path length. A simulation experiment was conducted with decisions where $d = n = 5$ and four different numbers of alternatives, $m \in \{5, 10, 15, 20\}$. Six different h_{stop} thresholds were set for

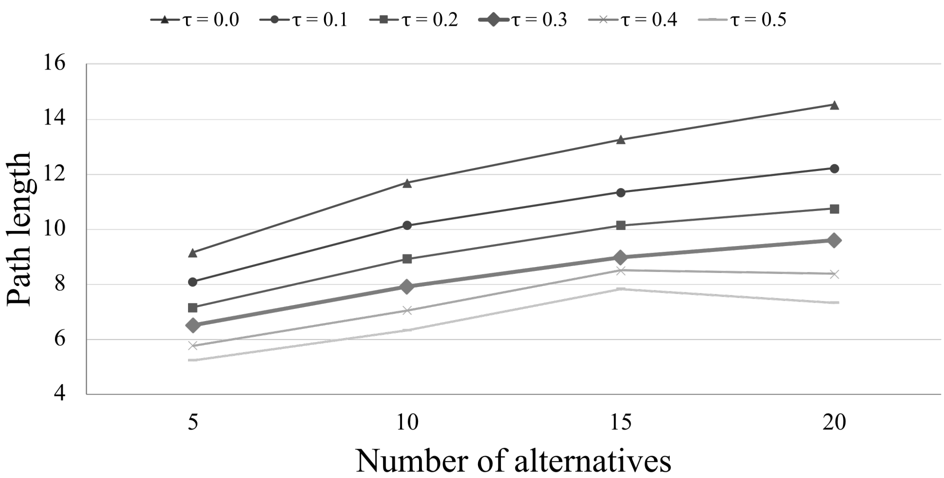


Figure 7.3: Mean consensus path length for different τ values

τ values ranging from 0.0 to 0.5 in 0.1 intervals. The Monte Carlo simulation included $K_{MC} = 10,000$ instances, with $1,000$ random decisions for each value of m. For each decision, 100 consensus paths were simulated using random resolutions at each step, resulting in a total of $100,000$ consensus path simulations for each decision size and each value of h_{stop}.

The mean consensus path lengths for the different thresholds are shown in Figure 7.3. The horizontal axis represents the four different problem sizes studied, which differ only in the number of alternatives. The vertical axis shows the consensus path length. Note that the axis starts at 4 to improve the readability of the closely spaced graphs. Six graphs are shown, which represent the different values of τ, with lighter graphs indicating higher τ values.

At $\tau = 0.3$, the mean consensus path is only about 66% to 71% as long as the path length for $\tau = 0.0$, which represents a significant efficiency improvement.

7.2.3 Rank-1 acceptabilities for different thresholds

This simulation study examines the rank-1 acceptability of the most-preferred alternative for different consensus thresholds τ. The problem size was set to $d = 6$, $m = 10$, and $n = 5$, and the SAW decision method (see Section 2.2.2) was used. The initial CMAA analysis was conducted with $K_{MC} = 1,000$. Consensus-building was performed for 1,000 random decisions with 100 different consensus paths each, resulting in 100,000 different consensus paths.

The consensus thresholds studied were $\tau = 0.0$, 0.1, 0.2, 0.3, 0.4 and 0.5. The resulting rank-1 acceptabilities of the most-preferred alternative were grouped into six buckets ranging from $[0.9, 1.0[$ to $[0.4, 0.5[$ and one bucket containing just the single value 1.0.

Figure 7.4 shows a histogram of the rank-1 acceptabilities of the most-preferred alternatives for each consensus threshold τ. The horizontal axis represents the buckets for the rank-1 acceptabilities and the rank-1 acceptability 1.0, while the vertical axis shows the number of occurrences for each bucket. Six differently shaded columns are shown, ranging from darkest for $\tau = 0.0$ to lightest for $\tau = 0.5$. In addition, the columns for $\tau = 0.3$ are diagonally striped.

The black column for $\tau = 0.0$ indicates that all 100,000 consensus paths resulted in

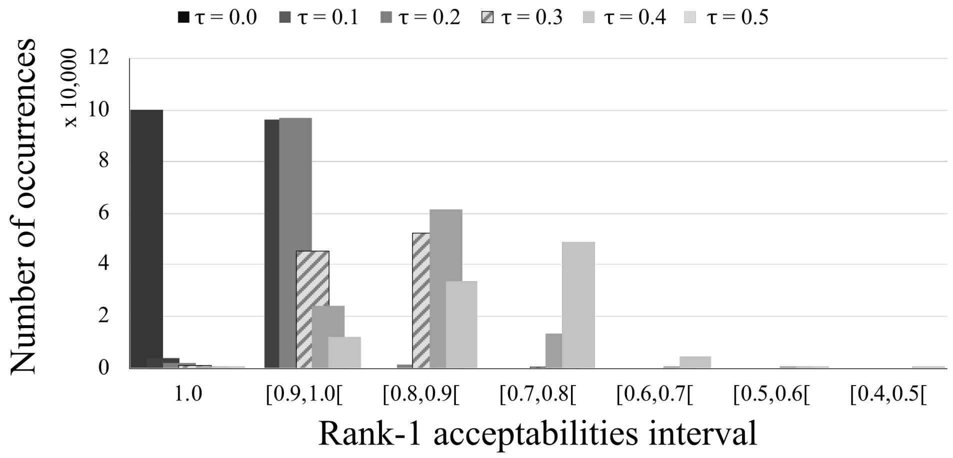


Figure 7.4: Distribution of rank-1 acceptabilities for various termination criteria τ

a rank-1 acceptability of 1.0 for the most-preferred alternative, as expected. The lightest columns, representing $\tau = 0.5$, show the results for a more relaxed soft consensus, where the most-preferred alternative's rank-1 acceptability can be as low as 0.4 (8 occurrences), which can no longer be considered a firm consensus. For $\tau = 0.3$, 99.9% of the rank-1 acceptabilities for the most-preferred alternative are greater than or equal to 0.8.

7.2.4 Recommendation

For groups that prioritise efficiency over achieving a revelation-invariant consensus, $\tau = 0.3$ is recommended for specifying the termination threshold, as it delivers a very firm soft consensus. If the group expresses doubt about the soft consensus, they can always continue the consensus-building process with additional clarification conferences.

Recommendation 2. *(Entropy threshold) If a soft, but very firm consensus is sufficient, using* $\tau = 0.3$ *as a stopping criterion for consensus-building is recommended.*

7.3 Performance comparison of two consensus metrics

The most commonly used metric in multi-criteria consensus-building are input metrics (see Section 2.5.2). The next simulation study compares the convergence speeds of the consensus-building process using the CMAA consensus-building of Algorithm 2 and the MACRP 3 consensus-building algorithm by Zhang et al. (2019) (see Section 5.4.4), which uses the input metric CM_I from Equation 2.2.

The following simulations were similarly set up. The decisions were configured at $d = n = 5$ and four values of $m \in \{5, 10, 15, 20\}$. Weights were equal for all criteria and contained no discrepancies.

Judgements were random and uniformly distributed in the range $[0.0, 1.0]$. The decision model employed was SAW (see Section 2.2.2). Monte Carlo simulation was used to simulate the consensus paths of CMAA and MACRP 3 for $1,000$ decisions for each problem size. Both methods were given identical decisions to process. The CMAA analysis was conducted using a Monte Carlo simulation with $K_{MC} = 10,000$ instances.

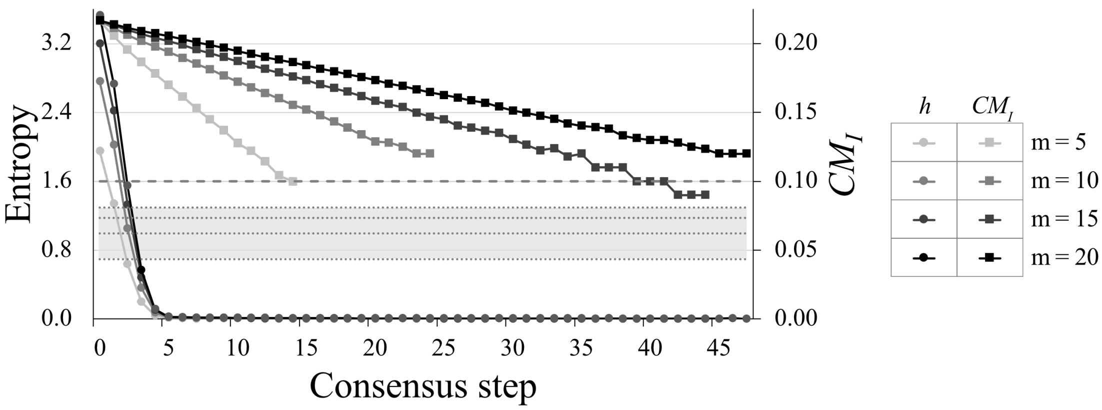


Figure 7.5: CMAA consensus-building with CAE and CM_I metric

7.3.1 CMAA consensus-building compared to an input metric

A consensus path simulation using the CMAA Algorithm 2 was conducted with EO-resolutions until a hard consensus was reached. At each step, the current entropy and the input-based consensus metric CM_I were measured.

Figure 7.5 illustrates the results of this simulation. The horizontal axis represents the consensus steps. The primary vertical axis shows entropy h, while the secondary vertical axis displays the input metric CM_I. Both metrics indicate a hard consensus when $h = CM_I = 0$.

A dashed reference line marks the input-based consensus threshold at $CM_I = 0.1$, while dotted lines represent the soft consensus thresholds for the CAE metric determined by $\tau = 0.3$ in Equation 4.4. The bottom dotted line corresponds to $m = 5$ and the top line to $m = 20$. The entropy consensus threshold corridor is indicated by a light gray background.

The colour of the graphs corresponds to the problem size, with lighter colours indicating smaller problem sizes and darker colours representing larger problem sizes. Graphs marked with dots represent the mean entropy-based metric at each consensus step, while graphs marked with rectangles display the mean input metric CM_I for each step of CMAA's consensus-building process. The mean entropies begin at 1.95, 2.76, 3.20 and 3.53 (from smallest to largest number of alternatives), while the input metric CM_I starts at about 0.22 for all sizes.

The mean path lengths for reaching a hard entropy-based consensus are 7.2, 10.2, 13.4, and 16.8 for the different decision sizes from smallest to largest, respectively. For all problem sizes, the mean CAE-based metric dropped to a mean entropy below 30% of h_{max} after just two to three consensus steps and below 2% of h_{max} after only five consensus steps. This rapid improvement is notable, because the CAE consensus metric is much stricter than the input-based consensus metric.

By contrast, the mean input metric CM_I decreases only slowly. At step 5, the entropy has already signalled an extremely high level of output consensus, but the values for CM_I have only improved by a few percent from their starting levels and are still far from their termination threshold at $CM_I = 0.1$.

Two of the four input metric graphs reach their input-based threshold within 47 iterations and two do not. decisions with five alternatives reach the input-based threshold on average at step 14, which is more than half the number of discrepancies for this problem size. The second graph to meet the input-based consensus threshold is for decisions with 15 alternatives, reaching it at step 39, which is also more than half of the maximum number of discrepancies.

In this experiment, the input metric was measured alongside the CMAA consensus-building Algorithm 2. It is possible that the input metric performs better within a consensus-building process specifically designed for it, which will be examined next.

7.3.2 MACRP 3 consensus-building compared to the CAE metric

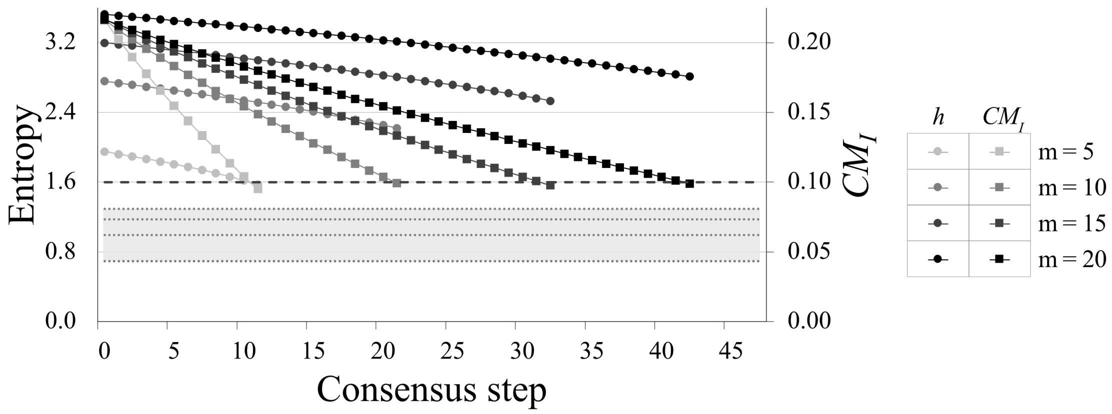


Figure 7.6: MACRP 3 consensus-building with CAE and CM_I

The MACRP 3 consensus-building algorithm was simulated with ideal compromise resolutions until a soft input-based consensus was reached at $CM_I = 0.1$. At each consensus step, a discrepancy was resolved to the average of its constituent evaluations. Figure 7.6 presents the results of this simulation, following the same presentation format as in Figure 7.5.

With the input-based consensus metric, hard consensus can only be achieved when all discrepancies have been resolved, which occurs at steps 25, 50, 75, and 100, for the smallest to largest problem size, respectively. However, a mean soft input consensus is already achieved at steps 11, 21, 31, and 43. At these steps, the mean CAE consensus metric measured a mean entropy of 1.57, 2.22, 2.56, 2.83, respectively, all of which are still far from a soft entropy-based consensus threshold.

The input metric CM_I is inadequate for detecting output consensus and terminates the iteration while there is still much disagreement about the most-preferred alternative. The reason lies in the MACRP 3 consensus-building algorithm. At each step, discrepancies are resolved to their average as a compromise. Although each resolution decreases the dissimilarity of the inputs, it does not necessarily change the output performance of the alternatives. Consequently, the rank-1 acceptabilities for the alternatives do not change significantly, and the CAE consensus metric does not improve.

7.3.3 Recommendations

Two insights can be drawn from this experiment. First, the input metric can only detect a hard consensus when all discrepancies are resolved, requiring a large number of steps which is both impractical and also significantly greater than is required by CMAA with the CAE metric.

Second, the input metric can only detect a revelation-invariant consensus when all discrepancies have been resolved. On the other hand, its soft consensus threshold can be misleading, when output consensus is desirable, because input-based consensus does not measure the degree of agreement on the most-preferred alternative. Consequently, for cooperative decisions, the input metric does not provide a useful measure to guide towards a correct decision.

This establishes the CAE metric as a reliable measure for reaching a revelation-invariant consensus. Moreover, in CMAA's consensus-building algorithm, consensus was reached with fewer than ten resolutions, and the large majority of discrepancies did not need to be treated. This efficiency is one of the primary advantages of CMAA in combination with the CAE metric. This supports Hypothesis 1 on page 10, indicating that the combinatorial space has enough Slack to allow for an efficient path to hard consensus.

Recommendation 3. *(Correct decision with CAE) Decision-maker groups that value a revelation-invariant consensus, should use the CAE consensus metric.*

7.4 Consensus-building efficiency

This section examines the efficiency of consensus-building within the CMAA framework. The consensus-building efficiency is compared for different decision sizes and different types of decision models. Further, studies investigate the effect of random and entropy-optimal resolutions, and the effect of cooperative and competitive resolutions.

7.4.1 Convergence efficiency for different decision sizes

In this section, consensus-reaching efficiency is examined from three different angles. First, a compensatory decision model and a non-compensatory decision model were compared in terms of their speed in achieving consensus. Following this, the convergence efficiency of the widely-used AHP decision method was analysed. Lastly, the distribution of consensus path lengths was investigated across ten different decision methods.

Comparing two types of convergence speeds for two decision models

In this simulation study, the convergence speeds of a compensatory method and a non-compensatory method are compared. Representing the compensatory method, the SAW decision method was chosen (see Section 2.2.2). As for the non-compensatory method, the Lexicographic method was selected (Fishburn, 1974) (see Section 2.2.2). It is expected that the Lexicographic method will converge faster than SAW, because it usually does not require all evaluations to determine the most-preferred alternative, whereas an alternative's performance in a compensatory method such as SAW depends on all evaluations.

For this simulation, the number of criteria was constant with $n = 6$, and the numbers of decision-makers and alternatives were $d \in \{3, 5, 7\}$ and $m \in \{5, 10, 15, 20\}$, respectively. To ensure the comparability of evaluations, SAW preferences and judgements were integers $[1, 9]$, while the Lexicographic method used nine equivalence classes for judgements, denoted as $[A, I]$. Preferences in the Lexicographic method were ordinal.

For each method, $10,000$ decisions were randomly generated, with preferences and judgements generated for each decision-maker by sampling from the possible evaluation set with equal probability. The CMAA analysis was conducted using a Monte Carlo simulation with $K_{MC} = 10,000$. Consensus-building was performed using EO-resolutions for two entropy thresholds, $\tau \in \{0.0, 0.3\}$.

Figures 7.7 and 7.8 show the convergence speed of the SAW and Lexicographic methods, respectively. The vertical axis shows the consensus path length, while the horizontal axis represents the number of alternatives. Each graph represents a different number of decision-makers d, ranging from lighter shades ($d = 3$) to darker shades ($d = 7$).

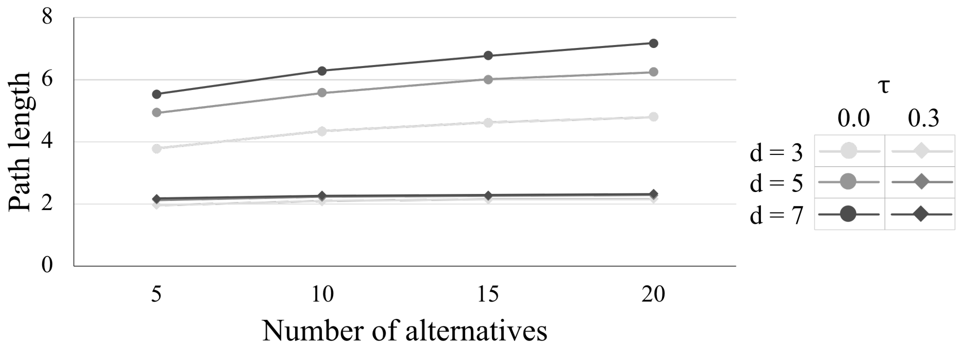


Figure 7.7: Consensus convergence for the SAW decision method

Figure 7.7 shows the results of the convergence simulation for SAW. The upper three graphs, marked with dots, represent the mean path lengths for a hard consensus for $d \in \{3, 5, 7\}$. The lower three graphs, marked with diamonds, show the mean path lengths for a soft consensus using $\tau = 0.3$.

For $\tau = 0.3$, the path lengths are independent of m and d. On average, the consensus path length is about two consensus steps, which is very fast.

The number of consensus steps for $\tau = 0$ increases slowly with m and d. For example, with three decision-makers, the mean consensus path length increases from four to just short of five consensus steps. When seven decision-makers are involved, the consensus steps range from about six to eight.

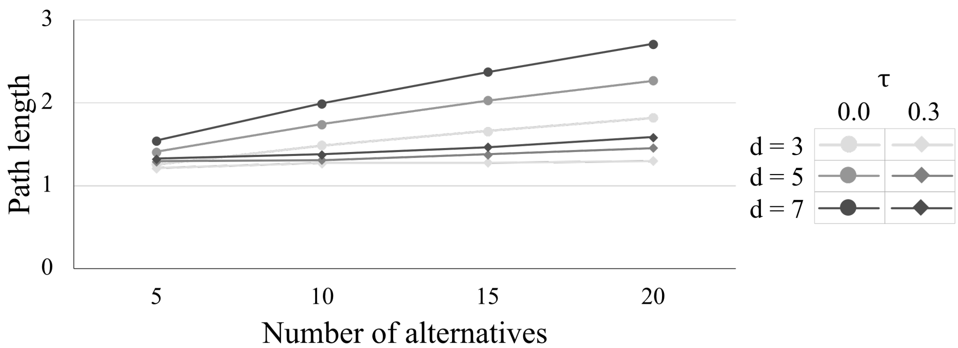


Figure 7.8: Consensus convergence for the Lexicographic decision method

Figure 7.8 presents the results for the Lexicographic method, following the same format as Figure 7.7. In general, the mean path lengths for the Lexicographic method are considerably shorter than when using SAW, confirming the initial assumption about the convergence speed of each method. Their overall range spans from about one consensus step to almost three consensus steps.

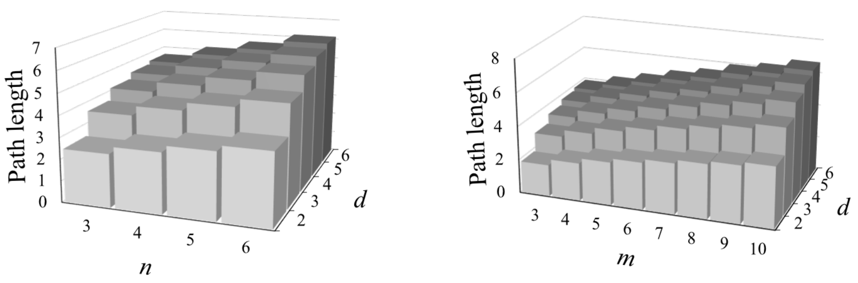
The soft consensus threshold at $\tau = 0.3$ improved the mean path length for SAW significantly, whereas the Lexicographic method does not benefit from it as much. This is because the Lexicographic method already produces very favourable path lengths for a hard consensus. However, there is still a noticeable advantage to be gained. The number of alternatives and decision-makers has a minor linear impact on the mean path lengths.

For $d = 7$ and $\tau = 0$, the Lexicographic method converges roughly three times faster than SAW. However, when a soft consensus is applied, this advantage for the Lexicographic method diminishes by about a half. This disparity in convergence speed may influence the choice of decision model for a group, depending on how their requirements align with each decision model type.

Consensus performance for minimal AHP

The results in this section are taken from a Bachelor thesis (Blumenthal, 2023) that was supervised by the author. The consensus-building efficiency of CMAA when combined with a modified AHP decision method was investigated. This modification is described in Section 8.2. The experimental setup involved varying problem sizes: $m \in \{3, 4, 5, 6, 7, 8, 9, 10\}$, $n \in \{3, 4, 5, 6\}$, and $d \in \{2, 3, 4, 5, 6\}$, with only a single-level of criteria.

The study consisted of 10,000 decisions, which were evaluated using Monte Carlo simulation with $K_{MC} = 1,000$ instances. Only EO-resolutions were considered. Each consensus path iteration continued until a hard consensus ($h_{stop} = 0$) was achieved.



(a) $m = 6$ (b) $n = 4$

Figure 7.9: Mean path lengths for the minimal AHP simulations

Figure 7.9 presents two three-dimensional histograms. Figure 7.9a shows the mean consensus path lengths for decisions with six alternatives. The horizontal axis represents the number of criteria, while the depth axis indicates the number of decision-makers. The vertical axis depicts the consensus path lengths for each combination of parameters. The columns represent the mean path length at the various number of alternative and decision-maker combinations. Figure 7.9b illustrates the mean consensus path lengths for decisions with four criteria. Here, the mean path lengths are depicted for different numbers of alternatives and decision-makers.

These diagrams demonstrate that the mean consensus path length grows linearly with the number of alternatives and criteria. This is to be expected, since the number of discrepancies grows linearly with these two parameters. On the other hand, the mean path length for an increasing number of decision-makers appears to grow asymptotically. This is also to be expected, because the probability of a discrepancy occurring at any location is $1 - 17^{1-d}$, which approaches 1 asymptotically as d increases.

A multiple regression analysis was performed on the data from Figures 7.9a and 7.9b. The resulting formula for the expected consensus path length L is:

$$L = 0.443n + 0.257m - 9.474e^{-0.48d} + 2.46 \ .$$

The fit to the data is extremely good, with a coefficient of determination R^2 of 0.974. The consensus path length grows modestly with the problem size.

7.4.2 Consensus-building efficiency for different decision models

Figure 7.10 illustrates the distribution of consensus path lengths for CMAA in conjunction with ten different decision methods. These methods, as detailed in Section 2.2.2, include SAW, FSAW, WPM, PROMETHEE, TOPSIS, FTOPSIS, AHP (with a single criterion level), SAW-NC, ABX-Lex, and the Lexicographic decision method. SAW-NC represents a non-compensatory variant of SAW, achieved by setting criteria weights to $[10; 100; 1,000; 10,000; 100,000]$. The problem size investigated consisted of $d = 3$ criteria, $n = 6$ decision-makers, and $m = 10$ alternatives.

Each decision model used five different judgements, except for the two fuzzy decision models, which employed seven classes according to Table 8.23. AHP adopted the evaluations outlined in Table 2.3, while ABX-Lex used three equivalence classes as described in Section 2.2.2. Across all decisions, only 50% of the judgement tasks were discrepancies. For each model, $10,000$ random decisions were generated, analysed with the Monte Carlo simulation employing $K_{MC} = 10,000$ instances. Consensus-building was executed using EO-resolutions until a hard consensus was reached applying the parameter $\tau = 0.0$.

Table 7.1: Mean path lengths for the ten decision methods

Decision method	Mean
SAW	4.3
FSAW	4.6
WPM	4.3
PROMETHEE	5.0
TOPSIS	4.8
FTOPSIS	5.1
AHP	3.9
SAW-NC	2.4
ABX-Lex	2.4
Lexicographic	1.9

Table 7.1 summarises the mean path lengths for a hard consensus for each decision method combined with CMAA. The Lexicographic method exhibits the shortest mean consensus path length, averaging 1.9 steps. Conversely, FTOPSIS has the longest mean consensus path length at 5.1 steps, followed closely by PROMETHEE at 5.0 steps on average. Compensatory decision methods (the first seven methods in the table) demonstrate approximately double the mean path length compared to non-compensatory methods (the last three methods).

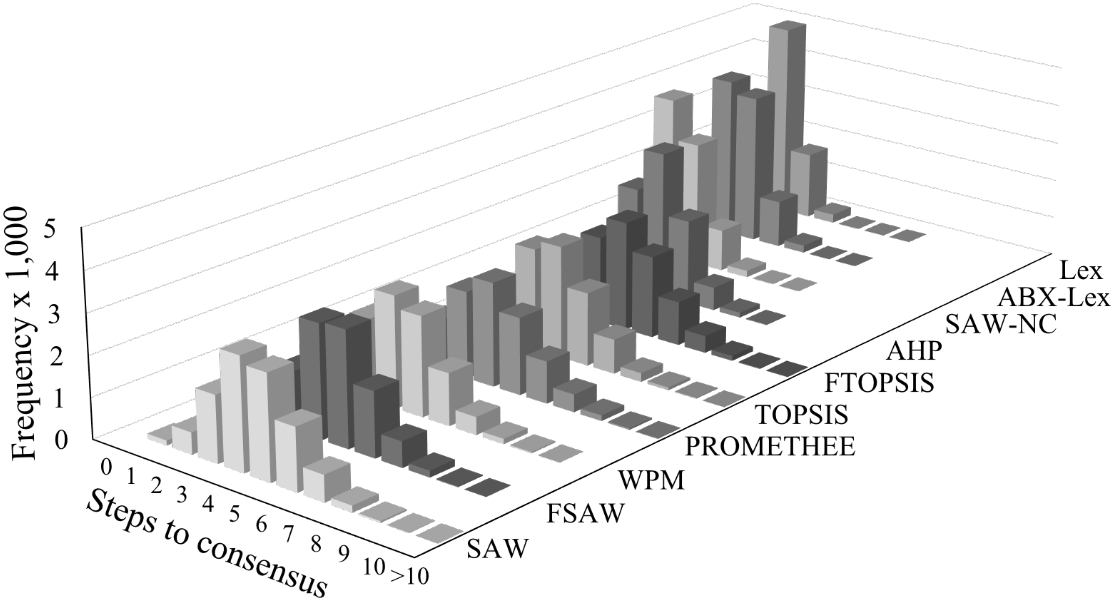


Figure 7.10: Histogram of consensus path length comparing ten different decision models

Figure 7.10 displays the histogram representing the number of steps required to reach hard consensus for all decision methods outlined in Table 7.1. The fastest decision methods (SAW-NC, ABX-Lex, and Lex) exhibit smaller variances compared to the compensatory decision methods. While the PROMETHEE, TOPSIS, and FTOPSIS methods occasionally require more than 10 consensus steps, this occurs in less than 0.1% of decisions. In summary, all types of decision methods demonstrate rapid convergence using the CMAA consensus-building approach.

Recommendation

All studies were conducted using EO-resolutions to reach a hard consensus. Although this is the ideal case that will seldom occur in practice, it can still highlight the differences in efficiency between the consensus-building methods.

A consistent observation across the studies in this section is that non-compensatory decision methods tend to reach a hard consensus more quickly. This underscores the impact of employing a dictator criterion. Such a criterion may alone suffice to determine the first-ranked alternative, reducing the reliance on all judgements and correspondingly on their discrepancies. Compensatory decision methods include judgements with respect to all criteria to compute preferredness, hence, more judgement discrepancies tend to have an impact.

Therefore, if achieving rapid convergence is a priority for a group, they may want to consider using a non-compensatory decision model.

Recommendation 4. *(Fast convergence) If fast consensus-building is required, a non-compensatory decision model should be used.*

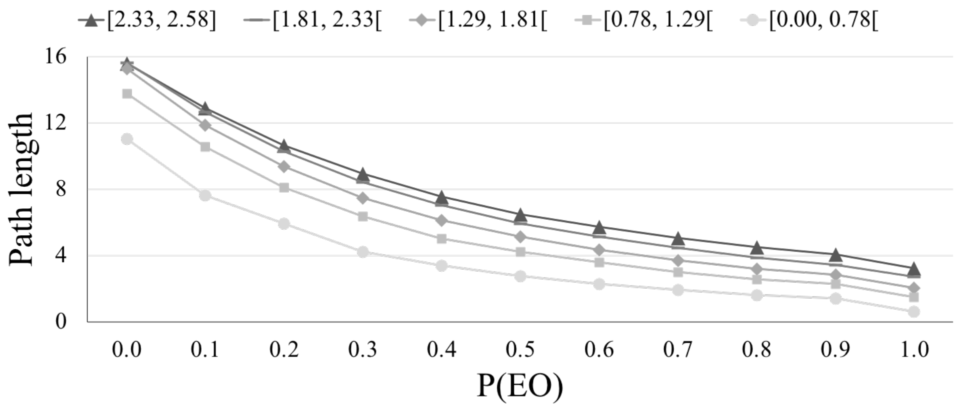


Figure 7.11: Mean consensus path lengths for P(EO)

7.4.3 Convergence performance for realistic resolutions

The CMAA consensus-building Algorithm 2 follows an optimistic strategy. Thus, consensus paths employing EO-resolutions at each step lead to fast consensus convergence. However, these resolutions may not necessarily be the ones chosen by the group after creating a shared mental model, making the EO convergence paths not entirely realistic.

Therefore, the next study examines the impact of choosing either entropy-optimal or non-entropy-optimal resolutions. To this end, the parameter $P(EO)$ is introduced. This parameter represents the probability that the EO-resolution is selected for each discrepancy.

In this study, 1,000 decisions of size $m = n = 6$ and $d = 3$ were simulated. The decision model used was a minimal AHP (see Section 8.2) with only a single level of criteria. For the CMAA analysis, $K_{MC} = 10,000$ instances were used, and consensus-building was terminated when a hard consensus was reached. Each decision was run with 100 random consensus paths. The resolution chosen at each step was random and governed by $P(EO)$. The simulated consensus path lengths were sorted into five buckets according to their initial entropy, indicating the entropy 'distance' the consensus-building process needed to traverse.

Figure 7.11 shows the results of this study. The horizontal axis shows the values of the parameter $P(EO)$, and the vertical axis shows the consensus path length. Each graph represents the average value for one of the five initial entropy buckets. The dark graph marked with triangles represents mean consensus path length for initial entropies in the interval $[2.33, 2.58]$, and the lightest graph marked with dots represent them for initial entropies in the interval $[0.00, 0.78[$. For $P(EO) = 1.0$, the entropy-optimal resolution is chosen at each consensus step. Hence, these values represent the shortest mean path length for each initial entropy.

As $P(EO)$ increases, the consensus path length decreases, indicating faster convergence when more EO-resolutions are chosen. When $P(EO) = 1.0$, the paths are shortest, showing the ideal scenario for fast convergence. The initial entropy buckets show that higher initial entropy requires more steps to reach consensus, but the impact of $P(EO)$ is consistent across different initial entropies. This is not surprising, because the initial entropy represents the amount of ground the consensus-building process has to cover to reach consensus. Even for the worst case $P(EO) = 0.0$, the mean path lengths are still less than 27% to 40% of the overall number of discrepancies.

However, it is unlikely that decision-makers will only resolve discrepancies in a non-entropy-optimal manner. Long consensus paths, occurring when $P(EO) \leq 0.2$, typically arise when entropy is already low and decision-makers repeatedly fail to select resolutions that lead to consensus. Such situations could be mitigated by applying a soft consensus threshold such as $\tau = 0.3$.

This study confirms that employing EO-resolutions accelerates the consensus-building process. The effect of $P(EO)$ provides insight into the flexibility of the consensus process, showing that even when non-entropy-optimal resolutions are occasionally chosen, the path length does not significantly lengthen.

7.4.4 Convergence performance for two types of conflict resolution

The CMAA framework was developed for cooperative decisions (Definition 2 on page 3), where discrepancies are caused by information asymmetries, which can be resolved by sharing information (see Section 2.3.2). However, in practice, discrepancies may also be caused by differences in taste or beliefs, or by conflicting interests, representing competitive decisions (Definition 1 on page 2). Discrepancies of this type can only be resolved by compromise.

A compromise resolves a discrepancy by agreeing to a resolution that lies between the extremes, such as an average. This approach does not reflect the effect of sharing a mental model, but rather the outcome of a negotiation. Such averaging seldom affects the performance of alternatives substantially; it merely decreases the dissimilarity of the input evaluations. Consequently, this type of resolution might produce a longer consensus path compared to one generated from cooperative resolutions.

The next study investigates the behaviour of the consensus path length when both types of conflict resolutions are present. The first type is the agreement on the resolution of a discrepancy, simulated by a random selection from the available preferences or judgements. The second type is a compromise value, simulated as the rounded average of all decision-maker preferences or judgements contained in the discrepancy. The following studies use the agreement ratio ρ as a parameter that describes the proportion of resolutions that are agreements:

$$\rho = \frac{||agreements||}{||agreements|| + ||compromises||} \ . \tag{7.1}$$

When $\rho = 0.0$, all resolutions were compromise resolutions, while for $\rho = 1.0$ all resolutions were agreements on one of the initial input evaluations. At $\rho = 0.5$, half the resolutions were agreements, and half were compromises.

Three simulations were performed:

1. The effect of the agreement ratio on consensus path length for different problem sizes: This simulation examined how the agreement ratio impacts the consensus path length across various problem sizes.
2. The effect of the agreement ratio with various consensus thresholds: This simulation explored the influence of the agreement ratio on the consensus path length for different consensus thresholds.
3. The effect of the agreement ratio on the input-based and CAE consensus metric: This simulation ran the CMAA consensus-building Algorithm 2 for various agreement ratios, comparing the development of the input metric with the CAE metric.

All studies used the SAW decision method (see Section 2.2.2). The CMAA analysis used $K_{MC} = 1,000$ instances in each case. The consensus-building simulations were conducted

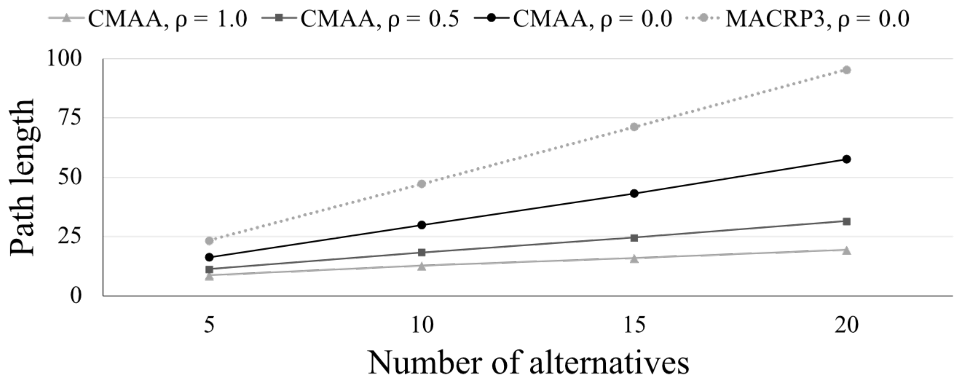


Figure 7.12: Mean path length for various agreement ratios ρ

with $1,000$ random decisions and 100 consensus paths each. During consensus-building, when an agreement is chosen, it is selected randomly. Compromise resolutions were deterministic, while agreement resolutions were chosen randomly.

Effect of agreement ratio for different problem sizes

Four different problem sizes were investigated, with $d = 6$, $n = 5$, and $m \in \{5, 10, 15, 20\}$. Additionally, three agreement ratios, $\rho \in \{0.0, 0.5, 1.0\}$, were studied. The consensus threshold for this study was set at $\tau = 0.3$, representing a very firm soft consensus.

The simulation results are depicted in Figure 7.12. The horizontal axis illustrates the four different decision sizes, while the vertical axis denotes the consensus path lengths. The three lower graphs illustrate the mean path lengths for the agreement ratios $\rho \in \{0.0, 0.5, 1.0\}$ when employing CMAA consensus-building. The topmost dotted graph represents the mean consensus path length using the MACRP 3 consensus-building algorithm (see Section 7.3) for comparison, which exclusively uses compromise resolutions.

Substantial improvement is observed in CMAA consensus-building as the proportion of agreement resolutions increases. Even when only half of the resolutions are agreements, the mean path length is notably shorter than in the all-compromise case. The improvement in mean consensus path length for CMAA ranges between 20% to 37% for $\rho = 1.0$ and between 60% to 70% for $\rho = 0.0$. MACRP 3 needs considerably longer consensus paths compared to even the all-compromise case for CMAA. Thus, even concerning compromise resolutions, CMAA consensus-building demonstrates greater efficiency than MACRP 3.

Effect of agreement ratio for two consensus thresholds

The simulation examines the impact of the agreement ratio on the consensus path lengths for two consensus thresholds h_{stop}, for the parameters $\tau \in \{0.0, 0.3\}$ when applying Equation 4.4. The decision size studied was $d = 6$, with $m = 10$ and $n = 5$, while the agreement ratios considered were $\rho \in \{1.00, 0.75, 0.50, 0.25, 0.00\}$. All other experimental parameters remained consistent with those stated at the beginning of this section.

Figure 7.13 illustrates the results of this simulation. The horizontal axis displays the various agreement ratios, while the vertical axis represents the consensus path length. Each graph corresponds to a different consensus threshold, with the light and dot-marked graph depicting the mean consensus path lengths for hard consensus. Hard consensus ($\tau = 0.0$)

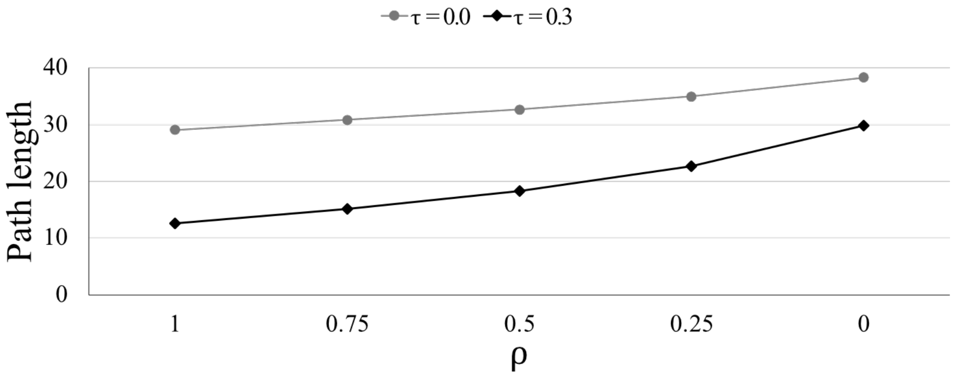


Figure 7.13: Consensus efficiency for two types of discrepancy resolutions

requires on average the longest consensus paths, ranging from nearly 30 steps when $\rho = 1$ to almost 40 steps when all resolutions are compromises.

As might be expected, consensus paths become shorter as the consensus threshold is relaxed and the agreement ratio increases. A higher agreement ratio ρ shortens the consensus path, because the resolutions in a discrepancy that affect the entropy the most are usually the extreme values, and the values in-between often have little effect. For $\tau = 0.3$, which defines the recommended consensus threshold, the consensus path with compromise-only resolutions requires an average of 58% more steps than when the group is in agreement. This diagram shows the efficiency gains in consensus-building that can be achieved when agreements, rather than compromises, are reached.

Effect of agreement ratio on the input-based and CAE consensus metric

The next study compared the behaviour of the input metric CM_I and the CAE metric during consensus-building with CMAA. The decision was of size $d = 6$, $m = 10$ and $n = 5$. Consensus-building was performed until a hard consensus was achieved. Agreement resolutions for $\rho = 1$ were chosen randomly. All other simulation parameters were used from the introduction of this section.

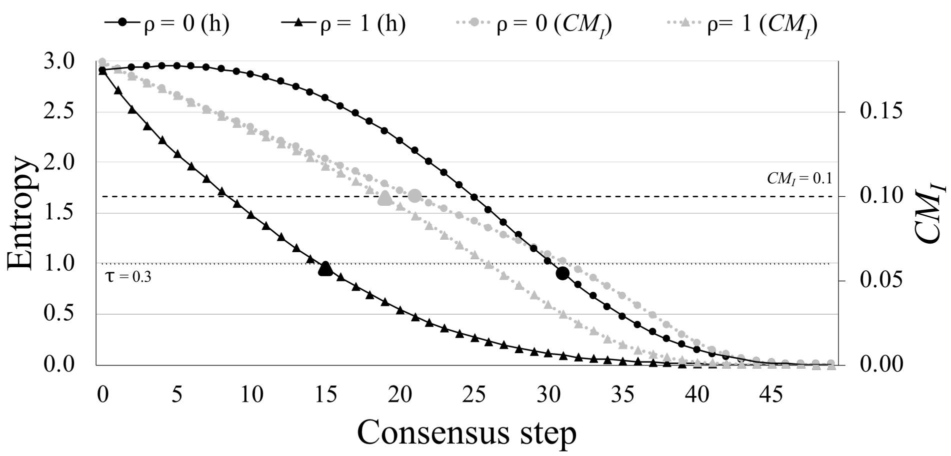


Figure 7.14: Consensus efficiency for two types of discrepancy resolutions

Figure 7.14 depicts the simulation results. The horizontal axis represents the consensus steps. The vertical axes display the consensus measures, with entropy on the primary vertical axis and the input metric CM_I on the secondary vertical axis. They indicate the mean current entropy or mean current CM_I at each consensus step for all simulated decisions. The vertical axes have been scaled so that the starting points of all graphs are aligned. Four graphs are displayed in total. The two darker graphs correspond to the mean current entropy, while the two lighter graphs represent the input-based consensus metric CM_I. Graphs marked with dots represent the parameter value $\rho = 0$, and graphs marked with triangles correspond to $\rho = 1$. The soft termination thresholds for the two consensus metrics are indicated by horizontal lines at $\tau = 0.3$ and $CM_I = 0.1$. The markers are bold where each consensus threshold is met.

The CAE metric reaches a soft consensus on average in 15 steps when $\rho = 1$ and in 31 steps when $\rho = 0$. The latter metric increases initially, before beginning to decrease at step 9. In contrast, for $\rho = 1$, the CAE metric improves on average rapidly from the start. This illustrates that CMAA responds to agreements with faster convergence than to compromises.

The two lighter graphs show the performance of the input metric CM_I during CMAA consensus-building. It indicates a soft consensus on average in 19 steps when $\rho = 1$ (marked by dots), and in 21 steps when $\rho = 0$ (marked by triangles). Both graphs exhibit a more constant improvement speed compared to the CAE metric (dark graphs).

For agreement resolutions ($\rho = 1$), the input metric CM_I requires more steps to reach a soft consensus and does so without reaching a hard CAE-based consensus. Conversely, for compromise resolutions ($\rho = 0$), the input metric indicates a soft consensus early, even though the decision is not close to a soft CAE-based consensus.

This study highlights that CMAA reaches consensus faster for a cooperative decision than a competitive decision. However, the algorithm is tolerant of compromise resolutions; although the consensus path is longer, consensus can still be reached. This simulation shows that the input metric CM_I cannot reliably detect a revelation-invariant output consensus when used in conjunction with CMAA. It tends to suggest stopping the consensus-building process either too early (in a competitive decision), or too late (in a cooperative decision).

Consequences

In practice, a group decision might encompass all causes for an evaluation conflict (see Section 2.3.2). This can lead to a mix of compromise and agreement resolutions. The type of discrepancy is unknown beforehand and is only revealed during a clarification conference. Nevertheless, the facilitator should always strive to reach an agreement, as it can shorten the consensus path.

It is important to note that the introduction of compromise resolutions changes the interpretation of the results. According to Corollary 1 on page 33, a cooperative group reaches a correct decision when they share their mental models. This assumption no longer holds when the group resolves discrepancies with compromises.

Recommendation 5. *(Invest in agreement resolutions) The larger the number of agreements in clarification conferences, the faster the consensus-building process converges. Therefore, it is worthwhile trying to reach agreement.*

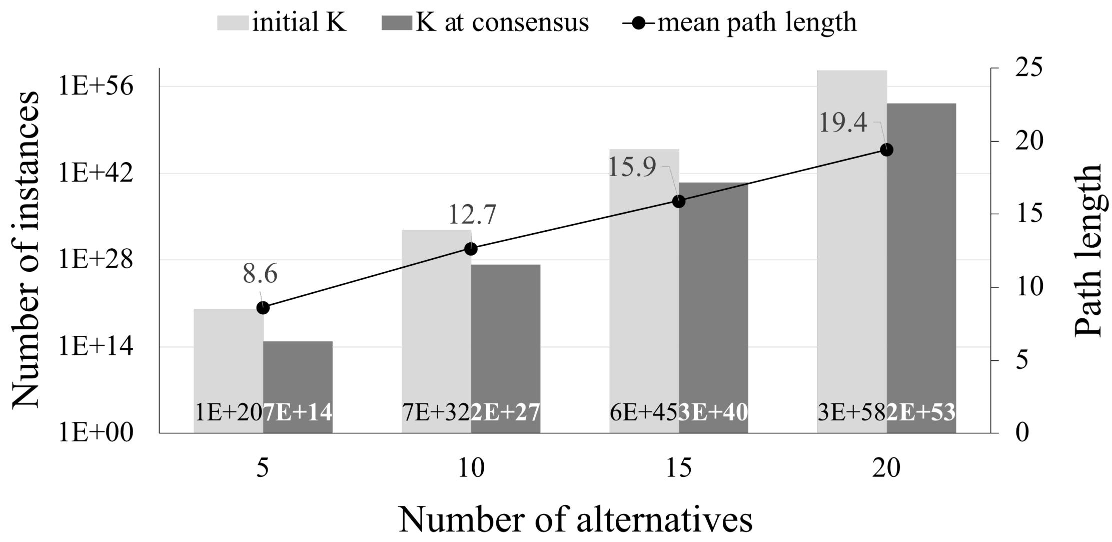


Figure 7.15: Mean number of combinations at the beginning and end of consensus-building

7.4.5 Size of combinatorial space after consensus-building

This work hypothesised (Hypothesis 1 on page 10) that the combinatorial space contains Slack, and not all discrepancies need to be resolved to arrive at a hard consensus. Therefore, the next study explores the size of the combinatorial space before and after consensus-building. It considers decisions of size $d = n = 5$ and $m \in \{5, 10, 15, 20\}$ modelled by SAW (see Section 2.2.2), with integer evaluations in the range $[1, 5]$ for preferences and judgements. The CMAA analysis was performed with Monte Carlo simulation of $1,000$ randomly generated decisions and 100 different consensus paths each, with $K_{MC} = 1,000$ instances at each step. The resolutions were randomly selected, and the consensus threshold was set to $\tau = 0$.

The results of the simulation are shown in Figure 7.15. The horizontal axis shows the four different problem sizes considered. The primary vertical axis shows the number of instances K on a logarithmic scale. The secondary vertical axis shows the consensus path length. The columns show the size of the combinatorial space. The lighter shaded columns shows the combinatorial space before consensus-building, and the darker shaded columns show it when hard consensus is reached. The graph with the dot markers shows the mean path length for each decision size.

On average, the size of the combinatorial space has been reduced by less than 0.01% when consensus is reached. Only a very small fraction of the discrepancies needed to be resolved, while the majority can remain untreated. Since a hard consensus was reached, no resolution of a remaining discrepancy can change the most-preferred alternative. These results confirm Hypothesis 1 on page 10.

Although a revelation-invariant consensus has been reached, the decision-makers might still feel uneasy about accepting the result. In this case, the facilitator should provide arguments to raise their confidence in the result.

Recommendation 6. *(Confidence in the result) When a hard consensus has been reached, no further resolution would change the consensus result. This might be hard to understand for the decision-makers and should be explained to them, if needed.*

7.5 Conclusion

The CMAA framework introduces a new and unique consensus metric, analysis algorithm and consensus-building strategy. Several simulation studies were performed in order to gain some experience into the method's behaviour and provide some recommendations for its application.

The combinatorial space of a decision can reach sizes that require too much computation time in a live group decision environment and are ultimately intractable. Therefore, Monte Carlo simulation can be used, and a sample size of $10,000$ yields a 95% confidence interval of half-width 0.01, which is sufficient in most cases. If a higher accuracy is required, the sample size can be increased to $100,000$ and can still be analysed in a few seconds. This confirms Hypothesis 2 on page 10.

A very firm soft consensus can be achieved with setting the parameter τ to 0.3. Relaxing the consensus threshold results in a shorter consensus path and can save up to half the iteration steps compared to hard consensus. This achieves an improved efficiency in consensus-building, supporting Goal 2 on page 7.

CMAA is much more efficient at building consensus than the representative input averaging algorithm MACRP 3 using its input consensus metric CM_I. The savings in the number of consensus iterations can be as much as 50%. Even in its worst case, in which all discrepancies are resolved to compromise, CMAA converges faster than MACRP 3. Furthermore, the input metric CM_I cannot detect an output consensus and may perform many additional consensus-building iterations after a hard revelation-invariant consensus has been reached. Conversely, the input metric CM_I might suggest stopping the consensus process prematurely (Definition 6 on page 29), before even a soft CAE-based consensus is reached. This shows that the new consensus metric enables an efficient guidance to a revelation-invariant consensus, serving Goals 1 and 2 on page 7.

On the other hand, the input-averaging consensus-building strategy used by MACRP 3 hardly improves the CAE metric at all. From an output metrics perspective, the input-driven changes can drive the output consensus in any direction, which remains hidden when no accompanying output metric is used. This is undesirable for cooperative decisions (Definition 2 on page 3).

CMAA consensus-building was tested with ten different decision methods and converged within a few steps in every case. It could be observed that non-compensatory methods are slightly faster than compensatory methods. The consensus path length grows only slightly, if the problem size or the number of decision-makers is increased. This confirms that the CMAA framework can be combined with other decision methods, which answers the Research Question 4 on page 11.

The entropy-optimal consensus path is unlikely to occur in practice. However, consensus-building remained efficient, even when non-entropy-optimal resolutions were present.

Research Question 2 on page 11 asked if consensus-building is still effective when compromise resolutions occur. A study found that even when compromise resolutions occur, the consensus-building is still more efficient compared to a similar input-averaging approach. The more agreement resolutions a consensus-building process is able to achieve, the faster the group decision converges. It also again shows that the input metric cannot be reliably used, when an output-based consensus is important.

The combinatorial space often contains a very large amount of Slack, and only a fraction of the discrepancies must be resolved to reach a hard and revelation-invariant consensus.

Figure 7.15 shows that in a decision containing 3E58 instances, 2E53 remain after consensus has been reached. This is strong experimental evidence in favour of the core hypothesis of the thesis (Hypothesis 1 on page 10).

8

Application of CMAA to previous case studies

The scientific literature contains few case studies on consensus-building in MCGDM. None provide sufficient data for a direct comparison with CMAA. However, numerous studies exist on single-step multi-criteria group decisions. In this chapter, eight such studies from the literature are compared to the initial analytical results of CMAA, and potential consensus paths are computed to demonstrate what CMAA could have achieved if it had been available at the time.

Two of the studies apply CMAA with decision methods that have non-standard input formats, addressing Research Question 4 on page 11. These studies demonstrate that CMAA consensus-building is efficient, achieving Goal 2 on page 7, and provide evidence for CMAA's analytical capabilities, addressing Research Question 3 on page 11. Five case studies were published in the articles by Goers et al. (2024) and Goers and Horton (2024c).

8.1 Green supplier selection

The three case studies in this section are taken from Banaeian et al. (2018). They describe the selection of green suppliers of olive oil, palm oil and sunflower and soybean oil for an Iranian manufacturer of edible vegetable oils.

The suppliers who were up for selection had been identified by an initial screening process. Data on each supplier was provided by the procurement and strategic sourcing department of the company. Thus, all decision-makers had access to the same information. For each decision, three internal experts were selected as decision-makers who provided a linguistic evaluation of the suppliers and the selection criteria.

The choice conducted in the study was based on the computed rankings by three different MCGDM methods: FTOPSIS, FVIKOR and Fuzzy GRA. In the following, the group FTOPSIS method is referred to as GTOPSIS. In this section, only FTOPSIS will be tested with CMAA. The fuzzy numbers used are the same as those shown in Table 8.23.

The same four criteria were used in all three studies:

c_1 *Service level: On-time delivery, after-sales service and supply capacity*

c_2 *Quality: Quality of material, labor expertise and operation excellence*

c_3 *Price: Product/service price, capital and financial power*

c_4 *EMS: Environmental prerequisite, planning and certificates*

The initial fuzzy criteria preferences by the three decision-makers were applied for all three supplier selections.

Table 8.1: Unified preferences for the green supplier selection decisions

c_1 (Service Level) $\{\mu_1, \mu_2\}$	c_2 (Quality) $\{\mu_1, \mu_2, \mu_3\}$	c_3 (Price) $\{\mu_1, \mu_2\}$	c_4 (EMS) $\{\mu_1, \mu_2, \mu_3\}$
{H, VH}	{H, MH, VH}	{MH, H}	{ML, M, MH}

The linguistic evaluations of the three decision-makers are shown in Table 8.1. Note that the 'green' criterion c_4 (EMS) receives on average the least important evaluations compared to all other criteria. The number of preference combinations is $||\mathbf{P}|| = 2^2 \cdot 3^2 = 36$ for all green supplier selections.

8.1.1 Green olive oil supplier selection

Table 8.2: The unified judgements for the olive oil selection example

Criteria	Olive Oil Supplier			
	$\{\lambda_1, \lambda_2, \lambda_3\}$	$\{\lambda_1, \lambda_2, \lambda_3\}$	$\{\lambda_1, \lambda_2, \lambda_3\}$	$\{\lambda_1, \lambda_2, \lambda_3\}$
c_1	{F, MP, P}	{G, G, VG}	{MP, P, VG}	{F, MP, P}
c_2	{F, G, MP}	{F, G, P}	{G, MP, VG}	{MG, P, VP}
c_3	{G, MP}	{F, G, MG}	{G, MG, VG}	{F, G, P}
c_4	{F, MG, VG}	{F, G, MG}	{G, VG}	{G, MG, MP}

Table 8.3: Performance indices for the olive oil suppliers

	CMAA						GTOPSIS	
	b_i^1	b_i^2	b_i^3	b_i^4	acc_i	R_O	CI	R_F
O_1	0.06	0.20	0.53	0.22	0.15	3	0.271	3
O_2	0.42	0.41	0.14	0.02	0.42	2	0.665	2
O_3	0.51	0.33	0.13	0.03	0.51	1	0.750	1
O_4	0.01	0.06	0.19	0.73	0.05	4	0.010	4

Four suppliers O_1, O_2, O_3 and O_4 were considered for selecting one olive oil supplier by the decision-makers. The unified linguistic judgement matrix is shown in Table 8.2. The alternatives' performance and corresponding ranking positions for GTOPSIS are shown in Table 8.3. Supplier O_3 is the most-preferred by a small margin.

The CMAA combinatorial space is made up of $||A|| = 2^2 \cdot 3^{14} = 19,131,876$ judgement combinations, the total number of instances for the combinatorial space is $||K|| = 36 \cdot 19,131,876 = 688,747,536$.

Table 8.4: Potential judgement entropies for the olive oil suppliers

	O_1 $\{\lambda_1, \lambda_2, \lambda_3\}$	O_2 $\{\lambda_1, \lambda_2, \lambda_3\}$	O_3 $\{\lambda_1, \lambda_2, \lambda_3\}$	O_4 $\{\lambda_1, \lambda_2, \lambda_3\}$
c_1	{1.48,1.31,1.19}	{1.38,1.27,1.23}	{1.40,1.30,0.50}	{1.41,1.32,1.29}
c_2	{1.24,1.50,1.20}	{1.32,1.05,1.28}	{1.24,1.45,1.11}	{1.44,1.27,1.27}
c_3	{1.45,1.19}	{1.35,1.28,1.35}	{1.34,1.39,1.29}	{1.32,1.42,1.26}
c_4	{1.26,1.34,1.42}	{1.38,1.30,1.34}	{1.36,1.34}	{1.40,1.32,1.31}

The initial CMAA analysis was performed using Monte Carlo simulation with $K_{MC} = 10,000$ samples of the combinatorial space. The rank-1 acceptabilities b_i^1 for the four olive oil suppliers were $(0.06, 0.42, 0.51, 0.01)$ (Table 8.3), with an entropy of 1.35. The maximum entropy for this decision would be 2.0. Ranking the olive oil suppliers according to their rank-1 performance gives $O_3 \succ O_2 \succ O_1 \succ O_4$. The Olympic holistic acceptabilities are $(0.15, 0.39, 0.42, 0.05)$, which gives the same ranking as the rank-1 acceptabilities. Both orders correspond to that of GTOPSIS.

The potential judgement entropies are shown in Table 8.4. Each potential judgement entropy corresponds to the judgements in Table 8.2. The resolution $\lambda_3(1,3) =$ *very good* would improve the entropy significantly to 0.50. By contrast, the potential preference entropies (not shown) are very similar to the current entropy of 1.35, indicating that no preference resolution would have much effect on consensus. However, if the group resolved this discrepancy to $\lambda_3(1,3) =$ *medium poor*, the entropy would worsen slightly to 1.40.

Table 8.5: Entropy-optimal consensus path for the olive oil supplier selection

Step	Act. resolution	h	b_1^1	b_2^1	b_3^1	b_4^1	Potential EO-resolution λ/μ	$\hat{h}$
0	-	1.35	0.06	0.42	0.51	0.01	$\lambda(1,3) = \{\text{VG}\}$	0.50
1	{VG}	0.52	0.01	0.09	0.90	1E-3	$\lambda(2,3) = \{\text{VG}\}$	0.03
2	{VG}	0.03	0.00	3E-3	1.00	0.00	$\lambda(1,2) = \{\text{F}\}$	0.00
3	{F}	0.00	0.00	0.00	1.00	0.00	-	-

The entropy-optimal consensus path for this decision is shown in Table 8.5. Only three steps are required to reach a hard consensus, resulting in O_3 as the most-preferred alternative. Just one entropy-optimal resolution is needed to achieve a soft consensus at $h_{stop} = 0.6$ (for $\tau = 0.3$), with O_3 as the strongest supplier. The remaining instances, which are all inactive at hard consensus, create a combinatorial space of size $K = (2^2 \cdot 3^2) \cdot (2^2 \cdot 3^{11}) = 25,509,168$.

In a simulation experiment, the consensus path lengths with random resolutions were investigated. For the CMAA analysis, $K_{MC} = 10,000$ was chosen, and $10,000$ random consensus paths were simulated. The results are shown in Figure 8.1. The horizontal axis shows the number of steps to consensus, and the vertical axis shows the number of times the simulation found this consensus path length.

The mean consensus path length for a hard consensus ($\tau = 0.0$) is 8.7 steps, and for a firm soft consensus ($\tau = 0.3$) is 4.0. In almost half the samples, a soft consensus could have been achieved with just one or two clarification conferences.

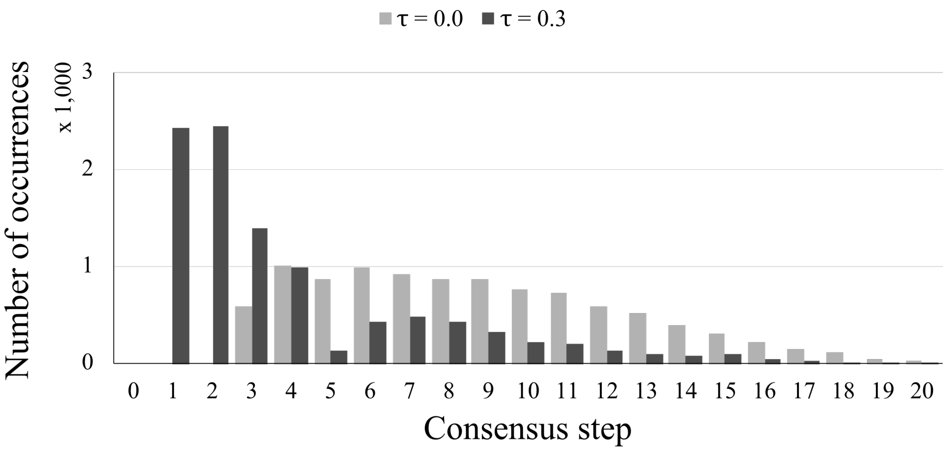


Figure 8.1: Consensus path simulation for the green supplier olive oil decision

8.1.2 Green palm oil supplier selection

Table 8.6: Unified judgement matrix for the palm oil supplier selection

Criteria	Palm Oil Supplier		
	P_1	P_2	P_3
	$\{\lambda_1, \lambda_2, \lambda_3\}$	$\{\lambda_1, \lambda_2, \lambda_3\}$	$\{\lambda_1, \lambda_2, \lambda_3\}$
c_1	{G, VG}	{MG, P}	{F, MG}
c_2	{MG, VG}	{F, MP, P}	{F}
c_3	{G, MG}	{MP, P}	{F, G}
c_4	{G, MG}	{MP, P}	{F, MG, MP}

Table 8.7: Rank acceptability indices for the palm oil suppliers

	CMAA					GTOPSIS	
	b_i^1	b_i^2	b_i^3	acc_i	R_O	CI	R_F
P_1	1.00	0.00	0.00	0.60	1	0.991	1
P_2	0.00	2E-3	1.00	0.10	3	0.000	3
P_3	0.00	1.00	2E-3	0.30	2	0.601	2

Three palm oil suppliers P_1, P_2 and P_3 were under consideration. The unified linguistic judgement matrix of the three decision-makers is shown in Table 8.6. Eleven of the twelve judgements are discrepancies. The GTOPSIS results from the original study are shown in Table 8.7 in the two right columns. Supplier P_1 performs more strongly than P_3, and P_2 has a performance of 0.

The total number of judgement combinations is $||A|| = 2^8 \cdot 3^3 = 6,912$, resulting in a combinatorial space of size $K = 248,832$, so the CMAA analysis was computed with the full combinatorial space. In a runtime simulation, the average time for computing the CMAA analysis for this decision's full combinatorial space with 1,000 replications was 0.5 seconds.

The rank acceptability indices and their corresponding Olympic holistic acceptabilities are shown in Table 8.7. Supplier P_1 has a rank-1 acceptability of 1, so the initial inputs were already at hard consensus; no consensus-building process was required.

CMAA gives a more definitive recommendation than GTOPSIS. This shows a weakness of aggregating preferences and judgements to a single value; it reduces the performance of otherwise good alternatives and increases the performance of otherwise weak alternatives. CMAA provides a much clearer argument for the group; no resolution would change the fact that P_1 ranks first.

It is interesting to note that the input metric CM_I from Equation 2.2 for this decision returns a value of 0.08, which is below the termination threshold. However, it implies that consensus is not yet perfect, although the CAE consensus metric shows that a revelation-invariant and output-based consensus had already been achieved.

8.1.3 Green sunflower and soybean oil supplier selection

Table 8.8: Unified judgement matrix for the sunflower and soybean oil selection decision

Criteria	Sunflower and Soybean Oil Supplier		
	S_1	S_2	S_3
	$\{\lambda_1, \lambda_2, \lambda_3\}$	$\{\lambda_1, \lambda_2, \lambda_3\}$	$\{\lambda_1, \lambda_2, \lambda_3\}$
c_1	{G, VG}	{MG, MP}	{G, P}
c_2	{F, G, MP}	{F, MG, MP}	{MP, P, VG}
c_3	{G, MG, VG}	{MG, VP}	{F, MG, MP}
c_4	{F, MG}	{F, MG, VP}	{F, MG}

Three sunflower and soybean oil suppliers S_1, S_2 and S_3 were being considered. The unified linguistic judgement matrix is shown in Table 8.8. The last two columns of Table 8.9 show the GTOPSIS results from the original study. According to these results, S_1 is clearly the best choice: it receives almost four times the performance value of the second-ranked alternative S_3, which is closely followed by supplier S_2.

All judgements are discrepancies, resulting in $||A|| = 2^6 \cdot 3^6 = 46,656$ and an overall combinatorial space of size $K = 1,679,616$. The initial CMAA analysis was computed with the full combinatorial space. Although the combinatorial space is about 17 times as large as the recommended K_{MC} (Recommendation 1 on page 80) for a Monte Carlo simulation, the runtime is still short enough with 5.5 seconds.

The initial entropy is 0.58, which is already close to the termination threshold of $h_{stop} = 0.48$ for $\tau = 0.3$. Table 8.9 shows the acceptability indices and their Olympic holistic acceptabilities, which yield the same ranking as GTOPSIS.

The potential judgement entropies are shown in Table 8.10. The smallest value is 0.13, located at discrepancy $\lambda(2,1)$, when the group would resolve to judgement $\lambda_2(2,1) = $ *good*. This represents a substantial improvement to the entropy and is below the termination threshold for the consensus iteration. It is interesting to note that the same discrepancy also has the potential to degrade the entropy to 0.85, when the group would resolve to $\lambda_3(2,1) = $ *medium poor*. A similar situation can be found at $\lambda(2,3)$, which also holds the worst achievable value of the entropy of 0.88. By contrast, no resolution of the discrepancy at $\lambda(4,3)$ would have a significant effect on consensus.

Table 8.9: Rank acceptability indices for the sunflower and soybean oil suppliers

	CMAA					GTOPSIS	
	b_i^1	b_i^2	b_i^3	acc_i	R_O		R_F
S_1	0.89	0.10	3E-3	0.57	1	1.000	1
S_2	0.02	0.38	0.60	0.19	3	0.220	3
S_3	0.09	0.51	0.40	0.25	2	0.267	2

Table 8.10: Initial potential judgement entropies for the sunflower and soybean oil suppliers

	S_1 $\{\lambda_1, \lambda_2, \lambda_3\}$	S_2 $\{\lambda_1, \lambda_2, \lambda_3\}$	S_3 $\{\lambda_1, \lambda_2, \lambda_3\}$
c_1	{0.63,0.49}	{0.65,0.45}	{0.80,0.13}
c_2	{0.56,0.13,0.85}	{0.53,0.66,0.46}	{0.22,0.16,0.88}
c_3	{0.52,0.72,0.42}	{0.67,0.42}	{0.56,0.69,0.42}
c_4	{0.65,0.47}	{0.55,0.65,0.45}	{0.51,0.61}

The entropy-optimal consensus path only takes two resolutions to reach a hard consensus. After one consensus step, a very firm soft consensus would have been reached. The entropy-optimal path strengthens the initially strongest supplier S_1 and leads to their selection. After two judgement resolutions, the combinatorial space is of size $K = (2^2 \cdot 3^2) \cdot (2^5 \cdot 3^5) = 279.936$. For all these instances, supplier S_1 is the only one to achieve first rank.

A Monte Carlo simulation of 10,000 CMAA random consensus paths was performed with $K_{MC} = 10,000$ at each step. The histogram of the consensus path lengths is shown in Figure 8.2. The mean consensus path length was 4.7 for a hard consensus and only 1.6 for the firm soft consensus using $\tau = 0.3$.

8.1.4 Conclusions

CMAA was used to recreate three case studies for selecting a green supplier that used FTOPSIS. In the study for selecting the palm oil supplier, CMAA demonstrated that no consensus-building process was required, as the decision already provided a hard, revelation-invariant consensus. Although an input-based consensus metric would have indicated a soft consensus, it does not reflect the same degree of consensus. For a cooperative decision, a soft input-based consensus is misleading, especially when a revelation-invariant consensus has already been reached, which is significantly stricter than an input-based consensus.

For the two other cases, CMAA was able to show that both a soft and a hard consensus could have been reached with a small number of steps. The number of instances remaining at the end of an entropy-optimal consensus path was only about $\frac{1}{27}$ ($\frac{1}{5}$ for random paths) or $\frac{1}{6}$ ($\frac{1}{8}$ for random paths) of the original number of combinations. Thus, Hypothesis 1 on page 10 can be confirmed for these decisions.

8.2 Two selection studies with AHP

CMAA cannot be combined with AHP in its standard form as described in Section 2.2.2. There are three reasons for this:

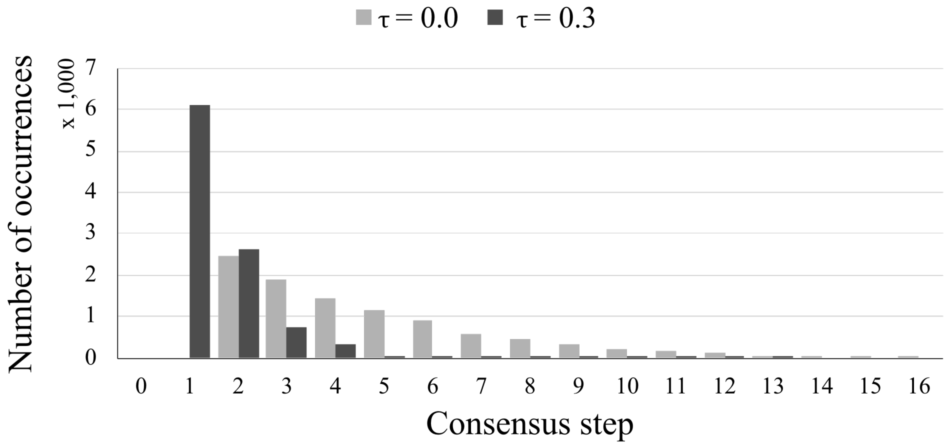


Figure 8.2: Consensus path simulation for the green supplier sunflower and soybean oil study

1. AHP uses pairwise comparisons for preferences and judgements rather than weights and scores.
2. AHP has more than one level of criteria.
3. All criteria on a given level are compared to each other, and all alternatives are compared to each other at the lowest level of the hierarchy.

The CMAA input format must be adapted to accept this additional data.

Furthermore, AHP requires a certain degree of consistency across all pairwise comparisons. In CMAA, the combination of evaluations by different decision-makers to form an instance would destroy this consistency and invalidate the instance.

There have been several approaches to modifying AHP that reduce the number of inputs (Spicelogic Inc., 2022; Abastante et al., 2019; Oliva et al., 2017; Harker, 1987). Harker (1987) introduced a minimalistic version of AHP that requires only the minimum number of pairwise comparisons and simultaneously solves the consistency problem for CMAA.

8.2.1 Minimal AHP

In this section, a special characterisation for deriving priorities from the pairwise comparison matrix is used to generate a minimal pairwise version of AHP that can be used with CMAA.

AHP uses pairwise ratio comparisons to capture relative evaluations of criteria and alternatives in a decision. For these pairwise comparisons, a 17-point scale from $\frac{1}{9}$ to 9 is used, as described in Table 2.3. All pairwise comparisons are captured in pairwise comparison matrices. Each criteria comparison on each hierarchy level generates a pairwise comparison matrix, and alternatives are compared with respect to each set of criteria at the lowest level of the hierarchy. The eigenvectors are computed for each pairwise comparison matrix A, which are called priorities.

If a pairwise ratio comparison matrix $A_{n\times n}$ is positive-reciprocal, then the following is valid:

$$a_{ij} = \frac{1}{a_{ji}}, \text{ for } 1 \leq i, j \leq n \ . \tag{8.1}$$

In AHP, the upper triangular elements of A are the decision-makers' pairwise comparisons, and the lower triangular elements are computed automatically according to Equation 8.1.

The largest eigenvalue of A, λ_1, is unique and real-valued. The vector p is its corresponding eigenvector: $Ap = \lambda_1 p$. The elements of the normalised eigenvector can be used as preferences at the corresponding level of the hierarchy or as judgements at the base of the hierarchy. These values are fed into SAW, the most commonly used decision model in conjunction with AHP.

The pairwise ratio comparison matrix A is *consistent*, when the transient property holds for all its elements:

$$a_{ik} = a_{ij} \cdot a_{jk} \ . \tag{8.2}$$

For a transient matrix A, the following holds (Bebiano et al., 2020):

$$A = pp^{(-1)}, \tag{8.3}$$

where p is the principal right eigenvector of A, and $p^{(-1)}$ denotes the transpose of p containing the reciprocals of each element. So, p can be found simply by reading off any column of A and scaling the resulting vector.

With a set of inputs u_l located at the $(i, i+1)$ matrix elements, a *minimal input matrix* can be created with the following structure:

$$A = \begin{pmatrix} 1 & u_1 & t_{1,3} & \dots & t_{1,n} \\ r_{2,1} & 1 & u_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & t_{n-2,n} \\ \vdots & & \ddots & \ddots & u_{n-1} \\ r_{n,1} & \dots & \dots & r_{n,n-1} & 1 \end{pmatrix} . \tag{8.4}$$

The elements t_{ij} in the upper triangular part of A are the transitive closure of the vector u_l, which forms the superdiagonal. The elements r_{ij} are the reciprocals of the upper triangular elements. This matrix is transitive and consistent. Since the matrix is completely specified by the elements u_l, the vector u can be used as a compact representation of the entire matrix.

Goers et al. (2024) show that the non-normalised priority vector p of a minimal input matrix can be obtained directly from the vector u by the following recursion:

$$p_l = \begin{cases} u_l \cdot p_{l+1}, & 1 \leq l \leq n-1 \\ 1 & l = n \end{cases} . \tag{8.5}$$

The priority vector then can be normalised.

When the minimal input matrix is used to represent decision-maker evaluations, AHP can be coupled with CMAA, which remains essentially unchanged. The pairwise ratio comparisons of the minimal input matrix are used to create the combinatorial space. Each combination now creates a valid instance, because every combination of elements of a minimal input matrix automatically creates a new minimal input matrix.

Small notational changes need to be performed to define the input evaluations. A preference task is denoted by (u_l) in the minimal matrix notation. A judgement task is denoted by (c_j, u_l) in the minimal matrix notation. Pairwise preference comparisons will be noted as $\mu_k(l)$, and pairwise judgement comparisons will be denoted as $\lambda_k(j, l)$. All other variables in Section 3.2 need to be adapted accordingly.

8.2.2 Site selection for aquaculture at the Canadian coast

In this experiment, a previously published case study that used single-step AHP for a group decision is analysed with the CMAA framework, and its consensus-building potential is examined. This study was part of a Bachelor's thesis of Eckardt (2023), from which some of the results in this section are taken.

Ozer (2007) presents a study about selecting a site for expending aquaculture in the coastal zone of Grand Manan Island in Canada. The study was conducted with five stakeholder groups in the region:

DM_1 Federal Scientists

DM_2 Provincial Government

DM_3 Non-Governmental Organisations

DM_4 Industrial Organisations

DM_5 Local Communities

Within each group, individual evaluations were collected and averaged, resulting in single input values for each group.

The four alternatives evaluated were a combination of two coastal areas and two aquaculture operations:

a_1 A1FF: Area 1 with fishing farms

a_2 A2FF: Area 2 with fishing farms

a_3 A1noFF: Area 1 without fishing farms

a_4 A2noFF: Area 2 without fishing farms

The criteria hierarchy originally applied consisted of four level-1 criteria. Each level-1 criterion c_1 *Resources*, c_2 *Habitat*, c_3 *Effluents* and c_4 *Activities* considers several level-2 criteria. The priorities of all level-2 criteria were pre-computed, so only the first-level criteria needed to be considered.

Table 8.11: The unified preference vector for the aquaculture study

	μ_1	μ_2	μ_3	μ_4
u_1: Resources vs Habitat	1.05	1.07	1.22	1.64
u_2: Habitat vs Effluent	1.80	1.96	3.11	4.29
u_3: Effluent vs Activities	0.20	0.25	0.55	1.35

The preference and judgement matrices from the study were converted to their equivalent minimal forms, which are shown in Table 8.11 and Table 8.12, respectively. Note that the values are not integers. This is for two reasons. First, the data was only available from the study containing the averages across the groups of decision-makers. Second, the preference and judgement matrices from the study were converted to their equivalent minimal forms. In total, this decision contained eleven judgement discrepancies and three preference discrepancies. The number of instances in the combinatorial space is $K = 4^3 \cdot (3^2 \cdot 4^3 \cdot 5^6) = 576,000,000$.

The initial CMAA analysis was computed using Monte Carlo simulation with $K_{MC} = 10,000$ random instances. Table 8.13 shows the rank-1 acceptabilities, the Olympic holistic

Table 8.12: The unified judgements for the aquaculture study

	A1FF vs A2FF $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$	A2FF vs A1noFF $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$	A1noFF vs A2noFF $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$
c_1	{0.14,0.16,0.26,0.27,0.39}	{3.05,4.30,4.54,7.41,8.36}	{0.12,0.14,0.22,0.23,0.33}
c_2	{249.50, 499.00, 24,950.00}	{4E-5,2E-3,4E-3}	{500.00}
c_3	{0.84,0.85,1.01,1.09}	{0.92,0.99,1.18,1.19}	{0.84,0.85,1.01,1.09}
c_4	{0.22,0.28,0.52,0.59,0.60}	{1.98,2.09,2.70,2.90,3.04}	{0.33,0.34,0.37,0.48,0.50}

Table 8.13: Comparison of results for the aquaculture decision

	CMAA-AHP						GAHP	
	b_i^1	b_i^2	b_i^3	b_i^4	acc_i	R$_C$	Pri.	R$_G$
a_1: A1FF	0.276	0.167	0.164	0.393	0.232	3	0.2484	3
a_2: A2FF	0.227	0.373	0.323	0.077	0.280	2	0.2546	2
a_3: A1noFF	0.229	0.143	0.195	0.433	0.200	4	0.2422	4
a_4: A2noFF	0.268	0.318	0.318	0.097	0.288	1	0.2548	1

acceptabilities and the performance according to the case study (GAHP aggregates each conflict using the geometric mean). R_C denotes the ranking positions of each alternative based on the holistic acceptability, and R_G denotes the ranking based on the GAHP priorities. Both approaches computed the same ranking of A2noFF $\succ$ A2FF $\succ$ A1FF $\succ$ A1noFF. Because the GAHP priorities are so close, a recommendation should not be provided based on them. The rank-1 acceptabilities are also very close, resulting in an entropy of 1.99, which is very close to the maximum of 2.0 for four alternatives.

The potential preference entropies are shown in Table 8.14, and the potential judgement entropies are shown in Table 8.15. Most values are equal to or little better than the initial current entropy of 1.99. The lowest potential entropy would be obtained if the judgement discrepancy at $\lambda(2,1)$ was resolved to $24,950$, giving a potential entropy of 1.45.

In the original study, AHP applying the geometric average was unable to generate a recommendation, because the resulting priorities were too close (see Table 8.13). A consensus path simulation shows that CMAA-AHP would have been able to create consensus with little additional clarification effort. In Table 8.16, the entropy-optimal consensus path is presented. It would have taken just three clarification steps to achieve a very firm soft consensus (indicated with dashed line) and just four steps to reach a hard consensus. A Monte Carlo simulation using $1,000$ different consensus paths and a random resolution of discrepancies revealed that the expected path length for a hard consensus contained 7.86

Table 8.14: Initial potential preference entropies for the aquaculture study

	μ_1	μ_2	μ_3	μ_4
u_1: Resources vs Habitat	1.96	1.97	1.99	1.90
u_2: Habitat vs Effluent	1.96	1.98	1.98	1.94
u_3: Effluent vs Activities	1.86	1.94	1.96	1.80

Table 8.15: Initial potential judgement entropies for the aquaculture study

	A1FF vs A2FF $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$	A2FF vs A1noFF $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$	A1noFF vs A2noFF $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$
c_1	{1.99,1.99,1.99,1.99,1.94}	{1.70,1.93,1.97,1.91,1.80}	{1.70,1.86,1.98,1.96,1.78}
c_2	{1.75,1.98,1.45}	{1.62,1.93,1.60}	-
c_3	{2.00,2.00,1.99,1.99}	{1.99,1.99,1.99,1.99}	{1.99,1.99,2.00,2.00}
c_4	{1.98,1.99,1.99,1.98,1.97}	{1.92,1.95,2.00,1.99,1.97}	{1.95,1.96,1.99,1.98,1.97}

Table 8.16: Entropy-optimal consensus path for the aquaculture study

Step	Act. resolution	h	b_1^1	b_2^1	b_3^1	b_4^1	Potential EO-resolution λ/μ	$\hat{h}$
0	-	1.99	0.28	0.23	0.23	0.27	$\lambda(2,1) = \{24{,}950\}$	1.44
1	{24,950}	1.33	0.63	0.16	9E-4	0.21	$\lambda(2,2) = \{4\text{E-}3\}$	0.81
2	{4E-3}	0.81	0.83	0.06	0.00	0.11	$\mu(3) = \{1.35\}$	0.09
3	{1.35}	0.10	0.99	3E-4	0.00	0.01	$\mu(1) = \{1.05\}$	0.00
4	{1.05}	0.00	1.00	0.00	0.00	0.00	-	-

clarification steps.

8.2.3 Approving loans for agricultural irrigation equipment in Serbia

In this section, CMAA is used to analyse a case study of a group decision by Srdjevic et al. (2011). Some of the results are from the Bachelor's thesis of Eckardt (2023).

The case study describes a decision by the board of a Serbian fund for agricultural loans. Five applications had been made for financial loans for irrigation equipment, and the decision had to be made which applicant should be awarded a loan. Five decision-makers were on the board:

DM_1 President of the Fund Council

DM_2 Senior advisor to the Fund

DM_3 Fund manager

DM_4 External expert advisor

DM_5 Representative of the Ministry of Agriculture

The three evaluation criteria were:

c_1 Complete presentation of service costs

c_2 Loan history

c_3 Insurance coverage

The pairwise preferences according to the minimal AHP matrix structure were determined from the original study data and are shown in Table 8.17. The number of preference instances is $||\mathbf{P}|| = 5^2 = 25$.

Table 8.17: The unified preference vector for the agricultural loan study

	μ_1	μ_2	μ_3	μ_4	μ_5
u_1: Service costs vs Loan history	0.13	1.30	2.33	3.34	4.46
u_2: Loan history vs Insurance coverage	3.15	3.88	3.92	4.04	5.86

Table 8.18: The unified judgements for the agriculture loan study

	P_1 vs P_2 $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$	P_2 vs P_3 $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$
c_1	{1.62,1.81,1.83,1.94,2.16}	{0.24,0.26,0.27,0.30,0.38}
c_2	{0.07,0.10,0.10,0.17,0.19}	{1}
c_3	{1.46,1.59,1.96,2.26,2.58}	{1.57,1,68,2.20,2.46,2.71}

	P_3 vs P_4 $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$	P_4 vs P_5 $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$
c_1	{14.40,14.68,15.83,16.03,18.77}	{0.39,0.41,0.48,0.56,0.57}
c_2	{2.27,2.39,2.96,2.98,6.27}	{0.47,0.53,0.60,0.66,0.75}
c_3	{0.12,0.13,0.15,0.15,0.18}	{0.32,0.39,0.43,0.45,0.59}

Five producers $P_1 \dots P_5$ applied for the agricultural loans. Table 8.18 shows the pairwise judgement comparisons. These are the converted values from the original data into the minimal input matrix. The number of judgement instances is $||\mathbf{A}|| = 5^{11} = 48,828,125$, so the overall size of the combinatorial space is $K = 1,220,703,125$. CMAA analysis was conducted using the Monte Carlo simulation with $K_{MC} = 10,000$ randomly generated instances.

Table 8.19: Comparison of results for the agricultural loan decision

	CMAA-AHP							GAHP	
	b_i^1	b_i^2	b_i^3	b_i^4	b_i^5	acc_i	R$_C$	Pri.	R$_G$
P_1	0.000	0.503	0.203	0.111	0.183	0.171	3	0.168	3
P_2	0.000	0.413	0.585	0.002	0.000	0.182	2	0.208	2
P_3	0.964	0.036	0.000	0.000	0.000	0.589	1	0.394	1
P_4	0.000	0.000	0.000	0.183	0.817	0.000	5	0.078	5
P_5	0.036	0.048	0.212	0.704	0.000	0.057	4	0.153	4

In Table 8.19, the results of the initial computations for CMAA and GAHP are shown. The ranking R_C for CMAA was computed using Olympic holistic acceptabilities, and the ranking R_G was derived from the priorities produced by GAHP. The two rankings are identical. The most-preferred alternative produced by CMAA is already very clear and very close to consensus, with 96.4% of all instances yielding P_3 as the most-preferred applicant. The current entropy of the initial decision is very low with 0.23, which is well below the very firm soft consensus threshold of $h_{stop} = 0.70$ ($h_{max} = 2.32$). CMAA analysis shows that P_1, P_2 and P_4 cannot rank first. The priority vector of GAHP also shows P_3 with the highest priority, but the recommendation is less convincing. Thus, the initial CMAA analysis provides additional valuable information compared to the averaging approach.

Table 8.20: The potential preference entropies for the agriculture loan study

	μ_1	μ_2	μ_3	μ_4	μ_5
u_1: Service costs vs Loan history	0.68	0.00	0.00	0.00	0.00
u_2: Loan history vs Insurance coverage	0.35	0.23	0.22	0.20	0.08

Table 8.21: The potential judgement entropies for the agriculture loan study

	P_1 vs P_2 $\{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5\}$	P_2 vs P_3 $\{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5\}$
c_1	{0.22,0.23,0.22,0.22,0.23}	{0.21,0.22,0.21,0.23,0.25}
c_2	{0.23,0.22,0.22,0.23,0.23}	-
c_3	{0.22,0.23,0.22,0.22,0.22}	{0.24,0.25,0.22,0.21,0.19}

	P_3 vs P_4 $\{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5\}$	P_4 vs P_5 $\{\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5\}$
c_1	{0.23,0.22,0.23,0.22,0.22}	{0.23,0.23,0.23,0.21,0.22}
c_2	{0.43,0.36,0.10,0.09,0.00}	{0.44,0.32,0.16,0.07,0.01}
c_3	{0.24,0.24,0.23,0.21,0.19}	{0.30,0.24,0.22,0.21,0.15}

The potential preference and judgement entropies are shown in Tables 8.20 and 8.21, respectively. A hard consensus for P_3 can be achieved in just one step with four different preference resolutions and one judgement resolution (for $\lambda(2,3) = 6.27$). All five resolutions would result in twelve inactive discrepancies, with $244,140,625$ remaining instances that all return the same most-preferred alternative.

Table 8.22: Pairwise comparison u_1 comparing c_1 Service costs with c_2 Loan history

	$\mu_k(1)$	$\hat{h}(\mu(1))$
DM_1 President	1.304	0.000
DM_2 Senior advisor	3.339	0.000
DM_3 Fund manager	0.130	0.684
DM_4 External expert	2.326	0.000
DM_5 Repr. ministry	4.462	0.000

In Table 8.22, the potential entropies for the preference discrepancy $\mu(1)$ are shown. The evaluations of four decision-makers would lead to a hard consensus. However, agreement on the evaluation of 0.130 by the Fund manager DM_3 would increase the entropy to 0.684. The Fund manager is the only decision-maker for whom Loan history c_2 is more important than Service costs c_1. If the Fund manager held decisive information that would change the evaluations of all other decision-makers, it would decrease P_3's rank-1 acceptability to 0.82 and increase P_4's rank-1 acceptability to 0.18, while all other rank-1 acceptabilities would remain the same. This is an example of the potential value of minority dissent (Nijstad et al., 2014). With averaging, this possibility would probably have been overlooked.

A Monte Carlo simulation using 1,000 random consensus paths showed that the mean consensus path length for this decision is 1.97 steps. The 11 remaining inactive discrepancies create about 48 million instances, all of which have no effect on the most-preferred applicant.

8.3 Objective measurements and subjective threshold decisions

In this section, a new decision method is proposed, which enhances consensus-building significantly (Goers and Horton, 2024c) for a special class of decisions. The condition is that objective measurements are available for all alternative/criterion-pairs.

8.3.1 Satisfaction threshold-based decision method

Table 8.23: Linguistic variables for FTOPSIS example

Preferences		Judgements	
Linguistic value	TFN	Linguistic value	TFN
very low (VL)	(0.0, 0.10, 0.2)	very poor (VP)	(0, 1.0, 2)
low (L)	(0.1, 0.20, 0.3)	poor (P)	(1, 2.0, 3)
medium low (ML)	(0.2, 0.35, 0.5)	medium poor (MP)	(2, 3.5, 5)
medium(M)	(0.4, 0.50, 0.6)	fair (F)	(4, 5.0, 6)
medium high (MH)	(0.5, 0.65, 0.8)	medium good (MG)	(5, 6.5, 8)
high (H)	(0.7, 0.80, 0.9)	good (G)	(7, 8.0, 9)
very high (VH)	(0.8, 0.90, 1.0)	very good (VG)	(8, 9.0, 10)

Often, decisions are made on the basis of subjective judgements, even when objective performance measurements are available. These approaches will be referred to as OMSJ (objective measurement, subjective judgement) methods. This section introduces a new approach to modelling OMSJ decisions that can be combined with CMAA and studies its application for three previously published case studies from engineering and forestry. Some of the results are taken from a Master's thesis by Rolle (2022).

In an OMSJ decision, decision-makers input their subjective evaluations of the value or utility of an alternative's measured performance. For example, in an engineering decision about a machine, a decision-maker might judge a power consumption of 10kW and one of 12kW to both be *very good*. They are equivalent for the purposes of the decision, and there is no need to distinguish between them. An OMSJ approach is unavoidable for methods such as AHP, which can only accept subjective judgements. An example can be found in Srdjevic et al. (2011).

One reason for using an OMSJ approach is that the relationship between an objective measurement and the value it provides may not be linear (Salo and Hämäläinen, 1997). In the decision of Section 8.2.3, objective data was available on each applicant, which decision-makers interpreted with pairwise comparison judgements. Another example is a case study by Noori et al. (2018), who use FTOPSIS with a group for deciding a site for a dam in Iran. The decision-makers were provided with objective measurements for each site

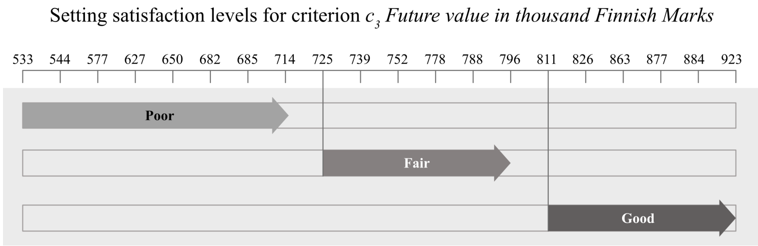


Figure 8.3: Illustration of subjective threshold selection in the forestry study

alternative regarding important criteria such as cost, reservoir safety and water quality. Then the decision-makers delivered their evaluations in the form of linguistic terms as described in Table 8.23.

When OMSJ methods are applied in groups, there is a risk of creating inconsistencies between judgements. These inconsistencies can be prevented, if the format for providing subjective judgements is changed. Instead of judging each criterion/alternative (c_j, a_i) pair, decision-makers choose thresholds that separate the measurements into satisfaction levels, which are then applied to all alternatives automatically.

Figure 8.3 illustrates setting satisfaction levels for all objective measurements in the case study in Section 8.3.3. This example shows 20 objective measurements and three satisfaction levels *poor*, *fair*, and *good* for the criterion *Future value* in thousands of Finnish Marks. The decision-makers set the thresholds that separate *poor* from *fair* at 725 Marks and *fair* from *good* at 811 Marks, thereby assigning each of the 20 measurements to a subjective evaluation.

Table 8.24: Intra-decision-maker inconsistencies

		Measurements		
		$M_1 < M_2$	$M_1 = M_2$	$M_1 > M_2$
	$g_1 \prec g_2$	C	I	I
Levels	$g_1 \equiv g_2$	C	C	C
	$g_1 \succ g_2$	I	I	C

Subjective estimates can create both intra-decision-maker and inter-decision-maker judgement inconsistencies, which can decrease the reliability of the recommendation. Intra-decision-maker inconsistencies are created when two alternatives with the same objective measurement receive different satisfaction levels from a single decision-maker. In Table 8.24, M_1 and M_2 represent the measurements of two alternatives a_1 and a_2, and g_1 and g_2 represent the satisfaction levels to which they have been assigned. Inconsistencies (I) occur when alternatives with the same measurement are assigned to different satisfaction levels or when an alternative with a superior measurement is assigned to an inferior satisfaction level.

Inter-decision-maker inconsistencies occur when two or more decision-makers submit contradictory judgements, as illustrated in Table 8.25. Judgements g_1 and g_2 by decision-makers DM_1 and DM_2 with respect to alternatives a_1 and a_2 contradict each other. For example, it is inconsistent for one decision-maker to judge a ranking position of 1 to be ‘good’ and a ranking position of 5 to be ‘fair’, while another decision-maker judges ranking position 1 to be ‘fair’ and position 5 to be ‘good’. (If the ranking measures a quality that is unambiguously beneficial or detrimental, then one of the decision-makers is also committing

		DM$_1$		
		$g_1 \prec g_2$	$g_1 \equiv g_2$	$g_1 \succ g_2$
	$g_1 \prec g_2$	C	C	I
DM$_2$	$g_1 \equiv g_2$	C	C	C
	$g_1 \succ g_2$	I	C	C

Table 8.25: Inter-decision-maker inconsistency

an intra-decision-maker inconsistency.)

These inconsistencies can be prevented, if the subjective judgements are based on measurement-satisfaction levels rather than criteria/alternative pairs.

8.3.2 Notation and adaptations

In an OMSJ decision, preferences have the standard format and can be treated in the usual way by CMAA.

Objective performance measurements are denoted by the matrix $\mathbf{M}_{n\times m}$, and the corresponding matrix of subjective satisfaction levels is denoted by $\mathbf{G}_{n\times m}$. The number of satisfaction levels is $Z+1$. Then, the (subjective threshold) judgements for criterion c_j create a vector $[t_{j1}, ..., t_{jZ}]$, using the units for the corresponding criterion. The threshold judgements for criterion c_j are denoted by $\lambda_k(j)$. Each threshold defines the boundary between two satisfaction levels, where satisfaction level $z+1$ is more valuable than satisfaction level z. The threshold values are sorted in ascending order for beneficial criteria $[t_{jz} \preceq t_{j(z+1)}]$ and in descending order for non-beneficial criteria $[t_{j(z+1)} \preceq t_{jz}]$. It is recommended to use semantic levels such as 'Fair' or 'Very Poor', rather than numerical values. A typical number of satisfaction levels might be the five-point Likert scale with $Z+1=5$.

For a beneficial criterion c_j, the satisfaction matrix $\mathbf{G}$ is defined as follows:

$$\mathbf{G}_{ji} = \begin{cases} 1, & \text{if } \mathbf{M}_{j,i} \prec t_{j,1} \\ z, & \text{if } t_{j,z-1} \preccurlyeq \mathbf{M}_{j,i} \prec t_{j,z} \\ Z+1, & \text{if } \mathbf{M}_{j,i} \succcurlyeq t_{j,Z} \end{cases} . \tag{8.6}$$

It is reversed for cost criteria accordingly.

For OMSJ decisions, judgement tasks will be denoted by (c_j) and subjective threshold judgements by $\lambda_k(j)$, where the index j refers to the criterion and all subjective thresholds are set by the decision-maker. All judgements are collected in the *unified judgement matrix* $\mathbf{A}$ in the usual manner. The matrix of subjective thresholds $\mathbf{A}$ has dimension $n \times 1$ and is referred to as the *unified judgement vector*. Here, each element in $\mathbf{A}_j$ collects all individual judgements $\lambda_k(j)$ for a criterion c_j:

$$\mathbf{A}_j = \bigcup_k \lambda_k(j),$$

where k is the local index to each criterion. When two decision-makers submit different threshold judgements $\lambda_{k1}(j) \neq \lambda_{k2}(j)$, a discrepancy is present. The cardinality ϕ_j for each judgement task equals the number of differing judgements.

An instance $\mathbf{A}^*$ of $\mathbf{A}$ is obtained by selecting one $\lambda_k(j)$ for each criterion c_j. This instance is used to create the satisfaction matrix $\mathbf{G}$ according to Equation 8.6. An example of a unified judgement vector is illustrated in Table 8.27.

Table 8.26: The triangular fuzzy evaluations

Preferences		Judgements	
Very Low (VL)	(0, 0, 2)	Poor (P)	(0, 0, 1)
Low (L)	(1, 3, 5)	Medium (M)	(0, 1, 2)
Medium (M)	(3, 5, 7)	Good (G)	(1, 2, 2)
High (H)	(5, 7, 9)	-	-
Very High (VH)	(8, 10, 10)	-	-

Table 8.27: Example unified judgement vector **A** with an example instance $\mathbf{A}^*$

	$\{\lambda_1,$	$\lambda_2,$	$\lambda_3\}$	$\mathbf{A}^*$
Net income	{[105, 133],	[73, 123],	[87, 102]}	[105, 133]
Biodiversity	{[0.16, 0.28],	[0.44, 1.11],	[0.42, 1.11]}	[0.16, 0.28]
Future value	{[544, 826],	[752, 884],	[796, 923]}	[752, 884]
Scenery	{[5.74, 6.05],	[6.11, 6.24],	[6.04, 6.22]}	[6.04, 6.22]
Blueberries	{[6, 9.1],	[5.1, 8.2],	[5.2, 8.2]}	[6, 9.1]

Algorithm 3: Initial threshold combinatorial acceptability analysis

Given: Decision algorithm
Input: **M**, all $\lambda_k(j)$ and $\mu_k(j)$
Construct **P** and **A**;
for *each instance* $[\boldsymbol{P}^*; \boldsymbol{A}^*]$ **do**
  Construct **G** using $\mathbf{A}^*$ and **M**;
  Apply the decision algorithm to $[\mathbf{P}^*; \mathbf{G}]$;
  Update counter variables;
end
Compute statistics from the counter variables;
Output: Acceptability indices

The CMAA analysis algorithm needs to be modified slightly to account for the generation of the satisfaction matrix $\mathbf{G}$ for each instance, as shown in Algorithm 3. Two minor changes have been made. Apart from the different formats of the judgements to $\lambda_k(j)$, the first change was added at the beginning of the loop to construct $\mathbf{G}$ from $\mathbf{A}^*$ and the objective measurement matrix $\mathbf{M}$. And the decision method is applied to $[\mathbf{P}^*; \mathbf{G}]$, instead of $[\mathbf{P}^*; \mathbf{A}^*]$. The remainder of the algorithm is identical to Algorithm 1.

8.3.3 Selection of a forest management plan in Finland

Table 8.28: Objective measurements for a forest planning decision (Kangas et al., 2006)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Net income	40	15	97	152	102	45	73	135	66	154
Biodiversity	1.14	1.44	0.42	0.12	0.41	1.11	0.44	0.14	0.70	0.12
Future value	877	923	739	544	752	863	778	627	826	533
Scenery	6.24	6.30	6.05	5.72	5.86	6.22	6.19	5.32	6.11	5.36
Blueberries	10.7	11.6	8.2	5.2	8.0	10.6	8.7	6.0	9.6	5.1

	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
Net income	87	133	37	91	82	115	123	62	105	95
Biodiversity	0.47	0.17	1.11	0.28	0.54	0.29	0.16	0.44	0.23	0.27
Future value	796	650	884	725	788	682	577	811	685	714
Scenery	5.91	5.74	6.24	6.12	6.05	5.99	6.04	6.24	6.05	6.11
Blueberries	8.7	6.0	10.8	7.3	8.9	6.8	5.9	9.1	6.7	7.3

The decision was to decide how to manage an area of forest in Finland that was owned by three different stakeholders who each had different priorities. This decision has been studied by Kangas and Kangas (2003), Laukkanen et al. (2005) and Kangas et al. (2006). The decision methods used were SMAA-2, SMAA-O in conjunction with Multicriteria Approval, AHP and PROMETHEE.

The following five criteria were considered:

- c_1 Net income: income from timber production in thousands of Finnish Marks (TFIM)
- c_2 Biodiversity: provided as an expert ranking
- c_3 Future value: estimated value of timber production over 20 years in TFIM
- c_4 Scenery: scenic beauty, provided as an expert ranking
- c_5 Blueberries: the yield of blueberries in kilograms per hectare (kg/ha)

The preference vector $\mathbf{P}$ of this decision is (VH, H, M, L, VL) for criteria c_1 to c_5, respectively. It does not contain any discrepancies. The triangular fuzzy representations of these linguistic terms are shown in Table 8.26. The decision contains 20 different management proposals P1 to P20. Their objective measurements are shown in Table 8.28, which have been abbreviated and rounded for clarity. Alternative P2 has the best performance for four out of the five criteria.

Table 8.29: Part of the satisfaction level matrix **G** for the example instance in Table 8.26

	P1	P2	P3	P4	P5	···	P20
Net income	P	P	P	G	P	···	P
Biodiversity	G	G	G	P	G	···	M
Future value	M	G	P	P	P	···	P
Scenery	G	G	M	P	P	···	M
Blueberries	G	G	M	P	M	···	M

To perform the comparison, CMAA was used with the Fuzzy SAW decision model, with the three satisfaction levels *poor*, *medium* and *good* for the objective measurements. An example of a unified judgement vector **A** is shown in Table 8.27, as well as one instance of a judgement vector $\mathbf{A}^*$ that was derived from it. Each vector contains a set of judgements $\{\lambda_1, \lambda_2, \lambda_3\}$ for the three decision-makers. Each judgement contains the two measurement thresholds that separate the three satisfaction levels. The size of the combinatorial space is $K = 1^5 \cdot 3^5 = 243$. From the instance $\mathbf{A}^*$ in Table 8.27, the satisfaction level matrix **G** was generated, which is shown in part in Table 8.29.

Table 8.30: Initial acceptabilities using judgement thresholds from Table 8.27

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
b_i^1	0.198	0.708	0.012	0.000	0.111	0.123	0.074	0.000	0.000	0.000
acc_i	0.237	0.458	0.017	0.000	0.078	0.164	0.061	0.000	0.001	0.000
Ranks (acc_i)	2	1	13	17	4	3	5	17	16	17

	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
b_i^1	0.012	0.074	0.025	0.012	0.074	0.074	0.029	0.000	0.025	0.000
acc_i	0.035	0.059	0.039	0.016	0.058	0.060	0.026	0.002	0.021	0.000
Ranks (acc_i)	10	7	9	14	8	6	11	15	12	17

Table 8.31: Initial judgement entropies for the artificial judgement thresholds

	λ_1	λ_2	λ_3
Net income	1.818	2.638	2.207
Biodiversity	3.016	1.587	1.896
Future value	2.931	1.876	1.463
Scenery	3.045	2.061	2.464
Blueberries	2.714	2.779	2.779

The CMAA analysis for the subjective judgement threshold yields the the rank-1 acceptabilities for the 20 alternatives shown in Table 8.30. Six alternatives (P4, P8, P9, P10, P18, P20) can be safely removed from the decision, because their rank-1 acceptabilities are 0.0, and their Olympic holistic acceptabilities are equal to, or very close to 0.0. The rank-1 acceptability and the holistic acceptability both produce the same ranking of the top-five alternatives: P2 $\succ$ P1 $\succ$ P6 $\succ$ P5 $\succ$ P7. The entropy is 2.76, which is quite low, because the maximum possible entropy for 20 alternatives is 4.32.

The initial judgement entropies for the artificial judgement thresholds are shown in Table 8.31. The judgement threshold at $\lambda_3(3)$ contains the resolution with the largest entropy reduction to 1.463, which is one third of the initial entropy and therefore already very close to h_{stop}.

A Monte Carlo simulation was performed to study the effect of the number of thresholds. The number of thresholds ranged from $Z = 1$ to $Z = 4$. The case $Z = 1$ corresponds to a simple pass/fail judgement, while $Z = 4$ corresponds to the typical five-point Likert scale. For each number of thresholds, $1,000$ random judgements were generated, and an entropy-optimal consensus path was computed.

The mean path lengths for the four different numbers of thresholds $z \in \{1, 2, 3, 4\}$ were very close, ranging from 2.3 to 2.2, with a slight tendency towards shorter path lengths as more thresholds are applied. The maximum number of discrepancies for the forestry planning example is equal to five: the number of criteria. The mean path length is about 2.2 for all values of Z. Adding judgement thresholds increases the probability of creating multiple rank-1 alternatives; the simulation calculated an average of 1.5 rank-1 alternatives at the end of the consensus paths.

By using thresholds rather than scores, the number of judgement discrepancies is reduced from 100 down to five. Furthermore, the threshold judgements reduce the risk of inconsistencies in the decision-maker inputs.

8.3.4 Selection of a wind turbine system

Table 8.32: The objective measurements for the four wind turbine alternatives in one area

Parameters	Nordex N80	Vestas V90	GE 3.6	REpower M5
	a_1	a_2	a_3	a_4
c_1 Rated Power in MW	2.5	3.0	3.6	5.0
c_2 Max capacity in MW	100.0	96.0	86.4	80.0
c_3 Net energy in 10^3MWh	356.4	360.2	334.0	327.7
c_4 Investment in 10^6Euro	359.0	341.1	314.2	290.4
c_5 CO_2 emissions in 10^3t	223.1	225.5	209.1	205.1

Bagočius et al. (2014) studied how a group of ten experts selected the best offshore wind turbine for a location in the Baltic Sea off the coast of Lithuania. The original study solved this problem with the WPM decision method (see Section 2.2.2). Five criteria were used to determine the selection, which were all in discrepancy and were applied as-is in the following studies. Criteria c_1 to c_3 were beneficial criteria and criteria c_4 and c_5 were non-beneficial criteria. The number of distinct preference vectors is $||\mathbf{P}|| = 10^5 = 100,000$. Four alternatives were considered. Their objective measurements are shown in Table 8.32. For the following simulations the decision method SAW was applied (see Section 2.2.2).

Initially, a random set of threshold judgements was generated for the ten decision-makers, as shown in Table 8.33. Two thresholds were used. The decision contains five preference and five threshold judgement discrepancies. Applying the CMAA analysis Algorithm 1

Table 8.33: Random threshold judgements

Criteria	λ_1	λ_2	λ_3	λ_4	λ_5
c_1 Rated power	{[3.6,3.6]	[3.6,3.0]	[3.6,2.5]	[5.0,2.5]	[5.0,3.0]
c_2 Max capacity	{[86,86]	[100,100]	[96,80]	[96,86]	[100,86]
c_3 Net energy	{[360,328]	[360,334]	[360,360]	[356,328]	[360,334]
c_4 Investment	{[359,359]	[314,314]	[290,314]	[314,341]	[290,290]
c_5 CO_2 emissions	{[205,209]	[209,223]	[209,209]	[205,223]	[223,226]

Criteria	λ_6	λ_7	λ_8	λ_9	λ_{10}
c_1 Rated power	[3.6,2.5]	[5.0,3.0]	[5.0,2.5]	[3.6,2.5]	[5.0,3.0]}
c_2 Max capacity	[96,96]	[96,96]	[80,80]	[86,86]	[100,80]}
c_3 Net energy	[356,334]	[356,328]	[360,334]	[360,328]	[360,328]}
c_4 Investment	[314,341]	[290,359]	[290,359]	[314,359]	[290,314]}
c_5 CO_2 emissions	[209,226]	[209,226]	[209,209]	[205,226]	[209,226]}

Table 8.34: Initial rank acceptability index for the random threshold judgements

	b_i^1	b_i^2	b_i^3	b_i^4	acc_i	R_1	R_{acc}	R_O
a_1	0.120	0.167	0.251	0.462	0.147	4	4	4
a_2	0.293	0.265	0.303	0.140	0.285	2	2	3
a_1	0.265	0.291	0.259	0.185	0.272	3	3	2
a_1	0.392	0.256	0.178	0.174	0.330	1	1	1

with a Monte Carlo simulation for $K_{MC} = 10,000$, this set of judgements yielded the rank acceptability indices, the rank-1 order R_1, the Olympic holistic acceptability rank order R_{acc}, and the rank order of the original study R_O, as shown in Table 8.34. Ranks 2 and 3 diverged from the original study, but their performance was close with both CMAA and in the original study. By using entropy-optimal resolutions, a hard consensus was achieved in three consensus steps, involving two threshold judgement resolutions and one preference resolution.

To review this study more comprehensively, a consensus path simulation was conducted. Threshold judgements with two thresholds were independently generated, creating a threshold judgement space of $||\mathbf{A}|| = 10^5 = 100,000$. Consequently, the size of the combinatorial space is $K = 100,000 \cdot 100,000 = 1E10$. The decision was simulated with 1,000 different threshold judgement sets. For each set, one consensus path was computed, varying the probability that each discrepancy would be resolved as entropy-optimal or not, expressed with $P(EO)$. For $P(EO) = 1.0$, all discrepancies are resolved entropy-optimally. For the CMAA analysis a Monte Carlo simulation was applied with $K_{MC} = 10,000$. The termination consensus was $\tau = 0.0$, representing a hard consensus.

The mean consensus path lengths for varying $P(EO)$ in this study are shown in Table 8.35. In the best-case scenario, where $P(EO) = 1.0$ the average consensus path length is 3.5, which is fast. Even in the opposite case, where the entropy-optimal resolution is never chosen, the path length averages 7.4. The maximum number of discrepancies for this problem is ten: five preference discrepancies and five judgement threshold discrepancies. In a typical

Table 8.35: Mean consensus path lengths for wind turbine selection

P(EO)										
0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
7.4	6.8	6.4	5.9	5.5	5.1	4.7	4.3	4.1	3.8	3.5

subjective judgement situation, decision-makers would evaluate all 20 judgement tasks, potentially creating 20 judgement discrepancies. Adding the five preference discrepancies, the decision would have consisted of 25 discrepancy evaluation tasks. Comparing the mean consensus path lengths to these 25 discrepancies, it is evident that the consensus path lengths are very efficient. This efficiency is particularly notable, considering that for $P(EO) = 0.0$, the entropy-optimal resolution is never chosen, but even the mean consensus path length of 7.4 for $P(EO) = 1.0$ would be still considered very efficient. This represents a significant advantage of threshold judgements compared to criterion/alternative-pair judgements.

8.3.5 Selecting the best setting for composite drilling

Osmond et al. (2021) used a multi-criteria method to choose machine settings for drilling holes into composite textiles, with the goal of minimizing the damage caused to the material. The decision was composed of eight alternatives, three criteria and three decision-makers. The original study applied the three decision methods TOPSIS, WPM and SAW. For the CMAA simulation, SAW was used.

The criteria describing the quality of the drilled hole were evaluated by the decision-makers. Only criterion c_1 is beneficial, while all others are non-beneficial. There were two criteria discrepancies, resulting in $||\mathbf{P}|| = 1^1 \cdot 2^2 = 4$ preference combinations. Alternatives A1 to A8 represent eight different drill settings, and their objective measurements were determined by test data.

Table 8.36: Random set of threshold judgements for selecting a drill setting

Criteria	λ_1	λ_2	λ_3
c_1 Residual tensile strength	{[45.54,45.35]	[50.16,50.16]	[56.04,56.04] }
c_2 Top delamination factor	{[1.25,1.29]	[1.21,1.25]	[1.08,1.29]}
c_3 Bottom delamination factor	{[1.08,1.14]	[1.08,1.2]	[1.17,1.17]}

A random set of threshold judgements was generated for two thresholds and is shown in Table 8.36. The combinatorial space had the size of $K = (1^1 \cdot 2^2) \cdot (3^3) = 108$. These threshold judgements were applied for an initial CMAA analysis with the full combinatorial space. The resulting rank acceptabilities, Olympic holistic acceptability, their respective ranks R_1, R_{acc} are shown in Table 8.37. CMAA produced the same most-preferred alternative as the original study, which did not provide the other ranking positions. The entropy-optimal consensus path required just two threshold judgement resolutions to reach a hard and revelation-invariant consensus.

A consensus path simulation was conducted to analyse this decision's mean consensus path length in a more comprehensive manner. The threshold judgements with two thresholds were independently generated, creating a threshold judgement space of $||\mathbf{A}|| = 3^3 = 27$, and

Table 8.37: Initial rank acceptability index for the random threshold judgements

	b_i^1	b_i^2	b_i^3	b_i^4	b_i^5	b_i^6	b_i^7	b_i^8	acc_i	R_1	R_{acc}
a_1	0.15	0.15	0.48	0.23	0.00	0.00	0.00	0.00	0.18	4	4
a_2	0.46	0.43	0.11	0.00	0.00	0.00	0.00	0.00	0.41	2	2
a_3	0.00	0.00	0.00	0.04	0.04	0.08	0.00	0.85	0.00	8	8
a_4	0.07	0.11	0.15	0.11	0.22	0.34	0.00	0.00	0.09	5	6
a_5	0.45	0.11	0.44	0.00	0.00	0.00	0.00	0.00	0.35	3	3
a_6	0.73	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.52	1	1
a_7	0.04	0.07	0.11	0.08	0.14	0.23	0.33	0.00	0.06	7	7
a_8	0.07	0.11	0.26	0.23	0.33	0.00	0.00	0.00	0.10	5	5

Table 8.38: Mean consensus path lengths for drilling setting selection

	P(EO)										
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
3+2	2.7	2.6	2.4	2.3	2.2	2.1	1.9	1.8	1.7	1.6	1.4

resulting in a complete combinatorial space of 108. This decision was simulated with 1,000 different decisions, applying random threshold judgements. For each decision, one consensus path was computed, varying the probability that each discrepancy would be resolved as entropy-optimal or not, expressed with $P(EO)$. For the CMAA analysis again the full combinatorial space was used with K instances. The termination parameter for consensus was $\tau = 0.0$, representing a hard consensus.

The mean consensus path lengths for the simulation are shown in Table 8.38. The maximum number of threshold judgement discrepancies is three, and the decision originally contained two preference discrepancies, totaling five discrepancies. In the best-case scenario, with $P(EO) = 1.0$, the average consensus path length is 1.4, indicating a rapid consensus-building. Even in the worst-case scenario, where the entropy-optimal resolution is never chosen, the path length averages 2.7. In a typical subjective judgement situation, decision-makers would evaluate all 24 judgement tasks, potentially resulting in 24 judgement discrepancies. Adding the two preference discrepancies brings the total to 26 discrepancy evaluations. When comparing the mean consensus path lengths to these 26 discrepancies, it is clear that the consensus path lengths are highly efficient. This highlights a significant advantage of using threshold judgements.

8.4 Conclusion

In this chapter, CMAA was applied to data from eight original case studies that used various single-step MCGDM approaches for determining a recommendation. The experiments highlight different abilities of CMAA and describe two adaptations of the method to accommodate decision models with non-standard input formats. The first adaptation showed how CMAA can be combined with the widely-used decision method AHP and tested it for two different case studies. The second modification showed how subjective thresholds for objective measurements can be treated with CMAA.

These experiments have shown:

1. *Efficiency:* CMAA provides a very efficient way to reach consensus in groups. The mean consensus path lengths are often just a fraction of the initial number of discrepancies contained in the decision. Retroactive application of CMAA to two cases revealed that an output consensus had already been reached without the need for a clarification conference. The standard MCGDM approaches were not able to recognise this. This demonstrates that CMAA builds consensus efficiently (Goals 1 and 2 on page 7), and that its combinatorial analysis provides a comprehensive and reliable measure of how close a group is to a revelation-invariant consensus (Goal 3 on page 7).
2. *Compatibility:* CMAA can be combined with most multi-criteria decision methods without modification. AHP is an exception, because of the need to maintain consistency in the evaluations. This difficulty was solved by using a minimal-input version of AHP, which also reduces the number of evaluations required from decision-makers. With this approach, CMAA only needs to adapt its input format to the hierarchical pairwise format of AHP. The consensus-building efficiency of CMAA with minimal-input AHP is comparable to that achieved with other multi-criteria methods, demonstrating CMAA's adaptability to different decision methods and addressing Research Question 4 on page 4.
3. *Minority dissent:* The study for awarding agricultural loans highlighted the importance of minority dissent. In this case, one of the five decision-makers had an opposing opinion that could have significantly altered the group's decision if adopted by the others. Without CMAA's ability to recognise the influence of every input on consensus, the potential effect of the minority opinion could not have been detected by any other approach to consensus-building. This achieves Goal 4 on page 7 by measuring each resolution's impact on consensus and fulfilling Requirement 1 on page 8, preserving the evaluation inputs of all decision-makers.
4. *Subjective value of measurements:* Many studies provide objective performance measurements for alternatives, which are then converted into subjective value judgements. A threshold approach to such decision models can prevent inconsistency, reduce the number of judgements needed, and minimise bias and the risk of decision manipulation. CMAA can generate very short consensus paths for this type of decision, achieving Goal 2 on page 7. This ability would be particularly beneficial for large decisions.

9
Case Studies

This chapter puts the CMAA framework to the test in practical applications. The objective is to determine whether groups will accept the new consensus-building approach, addressing research Question 5 on page 11. Two case studies were carried out; the first was described by Goers and Horton (2023b).

9.1 Project selection in a biotechnology startup

9.1.1 The decision context

Three post-doctoral researchers were planning to found a biotechnology startup. Their business idea was to offer two new technologies – a bioreactor and a purification device – for cultivating viruses for use in vaccine manufacturing and gene therapy. Both technologies held the promise of improving virus manufacturing significantly.

9.1.2 The decision

The decision task was to select the laboratory experiment that would deliver the greatest value for their startup at that moment in time. The six experiments to choose from were:

- a_1 Test whether the COVID-19 virus could be cultivated in the current bioreactor using cell-type A.
- a_2 Test whether the COVID-19 virus could 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 could 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 bioreactor was originally developed for cultivating adeno-associated viruses, which are widely used in gene therapy. However, the COVID-19 pandemic broke out during the project. So, for the startup it was an unplanned potential new business opportunity to improve manufacturers' ability to produce large amounts of COVID-19 vaccine.

The six criteria for selecting an experiment were:

- c_1 *Product attractiveness:* If successful, the resulting product will attract more potential customers.
- c_2 *Patentability:* If successful, the idea underlying the resulting product will be patentable.
- c_3 *Experiment success rate:* The experiment has a high probability of success.
- c_4 *Fast results:* The experiment will yield a result in four weeks or less.
- c_5 *MVP-relevance:* The experiment is relevant for developing a Minimum Viable Product (Moogk, 2012).
- c_6 *Attractiveness to investors:* A positive result would increase the chances of raising venture capital.

9.1.3 Inputs and initial analysis

Table 9.1: Individual (left) and unified (right) preference vectors from the founders

	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 9.2: Individual judgements from the three founders

	a_1 $\{\lambda_1,\lambda_2,\lambda_3\}$	a_2 $\{\lambda_1,\lambda_2,\lambda_3\}$	a_3 $\{\lambda_1,\lambda_2,\lambda_3\}$	a_4 $\{\lambda_1,\lambda_2,\lambda_3\}$	a_5 $\{\lambda_1,\lambda_2,\lambda_3\}$	a_6 $\{\lambda_1,\lambda_2,\lambda_3\}$
c_1	{A,A,A}	{A,A,A}	{A,B,B}	{B,X,B}	{B,X,X}	{A,A,B}
c_2	{B,A,A}	{B,A,A}	{A,A,A}	{A,A,A}	{B,A,B}	{A,B,B}
c_3	{B,B,B}	{B,B,B}	{A,B,B}	{B,A,B}	{B,A,B}	{A,A,A}
c_4	{B,B,X}	{B,B,B}	{B,B,A}	{A,A,B}	{A,A,B}	{B,X,B}
c_5	{A,A,A}	{A,A,A}	{A,A,A}	{A,A,A}	{A,A,B}	{A,A,A}
c_6	{A,A,A}	{A,A,A}	{A,B,B}	{B,B,B}	{X,B,X}	{B,B,B}

The decision method ABX-Lex (see Section 2.2.2) was used, with the three judgement equivalence classes shown in Table 2.4. The individual and unified preferences by the decision-makers are shown in Table 9.1. All six preference tasks resulted in discrepancies. Several inputs were inconsistent. For example, decision-maker DM_3 assigned priority 3 to criterion c_2, but also named it as a possibility for criteria c_3, c_5 and c_6. After removing all such inconsistencies, the remaining number of feasible preference vectors for each founder was quite small: for DM_1 four vectors, and for DM_2 and DM_3 two vectors each. The total number of feasible preference vectors was $||\mathbf{P}|| = 200$.

Table 9.2 shows the 36 decision-makers' judgements; 17 of these were unanimous, and 19 were bivalent discrepancies {A, B} or {B, X}, resulting in $||\mathbf{A}|| = 2^{10} = 524,288$ judgement combinations. The search space of combinations thus contained $K = 104,857,600$ instances.

Table 9.3: Rank acceptability indices and Olympic holistic acceptability

	b_i^1	b_i^2	b_i^3	b_i^4	b_i^5	b_i^6	acc_i
a_1	0.12	0.22	0.20	0.16	0.16	0.14	0.16
a_2	0.21	0.24	0.23	0.19	0.10	0.03	0.22
a_3	0.40	0.12	0.20	0.17	0.08	0.02	0.30
a_4	0.11	0.14	0.12	0.17	0.30	0.16	0.12
a_5	0.05	0.11	0.09	0.09	0.17	0.49	0.07
a_6	0.19	0.16	0.15	0.21	0.15	0.14	0.18
$\sum$	1.10	0.99	0.99	0.99	0.96	0.97	1.06

CMAA analysis with Monte Carlo simulation using $K_{MC} = 10,000$ instances resulted in the initial rank acceptability indices shown in Table 9.3. The rank-1 and holistic acceptabilities both yield the same ranking $a_3 \succ a_2 \succ a_6 \succ a_1 \succ a_4 \succ a_5$. The rank-1 acceptabilities return an entropy of 2.33, which is close to the maximum possible entropy of 2.58. Alternative a_3 is almost twice as strong as the second-ranked alternative a_2. Single-step multi-criteria group decision-making approaches would stop here and recommend a_3, because it receives the strongest support. However, as Section 9.1.4 will show, it turns out that this suggestion did not, in fact, reflect the most-preferred choice of the founders.

Tables 9.4 and 9.5 show the potential judgement acceptabilities for two of the alternatives. The results for alternative a_3 are shown in Table 9.4. The discrepancies at $\lambda(1,3)$, $\lambda(3,3)$ and $\lambda(4,3)$ reveal that a_3's rank-1 acceptability would be improved by a resolution to A and worsened by a resolution to B.

The potential judgement acceptabilities for alternative a_4 are shown in Table 9.5. The judgement discrepancies that influence a_4's rank-1 acceptability the most are at $\lambda(3,3)$, $\lambda(4,3)$, $\lambda(1,4)$, $\lambda(3,4)$ and $\lambda(4,4)$. Alternative a_4 benefits most from the resolution (to B) that is least favourable to a_3. This ability to identify particularly important discrepancies

Table 9.4: Potential 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	-	-	-	-	-	0.52	0.22	-	0.36	0.37	-	0.37	0.37	0.34	0.39	-
c_2	0.35	0.38	-	0.35	0.39	-	-	-	-	-	0.37	0.37	-	0.36	0.38	-
c_3	-	-	-	-	-	0.52	0.22	0.36	0.38	-	0.36	0.37	-	-	-	-
c_4	-	0.35	0.39	-	-	0.49	0.25	0.35	0.39	-	0.35	0.38	-	-	0.36	0.37
c_5	-	-	-	-	-	-	-	-	-	-	0.37	0.37	-	-	-	-
c_6	-	-	-	-	-	0.45	0.29	-	-	-	-	0.37	0.37	-	-	-

Table 9.5: Potential judgement acceptabilities for alternative a_4.

	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	-	-	-	-	-	0.09	0.11	-	0.13	0.08	-	0.10	0.11	0.09	0.11	-
c_2	0.10	0.10	-	0.10	0.11	-	-	-	-	-	0.10	0.10	-	0.09	0.11	-
c_3	-	-	-	-	-	0.07	0.13	0.15	0.06	-	0.10	0.11	-	-	-	-
c_4	-	0.10	0.11	-	-	0.05	0.15	0.17	0.03	-	0.09	0.11	-	-	0.10	0.10
c_5	-	-	-	-	-	-	-	-	-	-	0.10	0.11	-	-	-	-
c_6	-	-	-	-	-	0.09	0.11	-	-	-	-	0.10	0.10	-	-	-

can be very useful to the facilitator of the meeting and is one of the particular advantages of CMAA.

Table 9.6: Potential preference entropies for the biotechnology case study

	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

Table 9.7: Potential judgement entropies for the biotechnology case study

	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	-	-	-	-	-	-	2.07	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 potential preference entropies and potential judgement entropies are shown in Tables 9.6 and 9.7, respectively. The lowest entropy from either table is 1.56, and would be reached if the group decided to resolve c_3 to first position in the preference vector. Therefore, the greatest potential for consensus improvement lies in agreeing on the top-priority criterion. This is to be expected with a non-compensatory decision model.

The CMAA analysis concludes that there is a slight preference towards the alternative a_3, but the current entropy of 2.33 is too large to declare even a soft consensus. Thus, a consensus-building process needed to be performed.

9.1.4 Consensus-building

Consensus-building took place as a face-to-face meeting between the three founders and a facilitator, who used the consensus-building Algorithm 2 as an interactive digital assistant. Each consensus step was visualised for the founders, so that they were able to follow the changes that their resolutions caused in the acceptabilities of each alternative.

At each consensus step, the facilitator asked each founder to provide their reasons for their respective evaluations and clarify any open questions. Then, they were asked to provide a resolution that all three could agree on.

Table 9.8: Path to hard consensus of the biotechnology example

Step	Act. res.	Rank-1 acceptabilities							Pot. res.	
		h	b_1^1	b_2^1	b_3^1	b_4^1	b_5^1	b_6^1	λ/μ	$\hat{h}$
Initial	-	2.33	0.12	0.21	0.40	0.11	0.05	0.19	$\mu(3) = \{1\}$	1.56
1	$\mu(4) = \{1\}$	1.86	0.02	0.05	0.51	0.27	0.17	0.01	$\lambda(4,3) = \{A\}$	0.58
2	{B}	2.04	0.02	0.05	0.10	0.47	0.31	0.03	$\lambda(4,4) = \{A\}$	0.49
3	{A}	0.49	0.00	0.00	0.00	0.91	0.11	0.00	$\lambda(4,5) = \{B\}$	0.00
4	{A}	0.74	0.00	0.00	0.00	0.82	0.22	0.00	$\lambda(1,4) = \{B\}$	0.48
5	{B}	0.48	0.00	0.00	0.00	0.92	0.11	0.00	$\lambda(3,4) = \{A\}$	0.20
6	{A}	0.20	0.00	0.00	0.00	1.00	0.03	0.09	$\lambda(1,5) = \{X\}$	0.00
7	{B}	0.32	0.00	0.00	0.00	1.00	0.06	0.00	$\lambda(2,5) = \{B\}$	0.00
8	{B}	0.00	0.00	0.00	0.00	1.00	0.00	0.00	-	

Table 9.8 illustrates the resolutions chosen, the current entropy, the rank-1 acceptabilities and the entropy-optimal resolution for each consensus step. In total, the group took eight steps to reach a hard consensus with a_4 as the most-preferred alternative. Significantly, a_4 was not the most-preferred alternative based on the initial inputs.

At consensus step 1, the founders decided to make c_4 (Fast results) the most important criterion. This improved the rank-1 acceptability of the strongest alternative a_3 still further to 0.51, and negatively impacted the two COVID-19 tests a_1 and a_2, because these did not have category A performance for this criterion. The second consensus step reduced a_3 to only one-fifth of its previous rank-1 acceptability. The rank-1 acceptabilities for a_1 and a_2 remained the same, and the alternatives a_4 and a_5 became the strongest.

At step 3, alternatives a_1, a_2, a_3 and a_6 were eliminated from the decision. The remaining consensus steps were needed simply to determine whether a_4 or a_5 should be most-preferred. A firm soft consensus ($h < 0.3 \cdot h_{max}$) was reached at step 3. From step 6 on, a_4 already ranked first for the whole combinatorial space, and a few instances remained in which alternative a_5 was joint most-preferred with a_4. Four of the eight discrepancies were resolved towards the entropy-optimal resolutions, and four were not. The consensus-building process took about 60 minutes.

Figure 9.1 shows the development of the various types of judgement discrepancy (see Section 3.3). The height of the column at each step shows the number of judgement discrepancies that remain in the decision. Black shading refers to active discrepancies, medium grey to pivots and light grey to inactive discrepancies. Columns marked with a star show where the resolution chosen by the decision-makers was entropy-optimal. The line graph shows the entropy using the secondary vertical axis.

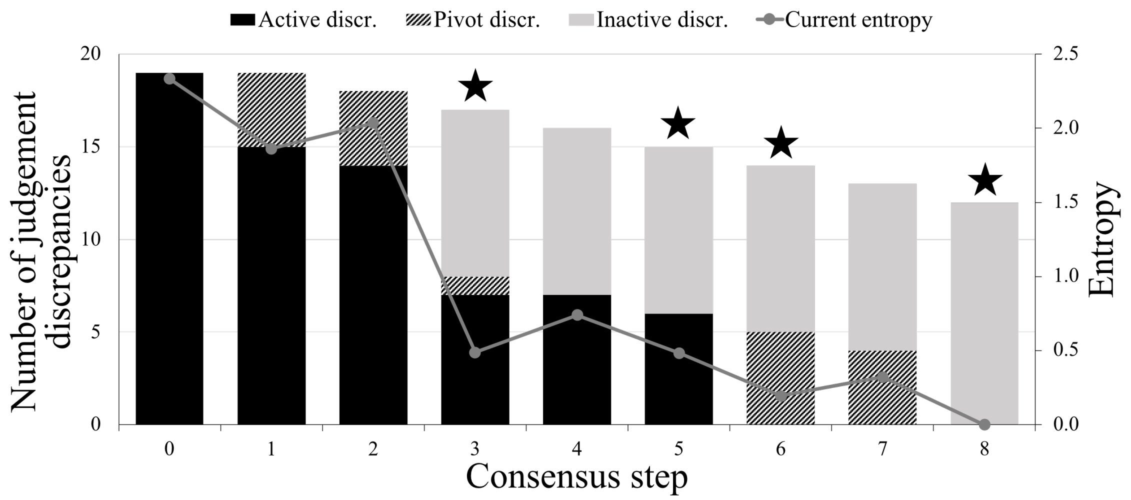


Figure 9.1: Development of the types of discrepancies during consensus-building

Initially, all 19 judgement discrepancies were active. In step 1, a preference discrepancy was resolved, so the 19 judgement discrepancies remained. The resolution of the preference caused four discrepancies to become pivots; these contained resolutions that would have immediately eliminated one or more alternatives. This is valuable information for the facilitator of the meeting. With the resolution of $\lambda(4,4)$ to A at step 2, four alternatives were eliminated, which led to the increase in inactive discrepancies from zero to nine at step 3 and the sharp improvement in entropy. Steps 4 and 5 did not contain any pivots, so no resolution would have lead to the elimination of further alternatives. The entropy-optimal resolution at step six turned all remaining active discrepancies into pivots. And, of course, at step 8, all discrepancies become inactive, because alternative a_4 reached a rank-1 acceptability of 1, and the entropy was 0.

Table 9.9: Unified 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	**B**	**B**	AB	$\{2,3,4,5\}$
c_2	AB	AB	A	A	**B**	AB	$\{2,3,4,5,6\}$
c_3	B	B	AB	**A**	AB	A	$\{2,3,4,5,6\}$
c_4	BX	B	**B**	**A**	**A**	BX	$\{\mathbf{1}\}$
c_5	A	A	A	A	AB	A	$\{2,3,4\}$
c_6	A	A	AB	B	BX	B	$\{2,3,4,5,6\}$

Table 9.9 shows the unified judgement matrix **A** and the unified preference vector **P** after consensus had been reached at step 8. The resolutions chosen by the decision-makers are marked with a bold typeface. Twelve judgement discrepancies and five preference discrepancies remain, yielding $||\mathbf{A}|| \cdot ||\mathbf{P}|| = 2^{12} \cdot 54 = 221,184$ instances. Alternative a_4 ranks first for all remaining instances, so this consensus is revelation-invariant.

In order to determine if this consensus-building result was typical, a comparison with similar decisions was made. First, 10,000 random consensus paths were generated for the

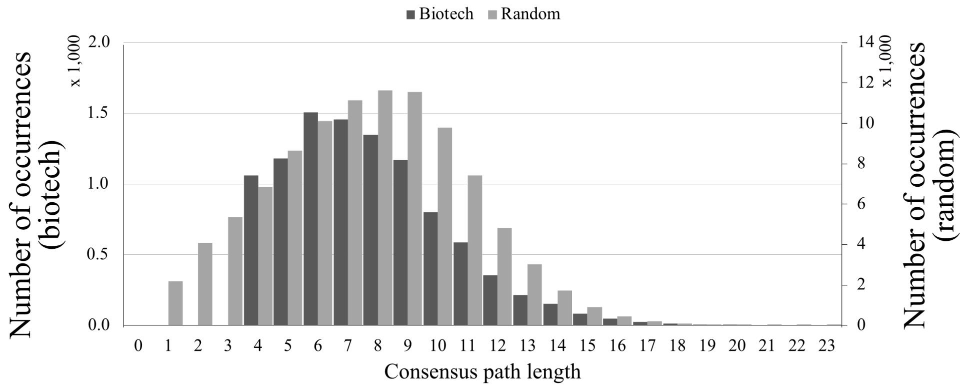


Figure 9.2: Distribution of path lengths for the biotechnology example and random decisions

original biotechnology decision. Then, 1,000 random decisions of the same size and with the same decision method and a similar number of discrepancies were generated, and each of these was solved with 100 random consensus paths using Monte Carlo simulation.

Figure 9.2 shows the distribution of consensus path lengths for the original decision (the darker shade, left vertical axis) and for the set of synthetic random decisions (lighter shade, right vertical axis). The original data resulted in a mean path length of 7.7, and the random decisions of 7.6. Note that several of the random decisions could have achieved consensus with only one or two resolution steps. It can be concluded that the real-life consensus path was typical for a decision of this size and type and with this number of discrepancies.

9.1.5 Observations and feedback from the decision-makers

In the decision meeting, the founders behaved exactly as CMAA needs them to: they shared their perspectives, interpretations, information and experience for each topic and created a shared understanding each time. This was particularly valuable, because the first discrepancy to be resolved concerned the dictator criterion, which by nature had an extremely strong influence on the decision result. These clarifications were not only valuable in the short term for reaching a decision, but also for creating a shared mindset as their project went forward.

The research Question 5 on page 11 is concerned with the acceptance by decision-makers of the CMAA consensus-building process. In order to obtain insight into this topic, the founders were asked the following five questions at the end of the meeting:

Q1 *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.*

Q2 *1) Did you understand the dictator property of non-compensatory criteria and 2) did it concern you?*
A: *1) Yes; 2) No*

Q3 *Does the SMM 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.*

Q4 *Was the consensus-building transparent and understandable?*
A: *Yes.*

Q5 *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.*

These answers suggest that decision-makers considered the CMAA approach to be appropriate and acceptable for their decision task. Recommendation 6 on page 96 did not need to be implemented.

9.2 Selection of a new product feature

9.2.1 The decision context and decision

Five junior software developers at a software company were tasked with adding a new feature to an in-house web application. This platform-independent tool supported software development teams in evaluating the effort and risks associated with development tasks. The project was undertaken as part of a Master's thesis by a Computer Science student at the University of Magdeburg (Blumenroth, 2024).

Together with the author, the five developers generated 18 ideas for new features, seven of which passed a suitability screening:

a_1 Chat-bot: Provide a chat-bot version.

a_2 Themes: Individualise the look and feel of the application.

a_3 Learning: Integrate dynamic learning.

a_4 Data feed: Allow for an information feed from an external software tool to their web application.

a_5 Import: Import items from a CSV-file.

a_6 Export: Export results to a CSV-file.

a_7 E-Mail: Forward the results via E-Mail.

Taking into account their limited time, experience and resources, the team agreed on five criteria for their decision:

c_1 *Value creation*: The more value the new feature creates for users, the better.

c_2 *Skill set*: The more they already possess the skills to implement the feature, the better.

c_3 *Effort*: The faster the feature can be implemented, the better.

c_4 *Testing*: The easier the new feature can be tested, the better.

c_5 *Maintenance*: The easier the feature can be maintained, the better.

Since the criteria *Effort* and *Value creation* might contradict each other, a compensatory decision model was used, which would allow a deficiency in value creation to be compensated by easier implementation. The FSAW decision model (see Section 2.2.2) was chosen, so that the decision-makers could use linguistic variables for their evaluations. The linguistic variables used are shown in Table 2.1.

Table 9.10: Initial unified preference vector for the feature case study

Criteria	$\{\mu_1, \mu_2, \mu_3\}$
c_1 (Value creation)	$\{VH, H\}$
c_2 (Skill set)	$\{H, M\}$
c_3 (Effort)	$\{M, L, VH\}$
c_4 (Testing)	$\{M, VL, H\}$
c_5 (Maintenance)	$\{H, M\}$

Table 9.11: Initial unified judgement matrix for the feature case study

	a_1 $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$	a_2 $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$	a_3 $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$	a_4 $\{\lambda_1, \lambda_2, \lambda_3\}$	a_5 $\{\lambda_1, \lambda_2, \lambda_3\}$	a_6 $\{\lambda_1, \lambda_2, \lambda_3\}$	a_7 $\{\lambda_1, \lambda_2, \lambda_3\}$
c_1	$\{G, F, P, VG\}$	$\{F, P, VG\}$	$\{G, F, VG\}$	$\{F, VG\}$	$\{G, VG\}$	$\{G\}$	$\{G\}$
c_2	$\{F, P, VP\}$	$\{G, F, VG\}$	$\{G, F, VG\}$	$\{F, P, VP\}$	$\{G, F\}$	$\{G, F, VG\}$	$\{G, F\}$
c_3	$\{P\}$	$\{G, F\}$	$\{G, F, P, VG\}$	$\{F, P, VP\}$	$\{F, P\}$	$\{G, F\}$	$\{G, F, P\}$
c_4	$\{G, F, P\}$	$\{G, F\}$	$\{G, VG\}$	$\{G, F, VG\}$	$\{G, F, VG\}$	$\{G, F\}$	$\{G, F\}$
c_5	$\{G, F, P\}$	$\{G, F, P, VG\}$	$\{G, VG\}$	$\{F, P, VG\}$	$\{G, F\}$	$\{F\}$	$\{G, F\}$

9.2.2 Inputs and initial analysis

The decision involved five decision-makers $d = 5$, five criteria $n = 5$, and seven alternatives $m = 7$. The developers completed 35 judgement tasks and five preference tasks. They input their evaluations separately via an online form. The resulting unified preference vector and unified judgement matrix are shown in Tables 9.10 and 9.11, respectively. All preferences were discrepancies with valency 2 or 3, leading to $||\mathbf{P}|| = 2^3 \cdot 3^2 = 72$ combinations. Only four judgement tasks were unanimous, while the remaining tasks resulted in $||\mathbf{A}|| = 1^4 \cdot 2^{14} \cdot 3^{14} \cdot 4^3 = 5{,}015{,}306{,}502{,}144$ combinations. The combinatorial space was very large, with $K = 72 \cdot 5{,}015{,}306{,}502{,}144 > 3.61 \cdot 10^{14}$ instances. Therefore, Monte Carlo simulation was used for the CMAA analysis with $K_{MC} = 100{,}000$. Although this is ten times larger that the recommended value of $K_{MC} = 10{,}000$ (Recommendation 1 on page 80), the computation time for the CMAA analysis is only 0.55 seconds.

Table 9.12: Rank acceptability indices and Olympic holistic acceptability

	b_i^1	b_i^2	b_i^3	b_i^4	b_i^5	b_i^6	b_i^7	acc_i	R_1	R_{acc}
a_1	8E-4	0.01	0.01	0.03	0.07	0.26	0.63	3E-3	7	7
a_2	0.16	0.17	0.12	0.12	0.20	0.17	0.06	0.159	2	3
a_3	0.63	0.18	0.09	0.06	0.03	0.01	1E-3	0.440	1	1
a_4	0.02	0.05	0.06	0.08	0.14	0.37	0.27	0.036	6	6
a_5	0.11	0.25	0.24	0.21	0.14	0.05	4E-3	0.165	3	2
a_6	0.07	0.23	0.26	0.24	0.15	0.04	3E-3	0.138	5	4
a_7	0.04	0.15	0.23	0.26	0.22	0.08	8E-3	0.092	4	5
$\sum$	1.04	1.04	1.02	0.99	0.96	0.98	0.98	1.03		

The initial entropy for this problem was 1.78, out of a maximum possible entropy of 2.81. In Table 9.12, the initial rank acceptabilities b_i^1 and the Olympic holistic acceptabilities acc_i

are shown. Both indicate a strong preference for alternative a_3 (*integrate dynamic learning*). The sums of the first three rank acceptabilities are greater than 1.0, indicating that these ranks involved ties, which are also carried over to the holistic acceptability.

The rankings of feature ideas differ slightly between the rank-1 acceptabilities (column R_1) and the holistic acceptability (column R_{acc}). Inclusion of the rank-2 and rank-3 acceptabilities switches the ranking positions of alternatives a_2 and a_5 and of a_6 and a_7. This occurs when a weaker rank-1 acceptability is compensated by a stronger performance in ranks 2 and 3.

Table 9.13: Potential preference entropies for the feature case study

	μ_1	μ_2	μ_3
c_1 (Value creation)	1.83	1.73	-
c_2 (Skill set)	1.77	1.78	-
c_3 (Effort)	1.75	1.77	1.77
c_4 (Testing)	1.68	1.89	1.62
c_5 (Maintenance)	1.72	1.83	-

Table 9.14: Potential judgement entropies for the feature case study

	a_1				a_2				a_3			
	λ_1	λ_2	λ_3	λ_4	λ_1	λ_2	λ_3	λ_4	λ_1	λ_2	λ_3	λ_4
c_1	1.77	1.77	1.77	1.80	1.65	1.53	1.65	-	1.70	2.32	0.90	-
c_2	1.79	1.78	1.77	-	1.75	1.70	1.80	-	1.75	2.20	1.16	-
c_3	-	-	-	-	1.79	1.75	-	-	1.49	1.94	2.26	1.05
c_4	1.78	1.78	1.77	-	1.79	1.76	-	-	1.91	1.63	-	-
c_5	1.79	1.77	1.77	-	1.78	1.72	1.62	1.77	2.01	1.50	-	-

	a_4			a_5			a_6			a_7		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
c_1	1.68	1.83	-	1.66	1.81	-	-	-	-	-	-	-
c_2	1.84	1.76	1.70	1.81	1.71	-	1.72	1.61	1.84	1.82	1.70	-
c_3	1.83	1.77	1.72	1.81	1.73	-	1.82	1.71	-	1.85	1.73	1.67
c_4	1.78	1.73	1.81	1.76	1.71	1.81	1.81	1.74	-	1.81	1.74	-
c_5	1.70	1.67	1.87	1.81	1.70	-	-	-	-	1.82	1.71	-

The potential preference and judgement entropies are shown in Tables 9.13 and 9.14, respectively. All preference resolutions have a very small impact on the entropy; the lowest potential preference entropy is at $\mu_3(4)$ with 1.62. In contrast, the potential judgement entropies show gains of up to 30% and losses as high as -50% – both at the discrepancy $\lambda_3(1, 3)$.

Table 9.15: Judgement sensitivities for the judgement task (c_1, a_3)

	a_1	a_2	a_3	a_4	a_5	a_6	a_7
δ_+	0.001	0.062	0.217	0.015	0.076	0.057	0.033
δ_-	-0.001	-0.077	-0.279	-0.016	-0.076	-0.054	-0.030
σ_i	-0.001	-0.077	-0.279	-0.016	-0.076	0.057	0.033

The judgement sensitivities for this discrepancy are shown in Table 9.15. Alternative a_3 might lose almost half of its rank-1 acceptability. For all other alternatives, no judgement resolution would significantly impact their rank-1 acceptability.

9.2.3 Consensus-building

The consensus-building process was conducted in a face-to-face meeting. Each junior developer had a protocol of their input evaluations. The unified preferences and judgements were visualised on a digital whiteboard, together with the rank-1 acceptabilities for each alternative, which were updated at each step of the process. A facilitator then directed the clarification conferences, allowing the group to share their mental models and reach a resolution agreement. The resolution was applied, the CMAA analysis computed the statistics, and an excerpt was shared with the group via a digital whiteboard. This process continued until the group reached a hard consensus.

Table 9.16: Path to hard consensus for the feature selection decision

Step	Act. res.	Rank-1 acceptabilities								Pot. res.	
		h	b_1^1	b_2^1	b_3^1	b_4^1	b_5^1	b_6^1	b_7^1	λ/μ	$\hat{h}$
Initial	-	1.75	8E-4	0.15	0.65	0.02	0.10	0.07	0.04	$\lambda(1,3) = \{VG\}$	0.89
1	{G}	1.67	4E-4	0.15	0.67	0.02	0.10	0.06	0.04	$\lambda(3,3) = \{VG\}$	0.65
2	{G}	1.19	1E-4	0.12	0.80	0.01	0.06	0.03	0.02	$\lambda(2,3) = \{VG\}$	0.38
3	{G}	0.97	0.00	0.11	0.84	0.01	0.04	0.01	0.01	$\lambda(5,3) = \{VG\}$	0.52
4	{G}	1.30	0.00	0.15	0.77	0.02	0.07	0.03	0.01	$\lambda(1,2) = \{P\}$	0.84
5	{G}	1.00	0.00	0.03	0.87	0.02	0.07	0.03	0.02	$\lambda(4,3) = \{VG\}$	0.56
6	{G}	1.34	0.00	0.05	0.82	0.03	0.11	0.05	0.03	$\mu(3) = \{VH\}$	0.99
7	{M}	1.31	0.00	0.04	0.82	0.03	0.11	0.05	0.03	$\lambda(4,5) = \{F\}$	1.03
8	{G}	1.20	0.00	0.05	0.86	0.03	0.06	0.05	0.04	$\lambda(3,5) = \{P\}$	0.92
9	{F}	1.34	0.00	0.04	0.80	0.03	0.12	0.05	0.03	$\lambda(1,5) = \{G\}$	0.91
10	{VG}	1.53	0.00	0.04	0.68	0.03	0.24	0.04	0.03	$\lambda(5,5) = \{F\}$	0.93
11	{F}	0.92	0.00	0.05	0.91	0.03	0.00	0.05	0.04	$\lambda(2,6) = \{F\}$	0.66
12	{G}	0.67	0.00	0.05	0.94	0.03	0.00	0.00	0.04	$\lambda(5,2) = \{P\}$	0.41
13	{G}	0.56	0.00	0.02	0.97	0.03	0.00	0.00	0.04	$\lambda(5,7) = \{F\}$	0.33
14	{F}	0.32	0.00	0.02	0.97	0.03	0.00	0.00	0.00	$\lambda(1,4) = \{F\}$	0.14
15	{VG}	0.45	0.00	0.02	0.95	0.06	0.00	0.00	0.00	$\lambda(5,4) = \{P\}$	0.15
16	{F}	0.16	0.00	0.02	0.99	2E-3	0.00	0.00	0.00	$\lambda(2,2) = \{F\}$	0.02
17	{G}	0.02	0.00	0.00	0.99	2E-3	0.00	0.00	0.00	$\lambda(2,4) = \{VP\}$	0.00
18	{VP}	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	-	-

The resulting consensus path is shown in Table 9.16. The group required 18 consensus steps to reach a hard consensus, which took about 75 minutes to conduct (4 minutes on average for each clarification conference). The initially most-preferred feature idea, *Learning* (a_3), was ultimately selected.

At termination, 18 of the 36 preference and judgement discrepancies remained. These were inactive, so the consensus was not only hard, but also revelation-invariant (Definition 3 on page 4). The number of preference instances at consensus was $||\mathbf{P}|| = 1^1 \cdot 2^3 \cdot 3^1 = 24$, and the number of judgement instances was $||\mathbf{A}|| = 1^{21} \cdot 2^6 \cdot 3^6 \cdot 4^2 = 746,496$, giving a total of $K = 17,915,904$ instances. A very firm soft consensus was reached at Step 12 with $h = 0.67$ ($\tau = 0.3$), which is 24% of the maximum entropy.

For 15 clarification conferences, the sharing of mental models led to resolutions in the middle of the evaluation spectrum ('moderate', 'fair' or 'good') rather than one of the extreme evaluations ('very poor' or 'very good'). This explains the very long consensus path, as a

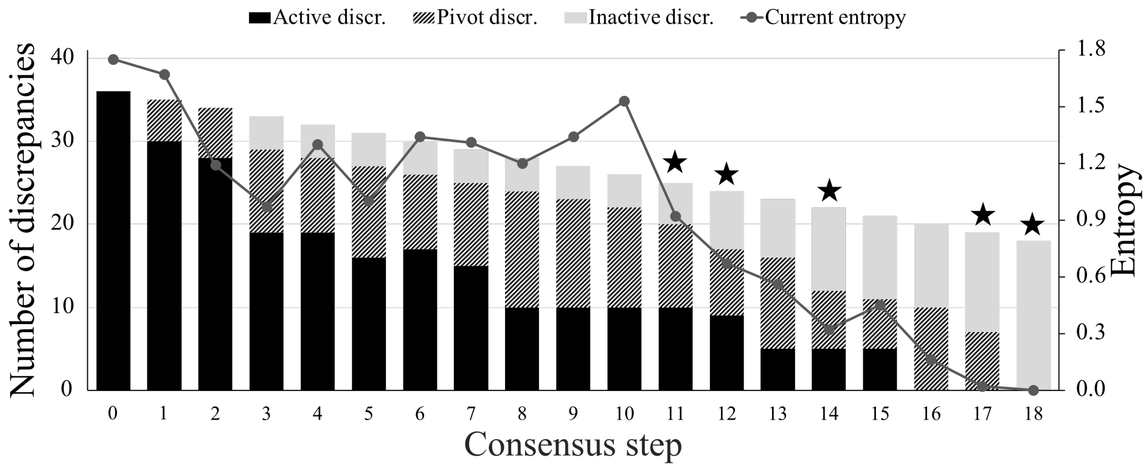


Figure 9.3: Development of the types of discrepancy during consensus-building

compensatory decision model tends to take longer when 'middle' evaluations are applied, because the performance of the alternatives gets less distinguishable (see Section 7.4.4).

The development of discrepancy types is illustrated in Figure 9.3. The horizontal axis represents consensus steps. The primary vertical axis indicates the number of remaining discrepancies. The shaded parts of each column represent the number of each type of discrepancy at each step. The solid black part of the column represents active discrepancies, while the striped parts denote pivots. Light gray parts of the column represent inactive discrepancies. The second vertical axis measures the entropy, with the dark gray line indicating the current entropy at each step. Stars over columns indicate when the group resolved a discrepancy optimally in terms of entropy at that step.

Four observations can be made from this diagram. First, the number of inactive discrepancies grows, which is expected because for each eliminated feature idea, all corresponding judgement discrepancies immediately become inactive. Second, there are pivots present at each step of the process, so it would have been possible to eliminate at least one alternative at every step. Third, at steps 16 and 17, all remaining active discrepancies contained a pivot, indicating that every available resolution would lead to the elimination of a feature idea. And, most importantly, the entropy would be significantly reduced, even if the group did not choose an entropy-optimal resolution. Even though the decision-makers chose disadvantageous resolutions at many steps, consensus was reached with half the discrepancies untouched. This is further support for the Slack-Hypothesis 1 on page 10.

9.2.4 Observations and feedback from the decision-makers

After the consensus-building, the junior developers discussed some insights. They particularly valued the clarification conferences regarding *Data feed* (a_4). It was one of the most requested features by their users, and some developers were particularly attached to this idea. However, through the clarification conferences, they discovered unshared mental models that changed the view of the other junior developers regarding the criteria *Value creation*, *Effort*, and *Skill set*. This shows that the facilitator should always address any concerns the decision-makers have regarding alternatives they are particularly attached to.

In retrospect, resolving the judgement discrepancies $\lambda(1, 4)$ and $\lambda(1, 5)$ was unnecessary.

Both could have been left unresolved, and a hard consensus would still have been reached. However, the consensus-building process required these resolutions because, at the time, they had the greatest potential to positively impact consensus. The resolutions were not entropy-optimal, which the algorithm could not have predicted.

After the consensus-building process, the decision-makers filled in a questionnaire. The first question asked whether the hard consensus provided by the CMAA algorithm reflected their subjective consensus. All junior developers agreed that it did. The second question inquired about the perceived efficiency in building the consensus. All of them found the consensus-building process to be efficient. The junior developers announced that they were very satisfied with the decision result and pointed out three benefits of the CMAA consensus-building process in their view:

1. The clarification conferences provided eye-openers for all of them.
2. Sharing their concerns and knowledge during the clarification conferences helped them to agree on resolutions, and they valued the transparency that CMAA provided.
3. Before consensus-building, the junior developers had preferences regarding certain feature ideas, but throughout the process they were able to review the ideas and criteria more objectively.

This feedback shows that CMAA consensus-building was accepted by this group, which provides an answer to research Question 5 on page 11 concerning the acceptance of the process by decision-makers.

9.3 Conclusion

In both case studies, the decision-makers declared their acceptance of the consensus-building process and satisfaction with the result. This provides a positive answer to research Question 5 on page 11 about whether decision-maker groups will accept the new consensus-building framework, despite its leaving the majority of the discrepancies unresolved.

Both groups valued the clarification conferences for uncovering important information that were crucial to reaching an improved decision. They played an important part in the group's subjective feeling that the consensus-building supported them finding the correct decision as defined by Definition 4 on page 5.

The first case study in this chapter illustrated an efficient path towards consensus using CMAA; the decision-makers needed only four steps to soft consensus and eight steps to a hard consensus, and the entropy-optimal path to hard consensus required only four steps. This is a practical demonstration of the large amount of Slack (Definition 5 on page 10) that is present in CMAA's large combinatorial spaces and of Hypothesis 1 on page 10, which claims that only a fraction of the discrepancies need to be resolved to reach consensus. The results also support the goals of Efficiency (Goal 2 on page 7) and Revelation-Invariance (Goal 1 on page 7).

The second case study resulted in a longer consensus path, which required the resolution of half of the discrepancies present in the overall decision. This was due to the group choosing intermediate-value resolutions at several steps. The entropy-optimal resolution is usually found at one of the two extremes in a discrepancy, but in this case, only three of the 18 resolutions were entropy-optimal, and the entropy actually increased at five steps. This corroborates the findings of the simulation study in Section 7.4.3 that the lower the proportion of entropy-optimal resolutions, the longer the consensus path becomes. For this particular case, a more conservative consensus-building strategy that favours discrepancies whose resolution would at least not increase the entropy significantly might have been more efficient.

In Section 7.4.3, it was hypothesised that realistic consensus paths are likely to contain more than 20% but fewer than 100% entropy-optimal resolutions. Both case studies provided evidence for this assumption: in the biotechnology case study, 50% of the resolutions were entropy-optimal, and in the feature idea case study, the proportion was 28%.

10
Summary and Conclusions

In this chapter, the insights provided by the research are summarised, and an overall conclusion is drawn. Furthermore, opportunities for future research topics are pointed out.

10.1 Summary

10.1.1 Revisiting the goals of the thesis

Multi-criteria group decisions fall into two categories: competitive decisions, where decision-makers have conflicting agendas, and cooperative decisions, where decision-makers share a common goal. Until now, research in multi-criteria decision-making has focused almost exclusively on competitive decisions. This thesis presents the first multi-criteria decision-analysis and consensus-building algorithm that is explicitly designed for use with cooperative decisions.

The goals of the thesis were:

- **Consensus-building**
  The main goal is to create a consensus-building algorithm that provides an efficient, revelation-invariant consensus for cooperative decisions (Goals 1 and 2 on page 7).
- **New consensus metric**
  The consensus-building approach should be driven by a new consensus metric (Goal 3 on page 7) that is output-based and can detect revelation-invariant consensus (Definition 3 on page 4).
- **Consensus degree prediction**
  Develop a decision analysis algorithm that can distinguish conflicts by their impact on the consensus degree and predict the new level of consensus of each potential resolution (Goal 4 on page 7).
- **Method independence**
  The new consensus-building approach should be compatible with any commonly used

MCDM method in order to give decision-makers the freedom to choose the most appropriate method for their situation (Goal 5 on page 8).

- **Automation**
  In addition to the formal specifications of the algorithms, a further goal was to identify some best practices for digital facilitation (Goal 6 on page 9).

The thesis set out to investigate two hypotheses on which the success of the new algorithms depended:

- **Slack**
  A hard consensus can be reached without having to resolve most of the conflicting evaluations (Hypothesis 1 on page 10).
- **Sampling accuracy**
  A small sample of the combinatorial space is sufficient to achieve accurate decision analysis using Monte Carlo simulation (Hypothesis 2 on page 10).

Five additional research questions were introduced, which can be found on page 11.

The thesis achieved all its goals, found evidence that strongly supported both hypotheses, and provided answers to all five research questions. The following sections briefly summarise these achievements.

10.1.2 Novel CMAA framework

CMAA is based on two fundamental insights for cooperative multi-criteria group decision-making. The first insight is that evaluation conflicts should be treated as indicators of information asymmetry. Each piece of unshared information has the potential to alter the performance of the alternatives. Thus, it is essential to preserve each input evaluation, rather than submerge it in an aggregated value. The second insight is that no currently established input or output metric is able to detect a revelation-invariant consensus, making them unsatisfactory for cooperative groups. Instead, a new metric is required that incorporates the advantages of both input and output metrics and detects a revelation-invariant consensus.

These two insights motivate the following developments. To preserve all input evaluations, a combinatorial space is created. Each combination represents a possible agreement among all decision-makers when all information has been shared. While it would be too laborious to consider each combination manually, a machine can analyse the combinatorial space and compute the preferredness for each alternative across all combinations in a few seconds. This analysis can reveal the overall performance of an alternative for the entire combinatorial space or for a partial combinatorial space.

Analysing the partial combinatorial space that is generated by a conflict resolution allows for the prediction of the new consensus degree, if the group were to choose that resolution. This variable reveals to the decision-makers which potential resolutions would create more or less clarity about the most-preferred alternatives.

The new entropy-based consensus metric is used to measure the definitiveness of the most-preferred alternative in two ways: for the entire combinatorial space and for the reduced spaces corresponding to each potential conflict resolution. This metric powers an efficient consensus-building process by indicating when a revelation-invariant consensus is achieved, how close the group is to achieving it, and which conflict resolutions have the greatest potential to advance consensus. The consensus-building process is structured in a way that allows for facilitation either by a computer or by a human facilitator who is supported by a digital assistant. This concept was elucidated through various ideas and mock-ups, ultimately fulfilling Goal 6 on page 9.

Table 10.1: Comparison of consensus metrics

	Input-based	Output-based	CAE-based
Considers all input evaluations	✓	✗	✓
Perceived consensus	✗	✓	✓
Revelation-invariant consensus	✗	✗	✓

Table 10.1 compares the two known types of consensus metric and the new CAE metric regarding the three key requirements for cooperative groups. Input metrics have the advantage of considering all input evaluations for their proximity measure, but the resulting consensus can be deficient. Note that the consideration of all input evaluations for input metrics means fusing inputs to some average value. The output metric, on the other hand, considers the proximity of individual preferredness of alternatives, but cannot identify the inputs that produced these outputs. Neither type of metric can detect a revelation-invariant consensus. The CAE metric fulfills all three requirements and therefore best meets a cooperative group's needs.

10.1.3 Performance studies

A series of simulation experiments was conducted to study the behaviour of the CMAA framework. One of the initial studies addressed Hypothesis 2 on page 10, investigating whether analysing a random sample of the combinatorial space instead of the entire space is sufficient. This was crucial, because even small decisions can generate combinatorial spaces of a size that makes analysis intractable. The study demonstrated that 10,000 combinations are sufficient to achieve a 95% confidence interval with a half-width of 0.01 for the rank-1 acceptabilities. The worst-case scenario extends the consensus-building process by only one step, thus supporting Hypothesis 2 on page 10.

The consensus path can become lengthy, if the decision-makers repeatedly fail to exclude a low-value alternative. However, in these cases, a most-preferred alternative has already been clearly identified. To expedite the consensus-building process, a soft consensus threshold was introduced for the CAE metric. Studies were performed to determine an appropriate value for the threshold that balances efficiency with the reliability of the recommendation, supporting the efficiency Goal 2 on page 7.

The performance of CMAA consensus-building was studied from four different perspectives. First, the influence of different problem sizes on consensus performance revealed that consensus-building is highly efficient, requiring the resolution of only a fraction of conflicting evaluations. The number of steps to consensus grows linearly with the number of alternatives and criteria and sub-linearly with the number of decision-makers. The second perspective focused on comparing optimal versus non-optimal resolutions at each consensus step. It was found that consensus was fastest for entropy-optimal resolutions, and was still satisfactory with a realistic proportion of non-optimal resolutions. A third experiment found that even when conflicts were resolved by compromises, consensus could be achieved fairly quickly. The larger the proportion of cooperative resolutions, the faster the consensus process converges.

Lastly, comparing consensus convergence across different MCDM methods indicated that all methods reached consensus quickly.

The performance of the CAE metric was compared to an input metric in two scenarios. In the first scenario, CMAA consensus-building was applied, revealing that the CAE metric converges significantly faster than the input metric, which even failed to achieve a soft consensus in some cases. In the second scenario, an input-based consensus-building process showed that this approach hardly improved the CAE metric and took even longer to achieve consensus compared to the CAE metric applied in CMAA consensus-building. The CAE metric's outstanding performance, achieving consensus faster despite being both a more demanding and a more useful metric, supports both Goal 3 on page 7 and Goal 2 on page 7.

Finally, a study was conducted to measure the amount of Slack in a group decision. It found that even after achieving a revelation-invariant consensus, a very large fraction of the initial combinatorial space remained (and did not need to be treated). On average, the combinatorial space was still at 99.99% of its initial size when consensus was reached. This provides additional evidence for the Slack-Hypothesis 1 on page 10.

10.1.4 Case studies and comparison with studies from the literature

CMAA was applied retroactively to eight previously published case studies of single-step group decisions from the scientific literature that employed a variety of decision methods. In each case, the CMAA initial analysis reproduced the most-preferred alternative fom the original study. Furthermore, it showed that a consensus-building process would have been able to achieve revelation-invariance in a small number of clarification steps.

Two new case studies were performed using the CMAA framework. The first case study was conducted with the founders of a biotechnology startup, who wanted to select their next development project. They achieved a hard consensus for one of the projects with eight clarification steps. It is worth noting that the most-preferred alternative at consensus was different to the one at the start, reflecting a significant alignment in their mental models. They were satisfied with the recommendation that CMAA produced, and they felt that the process was efficient. The second case study was conducted with a group of software developers, who had to select the next feature to be developed for a software tool. Here, the consensus threshold was reached in 12 steps, even though many of the resolutions chosen by the decision-makers were similar in character to compromises. They were also satisfied by the recommendation. Both case studies answer Research Question 5 on page 11 in the positive.

10.1.5 Recommendations for practical application

Seven recommendations were made for facilitators and decision groups concerning the practical application of CMAA:

1. **Unbiased clarification conferences**
   To avoid introducing bias into clarification conferences, it is imperative not to reveal to the decision-makers which resolution would produce the greatest entropy reduction.
2. **Instance sampling**
   $K_{MC} = 10{,}000$ samples are sufficient for the Monte Carlo simulation of the combinatorial space.

3. **Consensus threshold**
   If a soft, but very firm consensus is sufficient, perform consensus-building until the initial entropy has reached $\tau = 0.3$ of its initial size.
4. **Correct decision**
   Groups that require a revelation-invariant consensus should use the CAE consensus metric.
5. **Decision method**
   For faster consensus-building, a non-compensatory decision model such as ABX-Lex is recommended.
6. **Competitive decisions**
   The greater the proportion of agreements (as opposed to compromises) in clarification conferences, the faster the consensus-building process converges. Thus, investment in reaching an agreement is likely to be worthwhile.
7. **Confidence in consensus**
   When a hard consensus has been reached, no further resolution will change the consensus result. However, this might be difficult for decision-makers to understand, in which case the facilitator should explain it to them carefully in order to maintain their confidence in the recommendation.

10.2 Discussion of the results and their implications

10.2.1 Advantages and potential

The CMAA framework offers a unique approach to analysing multi-criteria group decisions and building consensus. The advantages and potential of the new framework are summarised below:

- **CMAA analysis**
  CMAA provides a detailed analysis with several new analytical variables, including the CAE consensus metric. The analysis offers information on each input evaluation and its contribution to the performance of all alternatives, highlighting interdependencies.
- **CAE consensus metric**
  CMAA introduces a new consensus metric that considers the impact of all inputs affecting each alternative's performance. Compared to existing metrics, it is the most strict. The CAE metric can detect a revelation-invariant consensus, measuring the actual agreement among decision-makers about the performance of alternatives. No current input or output metric can achieve this.
- **Analytical prediction**
  A key advantage of CMAA's analysis is its ability to predict the consensus degree for every potential conflict resolution. The differentiation between inactive discrepancies (which do not change the consensus degree) and active discrepancies is useful for facilitators to understand the decision structure better. For active discrepancies, the sensitivities can be computed, indicating which conflict resolutions have a significant impact on each alternative. The most important aspect of this prediction is identifying the evaluation conflict that can improve consensus by the greatest amount, because it enables an efficient consensus-building heuristic.

- **Consensus-building**
  CMAA's consensus-building algorithm efficiently finds short consensus paths for co-operative decisions. When considering compromises for competitive decisions, it also provides efficient paths. This thesis employs a greedy heuristic, but the analytical variables allow for other heuristics to be developed in the future which cater to a group's specific needs or can be adapted during the process.
- **In practice**
  CMAA consensus-building can be implemented as a digital assistant to the group's facilitator. This was tested multiple times, including in the two case studies in Chapter 9. In both cases, the decision-makers accepted the consensus-building process and its recommendation. Furthermore, the framework is directly compatible with most decision methods and requires only minor adjustments for others.
- **CMAA threshold decision method**
  Subjective judgements are necessary to determine when an objective measurement meets a satisfaction level. CMAA enables a new multi-criteria group decision method that offers significant advantages for this type of decision. Decision-makers assign satisfaction levels to objective measurements. This reduces the cognitive load on the decision-makers and can reduce the risk of inconsistencies and manipulation.
- **Potential**
  CMAA has the potential to become a standard approach for multi-criteria group decisions and consensus-building, especially for cooperative decisions. In practice, the consensus-building process can be fully automated for application in distributed and/or asynchronous group decisions.

10.2.2 Limitations

The CMAA framework was developed with some simplifications and assumptions, which are important for understanding the context and limitations of the thesis results.

- When creating the combinatorial space, only input evaluations were considered. However, in practice, a group may resolve a conflict with a new preference or judgement that is not one of the ones initially provided by the decision-makers. Consequently, the simulated consensus process may have produced somewhat optimistic path lengths.
- The current consensus heuristic is based on predicting the performance of all alternatives for the next consensus step. This prediction focuses on achieving a local optimum, which may not lie on a globally most efficient consensus path. Such a scenario was demonstrated in the originally published article on CMAA (Goers and Horton, 2023a).
- The case study in Section 9.2 revealed cautious group behaviour, where the group avoided choosing extreme input evaluations for most resolutions. In these situations, the consensus heuristic might not provide the most efficient path to consensus.
- The CMAA framework is designed for cooperative decisions, aiming to resolve conflicts through agreement to overcome information asymmetry. The underlying premise is that pooling all significant information within the group will lead to a correct decision. However, in competitive contexts where compromises are made, this premise may not hold, in which case the correct recommendation is not defined and therefore cannot be guaranteed.

- To enable the use of AHP in conjunction with CMAA, a minimal version of the method was developed. This version reduces the number of pairwise comparisons and does not require a consistency ratio. However, this non-traditional use of AHP may not be acceptable to some users.

10.3 Outlook

The CMAA framework is an original development by the author, and there still remain many areas for exploration and future improvements.

The Identification Rule used in the consensus-building experiments was an optimistic heuristic that selected the resolution with the greatest potential for improving the consensus metric. In practice, however, it is unlikely that the decision-makers will choose the entropy-optimal resolution at every step. A more conservative heuristic could be developed that should perform better in practice. One possibility could be to select the discrepancy that has the best expected value for the entropy reduction.

The entropy-optimal resolution of a discrepancy is often one of the two extreme values. For decisions that contain competitive discrepancies, it might be beneficial to also compute the potential entropies for compromise resolutions, rather than just for the judgements and preferences supplied by the decision-makers.

The consensus-building algorithm uses a greedy heuristic that only examines the next iteration. It is known that this does not always result in the shortest consensus path. One possibility would be to look ahead further. However, this would significantly increase the computational complexity of the consensus-building algorithm and may therefore only be suitable for asynchronous decisions.

Research in this thesis indicates that the greater the number of agreement resolutions, the quicker the consensus-building process. Therefore, rather than proposing just one discrepancy, the algorithm could present the human facilitator with a set of entropy-reducing discrepancies. The facilitator could then select the discrepancy with the chances of cooperative resolution, based on observation of the group.

Competitive decisions consist of different stakeholders with often conflicting goals. Some stakeholders may have more in common than others. If it was known beforehand, which decision-makers belong to which stakeholder groups, the consensus heuristic could estimate the probability of conflict resolution as a compromise or agreement. These values could then be taken into account when identifying the next evaluation conflict for resolution, which might provide a more efficient path to consensus.

Shared Mental Model studies have faced challenges in identifying Hidden Profiles, owing to the holistic discussion style. The positive results obtained with CMAA suggest that Hidden Profile research might benefit from multi-criteria decision models, as they generate discussions with a much more limited scope.

The CAE consensus metric and consensus-building algorithm are based solely on the rank-1 acceptabilities. For applications in which the top-n alternatives are being sought, it may be more appropriate to base the consensus metric on the rank-1 to rank-n acceptabilities. In addition, the potential acceptabilities could be computed for the top-n ranks. This generalisation would provide a greater selection of discrepancies for the Identification Rule to choose from and may lead to a faster consensus.

Bibliography

Abastante, F., Corrente, S., Greco, S., Ishizaka, A., and Lami, I. M. (2019). A new parsimonious AHP methodology: Assigning priorities to many objects by comparing pairwise few reference objects. *Expert Systems with Applications*, 127:109–120.

Aguarón, J., Escobar, M. T., Moreno-Jiménez, J. M., and Turón, A. (2019). AHP-group decision making based on consistency. *Mathematics*, 7(3):242.

Aruldoss, M., Lakshmi, T. M., and Venkatesan, V. P. (2013). A survey on multi criteria decision making methods and its applications. *American Journal of Information Systems*, 1(1):31–43.

Asuquo, M. P., Wang, J., Zhang, L., and Phylip-Jones, G. (2019). Application of a multiple attribute group decision making (MAGDM) model for selecting appropriate maintenance strategy for marine and offshore machinery operations. *Ocean Engineering*, 179:246–260.

Bagočius, V., Zavadskas, E., and Turskis, Z. (2014). Multi-person selection of the best wind turbine based on the multi-criteria integrated additive-multiplicative utility function. *Journal of Civil Engineering and Management*, 20(4):590–599.

Bana e Costa, C. A. (1988). A methodology for sensitivity analysis in three-criteria problems: A case study in municipal management. *European Journal of Operational Research*, 33(2):159–173.

Banaeian, N., Mobli, H., Fahimnia, B., Nielsen, I. E., and Omid, M. (2018). Green supplier selection using fuzzy group decision making methods: A case study from the agri-food industry. *Computers & Operations Research*, 89:337–347.

Bebiano, N., Fernandes, R., and Furtado, S. (2020). Reciprocal matrices: properties and approximation by a transitive matrix. *Computational and Applied Mathematics*, 39(50).

Beck, K., Beedle, M., Bennekum, A. v., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R. C., Mellor, S., Schwaber, K., Sutherland, J., and Thomas, D. (2001). Agile Manifesto. https://agilemanifesto.org/.

Bittner, E. A. C. and Leimeister, J. M. (2014). Creating shared understanding in heterogeneous work groups: Why it matters and how to achieve it. *Journal of Management Information Systems*, 31(1):111–144.

Blagojević, B., Nordstroem, E.-M., and Lindroos, O. (2023). A framework for defining weights of decision makers in group decision-making, using consistency between different multicriteria weighting methods. *International Journal of Forest Engineering*, 34(2):130–142.

Blagojevic, B., Srdjevic, B., Srdjevic, Z., and Zoranovic, T. (2016). Deriving weights of the decision makers using AHP group consistency measures. *Fundamenta Informaticae*, 144(3-4):383–395.

Blomkvist, S. (2002). Persona–an overview. *Retrieved November*, 22(2004):15.

Blumenroth, N. (2024). Praxistauglichkeit der CMAA Konsensbildung mit Fuzzy SAW. Master's thesis, Otto-von-Guericke-University of Magdeburg. to be submitted.

Blumenthal, E. (2023). Performanzanalyse des CMAA-AHP-Verfahrens für Gruppenentscheidungen. Bachelor Thesis at the Otto-von-Guericke University of Magdeburg.

Brans, J.-P. and Vincke, P. (1985). Note—a preference ranking organisation method: (the PROMETHEE method for multiple criteria decision-making). *Management science*, 31(6):647–656.

Briggs, R., de Vreede, G.-J., and Nunamaker, J. (2003). Collaboration Engineering with thinkLets to pursue sustained success with group support systems. *Journal of Management Information Systems*, 19(4):31–64.

Briggs, R., Kolfschoten, G., de Vreede, G.-J., and Douglas, D. (2006). Defining key concepts for Collaboration Engineering. In *AMCIS 2006 Proceedings*.

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.

Brodbeck, F. C., Kerschreiter, R., Mojzisch, A., and Schulz-Hardt, S. (2007). Group decision making under conditions of distributed knowledge: The information asymmetries model. *Academy of Management Review*, 32(2):459–479.

Cabrerizo, F. J., Alonso, S., and Herrera-Viedma, E. (2009). A consensus model for group decision making problems with unbalanced fuzzy linguistic information. *International Journal of Information Technology & Decision Making*, 8(01):109–131.

Cabrerizo, F. J., Pérez, I. J., Chiclana, F., and Herrera-Viedma, E. (2017). Group decision making: Consensus approaches based on soft consensus measures. *Fuzzy Sets, Rough Sets, Multisets and Clustering*, pages 307–321.

Chao, X., Kou, G., Peng, Y., and 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.

Chiclana, F., García, J. M. T., Del Moral, M. J., and Herrera-Viedma, E. (2015). Analyzing consensus measures in group decision making. *Procedia Computer Science*, 55:1000–1008.

Chiclana, F., García, J. T., del Moral, M. J., and Herrera-Viedma, E. (2013). A statistical comparative study of different similarity measures of consensus in group decision making. *Information Sciences*, 221:110–123.

Chou, S.-Y., Chang, Y.-H., and Shen, C.-Y. (2008). A fuzzy simple additive weighting system under group decision-making for facility location selection with objective/subjective attributes. *European Journal of Operational Research*, 189(1):132–145.

Cinelli, M., Kadzinski, M., Gonzalez, M., and Roman, S. (2020). How to support the application of multiple criteria decision analysis? Let us start with a comprehensive taxonomy. *Omega*, 96:102261.

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.

De Dreu, C. K. and West, M. A. (2001). Minority dissent and team innovation: the importance of participation in decision making. *Journal of applied Psychology*, 86(6):1191.

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.

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, Q. and Saaty, T. L. (2014). An analytic hierarchy process model of group consensus. *Journal of Systems Science and Systems Engineering*, 23:362–374.

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.

Ducassé, M. and Cellier, P. (2012). The LogicalMulticriteriaSort thinkLet: Logical navigation for fair and fast convergence in multicriteria group decision making. In *Group Decision and Negotiation Conference*.

Eckardt, M. (2023). Untersuchung der Anwendbarkeit des CMAA-AHP-Verfahrens anhand der Reevaluierung wissenschaftlicher Literaturbeispiele. Bachelor Thesis at the Otto-von-Guericke University of Magdeburg.

Evans, G. W. (2016). *Multiple criteria decision analysis for industrial engineering: Methodology and applications*. CRC Press.

Fasolo, B. and Bana e Costa, C. A. (2014). Tailoring value elicitation to decision makers' numeracy and fluency: Expressing value judgments in numbers or words. *Omega*, 44:83–90.

Fishburn, P. C. (1974). Exceptional paper—lexicographic orders, utilities and decision rules: A survey. *Management science*, 20(11):1442–1471.

Gao, J., Guo, F., Ma, Z., Huang, X., and Li, X. (2020). Multi-criteria group decision-making framework for offshore wind farm site selection based on the intuitionistic linguistic aggregation operators. *Energy*, 204:117899.

Goers, J., Eckardt, M., Blumenthal, E., and Horton, G. (2024). CMAA-AHP: Combinatorial Multicriteria Acceptability Analysis with the Analytical Hierarchy Process. *Central European Journal of Operations Research*, pages 1–28.

Goers, J. and Horton, G. (2023a). Combinatorial Multi-Criteria Acceptability Analysis: A decision analysis and consensus-building approach for cooperative groups. *European Journal of Operational Research*, 308(1):243–254.

Goers, J. and Horton, G. (2023b). Project selection in a biotechnology startup using combinatorial acceptability analysis. *Decision Making: Applications in Management and Engineering*, 6(2):828–852.

Goers, J. and Horton, G. (2024a). On the Combinatorial Acceptability Entropy consensus metric for Multi-Criteria Group Decisions. *Journal for Group Decision and Negotiation*. submitted for publication.

Goers, J. and Horton, G. (2024b). On the treatment of conflict resolutions in Multi-Criteria Group Decisions. *Information Fusion*. submitted for publication.

Goers, J. and Horton, G. (2024c). Threshold Combinatorial Multicriteria Acceptability Analysis for group decisions with subjective interpretations of objective measurements. In *Proceedings of the Modelling, Data Analytics and AI in Engineering*. accepted for presentation.

Harker, P. (1987). Shortening the comparison process in the AHP. *Mathematical Modelling*, 8:139–141.

Herrera-Viedma, E., Cabrerizo, F. J., Kacprzyk, J., and Pedrycz, W. (2014). A review of soft consensus models in a fuzzy environment. *Information Fusion*, 17:4–13.

Herrera-Viedma, E., Herrera, F., and Chiclana, F. (2002). A consensus model for multiperson decision making with different preference structures. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 32(3):394–402.

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.

Horton, G., Goers, J., and Knoll, S. W. (2016). How not to select ideas for innovations: A critique of the scoring method. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 237–246. IEEE.

Huang, Y.-S., Liao, J.-T., and Lin, Z.-L. (2009). A study on aggregation of group decisions. *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research*, 26(4):445–454.

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

Hwang, C.-L., Yoon, K., Hwang, C.-L., and Yoon, K. (1981). Methods for multiple attribute decision making. *Multiple attribute decision making: methods and applications a state-of-the-art survey*, pages 58–191.

Jeffries, R. (2008). How should user stories be written? https://ronjeffries.com/xprog/blog/how-should-user-stories-be-written/.

Kacprzyk, J. and Fedrizzi, M. (1988). A 'soft'measure of consensus in the setting of partial (fuzzy) preferences. *European Journal of Operational Research*, 34(3):316–325.

Kaliszewski, I. and Podkopaev, D. (2016). Simple additive weighting — a meta model for multiple criteria decision analysis methods. *Expert Systems with Applications*, 54:155–161.

Kangas, A., Kangas, J., Lahdelma, R., and Salminen, P. (2006). Using SMAA-2 method with dependent uncertainties for strategic forest planning. *Forest Policy and Economics*, 9(2):113–125.

Kangas, J. and Kangas, A. (2003). Multicriteria approval and SMAA-O in natural resources decision analysis with both ordinal and cardinal criteria. *Journal of Multi-Criteria Decision Analysis*, 12(1):3–15.

Keeney, R. (2009). The foundations of collaborative group decisions. *International Journal of Collaborative Engineering*, 1(1/2):4–18.

Kendall, M. G. and Smith, B. B. (1939). The problem of m rankings. *The annals of mathematical statistics*, 10(3):275–287.

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.

Knoll, S., Plumbaum, T., Hoffman, J., and De Luca, E. W. (2010). Collaboration ontology: Applying collaboration knowledge to a generic group support system. In *Group Decision & Negotiation 2010 Conference (GDN2010)*, pages 1–37.

Knoll, S. W., Chelvier, R., and Horton, G. (2007). Formalised online creativity using thinXels. In *10th European Conference on Creativity and Innovation*, volume 19.

Koksalmis, E. and Kabak, Ö. (2019). Deriving decision makers' weights in group decision making: An overview of objective methods. *Information Fusion*, 49:146–160.

Kroese, D. P., Brereton, T., Taimre, T., and Botev, Z. I. (2014). Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6):386–392.

Kumar, K. and Chen, S.-M. (2022). Group decision making based on advanced intuitionistic fuzzy weighted heronian mean aggregation operator of intuitionistic fuzzy values. *Information Sciences*, 601:306–322.

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

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

Lai, Y.-L. and Ishizaka, A. (2020). The application of multi-criteria decision analysis methods into talent identification process: A social psychological perspective. *Journal of Business Research*, 109:637–647.

Laukkanen, S., Palander, T., Kangas, J., and Kangas, A. (2005). Evaluation of the multicriteria approval method for timber-harvesting group decision support. *Silva Fennica*, 39(2):249–264.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Lin, C., Kou, G., Peng, Y., and Alsaadi, F. E. (2022). Aggregation of the nearest consistency matrices with the acceptable consensus in AHP-GDM. *Annals of Operations Research*, pages 1–17.

Liu, W., Xiao, S., Browne, J. T., Yang, M., and Dow, S. P. (2018). ConsensUs: Supporting multi-criteria group decisions by visualizing points of disagreement. *ACM Transactions on Social Computing*, 1(1):1–26.

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.

Majumdar, A., Sarkar, B., and Majumdar, P. (2005). Determination of quality value of cotton fibre using hybrid ahp-topsis method of multi-criteria decision-making. *Journal of the Textile Institute*, 96(5):303–309.

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

Mandler, M., Manzini, P., and Mariotti, M. (2012). A million answers to twenty questions: Choosing by checklist. *Journal of Economic Theory*, 147(1):71–92.

Memari, A., Dargi, A., Akbari Jokar, M. R., Ahmad, R., and Abdul Rahim, A. R. (2019). Sustainable supplier selection: A multi-criteria intuitionistic fuzzy TOPSIS method. *Journal of Manufacturing Systems*, 50:9–24.

Mi, X., Liao, H., Lia, Y., Lin, Q., Lev, B., and Al-Barakati, A. (2006). Green supplier selection by an integrated method with stochastic acceptability analysis and multimoora. *Technological and Economic Development of Economy*, 26(3):549–572.

Moogk, R. (2012). Minimum viable product and the importance of experimentation in technology startups. *Technology Innovation Management Review*, 2(3):23–26.

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

Mukherjee, N., Zabala, A., Huge, J., Nyumba, T. O., Adem Esmail, B., and Sutherland, W. J. (2018). Comparison of techniques for eliciting views and judgements in decision-making. *Methods in Ecology and Evolution*, 9(1):54–63.

Munier, N., Hontoria, E., et al. (2021). *Uses and Limitations of the AHP Method*. Springer.

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.

Noori, A., Bonakdari, H., Morovati, K., and Gharabaghi, B. (2018). The optimal dam site selection using a group decision-making method through fuzzy TOPSIS model. *Environment Systems and Decisions*, 38:471–488.

Oliva, G., Setola, R., and Scala, A. (2017). Sparse and distributed Analytic Hierarchy Process. *Automatica*, 85:211–220.

Olson, D. L., Mechitov, A. I., and Moshkovich, H. M. (1998). Cognitive effort and learning features of decision aids: Review of experiments. *Journal of Decision Systems*, 7(1-4):129–146.

Osmond, R., Mollahoseini, Z., Singh, J., Gautam, A., Seethaler, R., Golovin, K., and Milani, A. (2021). A group multicriteria decision making with anova to select optimum parameters of drilling flax fibre composites: A case study. *Composites Part C: Open Access*, 5:100156.

Ossadnik, W., Schinke, S., and Kaspar, R. (2016). Group aggregation techniques for Analytic Hierarchy Process and Analytic Network Process: A comparative analysis. *Group Decision and Negotiation*, 25(2):421–457.

Ozer, I. (2007). Multi-criteria group decision making methods using AHP and integrated web-based decision support systems. Master's thesis, University of Ottawa, Faculty of Graduate and Post-Doctoral Studies.

Pagone, E., Salonitis, K., and Jolly, M. (2020). Automatically weighted high-resolution mapping of multi-criteria decision analysis for sustainable manufacturing systems. *Journal of Cleaner Production*, 257:120272.

Palomares, I., Crosscombe, M., Zhen-Song, C., and Lawry, J. (2018). Dual consensus measure for multi-perspective multi-criteria group decision making. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 3313–3318.

Palomares, I., Martinez, L., and Herrera, F. (2013). A consensus model to detect and manage noncooperative behaviors in large-scale group decision making. *IEEE Transactions on Fuzzy Systems*, 22(3):516–530.

Pang, J., Liang, J., and Song, P. (2017). An adaptive consensus method for multi-attribute group decision making under uncertain linguistic environment. *Applied Soft Computing*, 58:339–353.

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.

RealtimeBoard Inc. (2024). Miro. Online collaborative white board software: https://miro.com/.

Rolle, L. (2022). Schneller einen harten Konsens bei ABX-Lex Gruppenentscheidungen erreichen - Vergleich zweier Combinatorial Multi-Criteria Acceptability Analysis-Heuristiken. Master's thesis, Otto-von-Guericke University of Magdeburg.

S. Dhiman, H., Deb, D., S. Dhiman, H., and Deb, D. (2020). Multi-criteria decision-making: An overview. *Decision and control in hybrid wind farms*, pages 19–36.

Saaty, T. L. (1972). An eigenvalue allocation model for prioritization and planning. *Energy management and policy center, University of Pennsylvania*, 28:31.

Saaty, T. L. (1990). How to make a decision: The Analytic Hierarchy Process. *European Journal of Operational Research*, 48(1):9–26.

Saaty, T. L. (2004). Decision making—the Analytic Hierarchy and Network Processes (AHP/ANP). *Journal of Systems Science and Systems Engineering*, 13:1–35.

Salih, M. M., Zaidan, B., Zaidan, A., and Ahmed, M. A. (2019). Survey on fuzzy TOPSIS state-of-the-art between 2007 and 2017. *Computers & Operations Research*, 104:207–227.

Salo, A. and Hämäläinen, R. (1997). On the measurement of preferences in the Analytic Hierarchy Process. *Journal of Multi-Criteria Decision Analysis*, 6(6):309–319.

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.

Schwaber, K. (1997). Scrum development process. In *Business Object Design and Implementation: OOPSLA'95 Workshop Proceedings 16 October 1995, Austin, Texas*, pages 117–134. Springer.

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

Sodenkamp, M. A., Tavana, M., and Di Caprio, D. (2018). An aggregation method for solving group multi-criteria decision-making problems with single-valued neutrosophic sets. *Applied Soft Computing*, 71:715–727.

Sorooshian, S. (2018). Group decision making with unbalanced-expertise. In *Journal of Physics: Conference Series*, volume 1028, page 012003. IOP Publishing.

Spearman, C. (1987). The proof and measurement of association between two things. *The American Journal of Psychology*, 100(3/4):441–471.

Specchia, M. L., Frisicale, E. M., Carini, E., Di Pilla, A., Cappa, D., Barbara, A., Ricciardi, W., and Damiani, G. (2020). The impact of tumor board on cancer care: evidence from an umbrella review. *BMC Health Services Research*, 20:1–14.

Spicelogic Inc. (2022). Analytic Hierarchy Process software. http://www.spicelogic.com. (Accessed: 09.12.2022).

Srdjevic, Z., Blagdjevic, B., and Srdjevic, B. (2011). AHP based group decision making in ranking loan applicants for purchasing irrigation equipment: a case study. *Bulgarian Journal of Agricultural Science*, 17(4):531–543.

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.

Susskind, L. E., McKearnen, S., and Thomas-Lamar, J. (1999). *The consensus building handbook: A comprehensive guide to reaching agreement*. Sage publications.

Tervonen, T. and Lahdelma, R. (2007). Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research*, 178(2):500–513.

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.

Thakkar, J. J. (2021). *Multi-criteria decision making*, volume 336. Springer.

Tindale, R. S. and Winget, J. R. (2019). Group decision-making. In *Oxford Research Encyclopedia of Psychology*. Oxford Research Encyclopedias, Psychology.

Tran, T. N. T., Felfernig, A., and Le, V. M. (2023). An overview of consensus models for group decision-making and group recommender systems. *User Modeling and User-Adapted Interaction*, pages 1–59.

Triantaphyllou, E. and Triantaphyllou, E. (2000). *Multi-criteria decision making methods*. Springer.

Tundjungsari, V., Istiyanto, J. E., Winarko, E., and Wardoyo, R. (2012). Achieving consensus with individual centrality approach. *arXiv preprint arXiv:1203.5570*.

Turcksin, L., Bernardini, A., and Macharis, C. (2011). A combined AHP-PROMETHEE approach for selecting the most appropriate policy scenario to stimulate a clean vehicle fleet. *Procedia - Social and Behavioral Sciences*, 20:954–965.

Uddin, S., Ali, S. M., Kabir, G., Suhi, S. A., Enayet, R., and Haque, T. (2019). An AHP-ELECTRE framework to evaluate barriers to green supply chain management in the leather industry. *International Journal of Sustainable Development & World Ecology*, 26(8):732–751.

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

Von Neumann, J. and Ulam, S. (1951). Monte Carlo method. *National Bureau of Standards Applied Mathematics Series*, 12(1951):36.

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

Xue, M., Fu, C., and Yang, S.-L. (2020). Group consensus reaching based on a combination of expert weight and expert reliability. *Applied Mathematics and Computation*, 369:124902.

Yee, M., Dahan, E., Hauser, J. R., and Orlin, J. (2007). Greedoid-based noncompensatory inference. *Marketing Science*, 26(4):532–549.

Youssef, M. I. and Webster, B. (2022). A multi-criteria decision making approach to the new product development process in industry. *Reports in Mechanical Engineering*, 3(1):83–93.

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

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—i. *Information Sciences*, 8(3):199–249.

Zeleny, M. (2012). *Multiple criteria decision making Kyoto 1975*, volume 123. Springer Science & Business Media.

Zhang, G., Dong, Y., and Xu, Y. (2014). Consistency and consensus measures for linguistic preference relations based on distribution assessments. *Information Fusion*, 17:46–55.

Zhang, H., Dong, Y., Chiclana, F., and 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.

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.

Zhang, X. and Xu, Z. (2014). Deriving experts' weights based on consistency maximization in intuitionistic fuzzy group decision making. *Journal of Intelligent & Fuzzy Systems*, 27(1):221–233.

List of Figures

List of Tables