Goers, Jana; Horton, Graham

On the Combinatorial Acceptability Entropy Consensus Metric for Multi-Criteria Group Decisions

Group Decision and Negotiation

Provided in Cooperation with:
Springer Nature

Suggested Citation: Goers, Jana; Horton, Graham (2024) : On the Combinatorial Acceptability Entropy Consensus Metric for Multi-Criteria Group Decisions, Group Decision and Negotiation, ISSN 1572-9907, Springer Netherlands, Dordrecht, Vol. 33, Iss. 5, pp. 1247-1268, https://doi.org/10.1007/s10726-024-09891-z

Group Decision and Negotiation (2024) 33:1247–1268
https://doi.org/10.1007/s10726-024-09891-z

On the Combinatorial Acceptability Entropy Consensus Metric for Multi-Criteria Group Decisions

Jana Goers[1] · **Graham Horton**[1]

Accepted: 11 June 2024 / Published online: 30 July 2024

Abstract

In group decisions, achieving consensus is important, because it increases commitment to the result. For cooperative groups, Combinatorial Multicriteria Acceptability Analysis (CMAA) is a group decision framework that can achieve consensus efficiently. It is based on a novel Combinatorial Acceptability Entropy (CAE) consensus metric. As an output measure, the CAE metric is unique in its ability to identify the evaluations that have the greatest impact on consensus and to prevent premature consensus. This paper is intended to complement the original CMAA publication by providing additional insights into the CAE consensus metric. The design requirements for the CAE algorithm are presented, and it is shown how these requirements follow from the properties of cooperative decisions. The CAE-based consensus-building algorithm is contrasted both qualitatively and quantitatively with a representative example of the conventional input distance and input averaging approach to multi-criteria consensus-building. A simulation experiment illustrates the ability of the CAE-based algorithm to converge quickly to the correct decision as defined for cooperative decisions. The metric is able to meet a new, more stringent definition of hard consensus. The CAE approach highlights the need to distinguish between competitive and cooperative group decisions. Attention in the literature has been paid almost exclusively to the former type; the CAE approach demonstrates the greater efficiency and effectiveness that can be achieved with an approach that is designed specifically for the latter.

Keywords Multi-criteria group decision-making · Combinatorial multicriteria acceptability analysis · Consensus-building · Information entropy · Shared mental models

✉ Jana Goers
jana.goers@ovgu.de

Graham Horton
graham.horton@ovgu.de

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

1 Introduction

In a multi-criteria group decision, several decision-makers collaborate to select the most-preferred alternative according to a set of performance criteria. Consensus is important because it creates a higher level of commitment to the decision and increases the chances of success of its implementation (De Vreede et al. 2013). Susskind et al. (1999) state: "When decision-makers' preferences and concerns are considered, they are much more likely to actively participate in the implementation of the obtained solution." For this reason, consensus measurement and consensus-building are topics that have given rise to much research, as can be seen in the survey of Cai et al. (2023).

In this paper, we discuss the consensus metric that was introduced in conjunction with the Combinatorial Multicriteria Acceptability Analysis (CMAA) framework (Goers and Horton 2023a). We will refer to it as the *Combinatorial Acceptability Entropy* (CAE) metric. It is unique, in that it is not based on a similarity measure. Although it is an output metric, it enables a cost-minimizing consensus-building heuristic that can reach a high level of consensus with significantly fewer iterations than a typical input averaging and similarity based technique. In the first published case study using CMAA/CAE (Goers and Horton 2023b), a group of engineers reached a hard consensus on a product development decision with just eight clarification steps. To the authors' knowledge, CAE is the first output metric that makes it possible to identify the decision-maker inputs that have the greatest impact on consensus and thereby provide a heuristic for optimizing the consensus-building iteration. The motivation for this work is to provide some background information about the CAE metric and to compare and contrast its properties with those of a conventional input consensus metric.

1.1 Input and Output Consensus Metrics

Consensus measures can be divided into input and output metrics. These are also known as *coincidence among preferences* and *coincidence among solutions*, respectively Herrera-Viedma et al. (2014). Input metrics measure the degree of similarity between the decision-makers' evaluations (Del Moral et al. 2018), while output metrics measure the difference between their rankings. If the values are identical, the consensus is said to be *hard*; if the difference is small, but non-zero, the consensus is said to be *soft* (Pérez et al. 2018). Soft consensus is interpreted as 'partial agreement' between the decision-makers (Zhang et al. 2019). Herrera-Viedma et al. provide an overview of soft consensus models (Herrera-Viedma et al. 2014). Kacprzyk and Fedrizzi (1988) were the first to point out that even in a 'consensus-committed' group, differences between decision-makers – for example between their value systems – may prevent hard consensus and a certain degree of soft consensus must be sought instead.

The most common type of input consensus metric is the similarity of the decision-makers' judgements to their average value (List 2011; Jj et al. 2015; Liu and Li 2019). Other metrics are based on pairwise distances between judgements (Dong

et al. 2018) or on the variance of each set of judgements (Moral et al. 2017). Del Moral et al. (2018) compare different input-based distance metrics with an output metric and obtained significantly different results for each. Del Moral et al. (2018) study the effect of aggregation operators and consensus metrics on the level of consensus and the speed of the consensus-building process. Tapia et al. (2023) study the Gini index as a dispersion-based, rather than distance-based consensus metric and conclude that it is appropriate for use in the MCGDM context. Tapia et al. (2022) show that the Theil entropy can be used as a consensus measure that behaves similarly to a standard input distance metric. Zhang et al. (2019) balance individual and global consensus measures to reduce the adjustment distance and thereby reduce non-cooperative behavior during consensus-building.

Output consensus is more useful than input consensus, because it is the agreement about the preferredness of alternatives which is ultimately of interest, not agreement about the individual inputs. Del Moral et al. (2018) remark that "[coincidence among solutions] provides a more realistic measure of consensus". Input consensus is thus a means to an end, but not an end in itself. Nevertheless, output consensus is seldom considered in the MCGDM literature. This is presumably because input metrics, which are computed from individual evaluations, provide clues which of these might be adjusted and in what manner. On the other hand, input adjustments with output metrics are limited to general instructions of the type: "If an alternative ranks lower than the group average then adjust the judgements to promote it, and if it ranks higher than the average then adjust the judgements to demote it" (Del Moral et al. 2018).

1.2 Consensus-building Processes

Often in practice, no attempt is made to build consensus. Instead, decision-maker evaluations are simply aggregated to average values (Bagoçius et al. 2014; Memari et al. 2019; Gao et al. 2020). In the case of AHP, the individual priority vectors, rather than the pairwise comparisons, may be averaged (Srdjevic et al. 2011). Some approaches apply weights to the decision-makers according to their expert-level (Ozer 2007), to their level of conformity (Asuquo et al. 2019) or other objective attributes of their inputs (Srdjevic et al. 2011; Koksalmis and Kabak 2019).

Neglecting to build consensus carries the risk that the resulting decision is less than optimal, because extreme evaluations (which are obscured by averaging) may be indicators of important information that is unknown to the majority of the group.

Consensus-building in multi-criteria group decision-making is an iterative process, in which decision-maker evaluations are successively adjusted until the value of the consensus metric reaches a pre-determined threshold. The structure of the process is shown in Fig. 1. Methods vary in their consensus measure (Step 1), the identification of the evaluations to be adjusted (Step 3) and the direction of adjustment (Step 4). Zhang et al. (2019) refer to the algorithms of steps 3 and 4 as the *Identification Rule* and *Direction Rule*, respectively.

Input-averaging methods attempt to improve the consensus measure by having the decision-makers adjust their inputs in the direction of the group average. In some

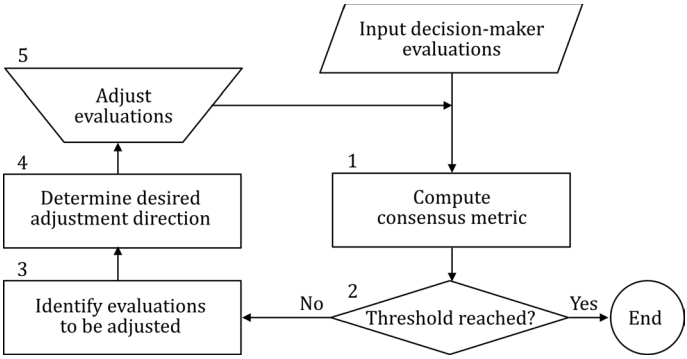

Fig. 1 Structure of a consensus-building process

cases, they are invited or encouraged to adjust their evaluations in the required direction (Herrera-Viedma et al. 2002; Pérez et al. 2018; Zhang et al. 2019; Dong and Saaty 2014), while in other cases, the adjustments may even be made by the algorithm automatically without referring back to the decision-makers at all (Xu 2009; Pang et al. 2017; Dong et al. 2016; Palomares et al. 2013; Lin et al. 2022). The evaluations of decision-makers who refuse to adjust their input may be penalized (Chao et al. 2021; Dong et al. 2016), or even ignored entirely (Kacprzyk and Fedrizzi 1988).

Many studies assume that adjusting an evaluation may incur a political or psychological cost for the affected decision-makers, and several approaches model the cost of consensus-building and propose methods for its reduction. Zhang et al. (2019) consider five different efficiency metrics and develop procedures to minimize them. Guo et al. (2023) present a 'minimum adjustment cost' consensus-building approach that takes the tolerance of each decision-maker towards changing their evaluations into account. Chao et al. (2022) simulate the efficiency of various consensus building approaches. Dong and Xu (2016) present a 'minimum expert consensus model' in order to reduce the number of times decision-makers are requested to adjust their judgements.

Other topics of importance concerning consensus-building are minority opinions, extreme opinions and non-cooperative behavior. Amenta et al. (2020) claim that extreme input values distort the overall group preference and propose an algorithm to de-emphasize their effects. Dong et al. (2016) propose a method to mitigate the negative effects of non-cooperative behavior during consensus-building. Xu et al. (2015) consider minority opinions within the decision-making group; their method increases the weight of minority decision-makers who are able to convince the majority of the validity of their arguments and also reduces the weight of decision-makers who are deemed to be non-cooperative. Gao and Zhang (2022) also reduce decision-maker weights if they are discovered to be behaving non-cooperatively. Due to the rise in social media, interest in decision-making in large groups has grown, and Labella et al. 2018 study the scalability of established consensus-building methods.

1.3 Contributions

The main contributions of this work are:

- Demonstrate that the CAE metric is an effective measure of the quality of output consensus.
- Compare and contrast the CAE metric with the conventional input distance-based approach.
- Demonstrate that, when used in conjunction with CMAA consensus-building, CAE can correctly identify consensus on the most-preferred alternative, even though the conventional metric is still far from reaching its termination criterion.
- Demonstrate that, when used in conjunction with conventional input averaging, CAE can correctly identify a lack of consensus on the most-preferred alternative, even though the conventional metric has reached its termination criterion.

The rest of this paper is organized as follows. Section 2 introduces some notation and terminology and discusses two important distinctions in MCGDM: cooperative versus competitive groups and input consensus versus output consensus. In Sect. 3, we discuss combinatorial acceptability as a measure of the preferredness of an alternative, the use of its entropy as a consensus metric and the resulting consensus-building process. In Sect. 4, we use a simple example to illustrate CAE-based consensus-building, and we present simulation results that compare the new approach with an input averaging consensus-building method. Finally, some concluding remarks are made in Sect. 5.

2 Preliminaries

2.1 Notation and Terminology

Throughout this paper, alternatives are denoted by $a_i, 1 \leq i \leq m$, criteria by $c_j, 1 \leq j \leq n$ and decision-makers by $\mathrm{DM}_k, 1 \leq k \leq d$. Subjective performance evaluations of alternatives with respect to criteria are referred to as *judgements*, and expressions of criteria importance are *preferences*. A multicriteria decision is composed of a set of judgement tasks and a set of preference tasks. The overall performance of an alternative will be referred to as its *preferredness*. The alternative with the greatest preferredness is the *most-preferred* alternative. In this paper, we consider the task of determining the most-preferred alternative. The algorithms can be easily extended, if a complete ranking is desired.

The judgement by decision-maker DM_k of alternative a_i with respect to criterion c_j is denoted by $\lambda_k(j, i)$, and the preference by decision-maker DM_k for criterion c_j by $\mu_k(j)$. When two or more decision-makers submit different judgements for a particular judgement task, a *judgement discrepancy* is created, and when they submit different preferences for a particular preference task, a *preference discrepancy* is created. For simplicity of presentation, we will only consider judgement discrepancies in this

paper. The algorithms can be easily extended to include preference discrepancies, which are treated in an analogous manner.

A mental model is a "mechanism whereby humans generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states" (Rouse and Morris 1984). If all members of a group have the same mental model of the relevant aspects of an issue, the model is said to be *shared*; if their private models differ, it is *unshared*. In the context of group decision-making, a shared mental model will lead to a unanimous evaluation.

2.2 Cooperative and Competitive Groups

In a group decision, initial judgements and preferences will inevitably differ. Briggs et al. (2005) give five causes of discrepancies:

1. People associate different concepts with the same words or symbols, or they use different words or symbols for the same concepts.
2. Individuals base their evaluations on different assumptions.
3. There are asymmetries in the information that individuals hold.
4. Decision-makers have mutually exclusive individual goals.
5. Decision-makers have different tastes.

These causes can be used to characterize two types of group decision:

Competitive decision. This type contains entries 4 and 5 in the list. Competitive decisions occur when discrepancies are caused by incompatible goals or by differences in fundamental beliefs. Fundamental beliefs may be based on personal taste or political positions, or they may simply be prejudices.
Cooperative decision. This category is composed of entries 1, 2 and 3 in the list. Discrepancies are caused (only) by differences in information and interpretation held by the decision-makers.

In competitive decisions, discrepancies cannot be resolved simply by an exchange of information, because decision-makers have conflicting agendas or irreconcilable beliefs. Instead, achieving consensus requires compromise, which can impose a psychological or political cost on the decision-makers. Furthermore, input consensus may not reflect actual agreement among the decision-makers: they adjust their judgements and preferences in order to reduce the numerical distance between them, but their opinions remain unchanged.

Several features that have been proposed for MCGDM methods were designed with competitive decisions in mind. These include decision-maker weighting, the assumption that making adjustments incurs costs, and, most importantly, promoting the group average as the desired consensus value. The first two features are not needed in cooperative decisions, and the third is counter-productive, since the

mental model behind an extreme judgement might be the one that convinces the whole group when shared.

By contrast, in a cooperative decision, decision-makers resolve discrepancies quickly, once they have heard all the relevant information and arguments. Examples of groups that match or come close to this ideal are physicians on a tumor board selecting a treatment for a cancer patient (Specchia et al. 2020) and a team of engineers and designers deciding on the next set of features to be added to a product.

For a cooperative group, a correct decision exists; it is the one that all decision-makers will agree on once they have created a shared understanding of the relevant issues (De Vreede et al. 2013). Keeney (2009) describes the cooperative consensus-building process as follows: "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 practice, decisions may be of mixed type: some judgement or preference tasks may be cooperative while others are competitive. The former can be resolved quickly and uncontroversially by creating a shared understanding, while the latter will require negotiation and compromise.

2.3 Input Consensus and Output Consensus

Input consensus measures the degree of similarity between the inputs of the decision-makers. A typical input consensus metric is the average of the distances between each decision-maker's judgement and the corresponding group mean:

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

where $\overline{\lambda}(j,i)$ is the average of the $\lambda_k(j,i)$ over all k.

Output consensus measures the degree of similarity between the results proposed by each decision-maker. The consensus degree can then be measured as the sum of the distance between each decision-maker's ranking and the group average ranking:

$$CM_O = \frac{1}{m \cdot d} \sum_{i=1}^{m} \sum_{k=1}^{d} |r_k(i) - \overline{r}(i)| \,, \tag{2}$$

where $r_k(i)$ is the rank determined by decision-maker DM_k for alternative a_i, and $\overline{r}(i)$ is the corresponding average rank over all decision-makers. Alternatively, Kendall's coefficient of concordance or Spearman's rank correlation coefficient may be used.

Clearly, input consensus is a sufficient condition for output consensus: if all decision-maker inputs coincide, then their rankings of the alternatives will also coincide. However, input consensus is not a necessary condition for output consensus; decision-makers may agree on the ranking of the alternatives, even though their inputs differ significantly.

Figure 2 (left) shows a minimally sized example of this phenomenon with $m = n = d = 2$. The decision model is Simple Additive Weighting (Kaliszewski and Podkopaev 2016), the criteria weights are both 0.5, and the range of permissible judgement values is [0, 1]. All judgements by decision-maker DM_1 are at the high end of the range, and all judgements by decision-maker DM_2 are at the low end. Both decision-makers prefer a_1 over a_2, and the metric CM_O would signal a hard output consensus. However, an input similarity measure would return a very poor degree of consensus; for example, the input averaging consensus index of Eq. (1) gives $CM_I = 0.35$ (from the range [0, 0.5]).

A second problem with input consensus is that it may fail to detect disunity between the decision-makers about the decision result. In Fig. 2 (right), both decision-makers submit judgements of similar size, resulting in $CM_I = 0.05$, which represents a high level of soft consensus. (Zhang et al. terminate the consensus-building process at $CM_I = 0.1$ (Zhang et al. 2019).) However, decision-maker DM_1 clearly prefers alternative a_1, while DM_2 prefers a_2, which an output metric such as Eq. (2) will correctly characterize as a low level of consensus.

3 The CAE Consensus Model

3.1 Design Considerations

The combinatorial acceptability entropy consensus model was developed as a result of the following observation:

Observation 1 In a cooperative decision, all decision-makers will submit the same judgement $\lambda(j, i)$ or preference $\mu(j)$, when their mental models for it have been shared.

The agreed value resulting from the shared mental model may correspond to one of the submitted ones, or it may be a new value that results from a synthesis of different arguments.

From Observation 1, the following 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*

	a_1 $\{DM_1, DM_2\}$	a_2 $\{DM_1, DM_2\}$
c_1	$\{0.9, 0.2\}$	$\{0.8, 0.1\}$
c_2	$\{0.9, 0.2\}$	$\{0.8, 0.1\}$

	a_1 $\{DM_1, DM_2\}$	a_2 $\{DM_1, DM_2\}$
c_1	$\{0.2, 0.1\}$	$\{0.1, 0.2\}$
c_2	$\{0.2, 0.1\}$	$\{0.1, 0.2\}$

Fig. 2 Comparisons of input averaging and acceptability entropy. Left: strong output consensus but weak input consensus; Right: strong input consensus but weak output consensus

judgement and preference. In this sense, a cooperative decision has a unique 'correct' solution.

The correct solution is predicated on the knowledge available to the decision-makers at the time the decision is made (Kahneman 2002). It does not imply absolute correctness in the sense of an oracle.

Corollary 2 *In a cooperative decision, any private 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.

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

This means that a decision method that prescribes adjustment corrections may lead to an incorrect solution in the sense of Corollary 1.

From these Corollaries follow three design requirements for a consensus-building algorithm: it must be able to produce the correct decision in the sense of Corollary 1, from which it follows that it must be unbiased in the sense of Corollary 3 and assign equal importance to each judgement or preference. From these requirements follows that the method will be able to pass the 'Unique Argument Test' (Horton and Goers 2021): If there is a clinching argument for a judgement or preference that is known to only 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.

3.2 Combinatorial Acceptability as a Measure of Preferredness

The CAE consensus metric is based on the assumption that any combination of judgements and preferences might be the one that remains when all mental models have been shared and the corresponding discrepancies resolved. Therefore, all decision-maker inputs are equally valid, and every combination of them, or *instance*, should, in principle, be taken into consideration when computing the preferredness of each alternative.

The *rank r acceptability* b_i^r of alternative a_i is the proportion of instances for which a_i achieves rank r. Some decision models can return multiple alternatives on the same rank, resulting in $\sum_i b_i^r > 1$. In this case, the values are normalized to a probability vector. The concept of rank acceptability was introduced by Lahdelma and Salminen (Lahdelma and Salminen 2001) in the context of Stochastic Multicriteria Acceptability Analysis, where it was computed using samples from continuous distributions representing uncertainty in the inputs.

The acceptability acc_i of alternative a_i is a measure of its preferredness. It is a function of its rank acceptabilities. One such function is the linear combination

$$\text{acc}_i = \sum_{r=1}^{m} \alpha_r \cdot b_i^r \,, \tag{3}$$

where the coefficients α_r are chosen to satisfy $\alpha_1 \geq \alpha_2 \ldots \geq \alpha_m$. One possibility are the inverse weights $\alpha_r = 1/r$ (Lahdelma and Salminen 2001). Since this paper is concerned with the identification of the most-preferred alternative only, we can use the simplest approach:

$$\alpha_r = \begin{cases} 1, \text{ if } r = 1 \\ 0 \quad \text{otherwise} \end{cases} . \tag{4}$$

We thus have $acc_i = b_i^1$, and we can use the terms 'acceptability' and 'rank-1 acceptability' interchangeably.

Figure 3 illustrates the acceptability computation. Instances of the decision problem are formed by combining decision-maker inputs. Each instance is passed to the chosen decision method, which returns a ranking of the alternatives for that instance.

Two sets of counter variables are required. The first records for each alternative a_i and ranking position r the number of instances in which a_i attains rank r. After normalization, these values are input into Eq. (3) to obtain the acceptabilities of each alternative.

The second set has cardinality $n \times m \times d$. It contains a counter for each alternative and decision-maker input in each discrepancy and records the number of instances in which alternative a_i is most-preferred when the discrepancy at (c_j, a_i) is resolved to $\lambda_k(j, i)$. These represent the contributions of each individual decision-maker input to the acceptability of each alternative and are referred to *current judgement acceptabilities*. They are also used to compute the proportion of instances for which each alternative would become most-preferred if the corresponding input were to be selected in the next step. These are referred to as *potential judgement acceptabilities*. When preference discrepancies are present, *current preference acceptabilities* and *potential preference acceptabilities* are defined analogously.

The number of instances generated by a decision can be extremely large. In such cases, Monte Carlo simulation can be used instead of complete enumeration of the instance space. 10,000 random instances are required to obtain a 95% confidence interval half-width of 0.01 for each rank acceptability value, which provides

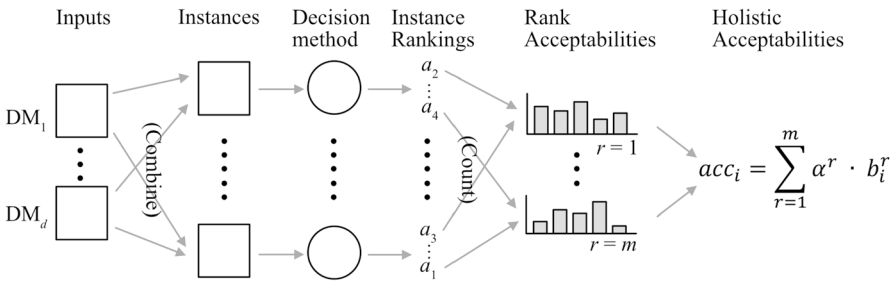

Fig. 3 Computation of the combinatorial acceptabilities

sufficient accuracy in practice (Goers and Horton 2023b). The computation time to analyse 100,000 instances is less than one second on a notebook computer and therefore does not present an obstacle in practice.

During the generation of the instances, repeated input values are only considered once. For example, if three decision-makers submit the judgements $\{1, 1, 3\}$, the algorithm treats them as the set $\{1, 3\}$ when generating the instances. This is a consequence of Corollary 2: input values should be treated equally.

Rank acceptabilities can also be interpreted probabilistically; b_i^r is the probability that alternative a_i will achieve rank r, if it is assumed that discrepancies are resolved randomly, where each judgement or preference in a discrepancy is equally likely to be chosen by the group.

The acceptability computation is independent of the decision method used; it simply generates combinations of the inputs required by the decision method and uses the rankings that it produces.

3.3 Acceptability Entropy as a Consensus Metric

The idea behind the CAE consensus metric is to measure the similarity of the vector of acceptabilities to a standard unit vector. This is achieved using the information entropy h:

$$h = -\sum_i p_i \cdot \log_2(p_i)\,, \tag{5}$$

where p is a probability vector. Entropy is a measure of the uncertainty in a discrete random variable (Shannon 1948). The maximum value $h_{max} = \log_2(p_i)$ is obtained when the p_i are equal. Its minimum value is 0, which is obtained when $p_{i^*} = 1$ for some i^* and all other p_i are 0. Entropy is a measure of the difference between a given probability vector and a standard unit vector. It is this property that makes it suitable as an output consensus metric, when the p_i represent the preferredness of the alternatives.

By substituting the acceptabilities acc_i into Eq. (5), h is the *current entropy* in the decision for the most-preferred alternative. This is a measure of the degree to which the acceptability of one alternative dominates that of the others; a value $h \ll 1$ indicates that the acceptability of one alternative is significantly greater than that of any of the others. Substituting the potential acceptabilities for a given decision-maker input $\lambda_k(j, i)$ into Eq. (5) gives the *potential (judgement) entropy* $\hat{h}(\lambda_k(j, i))$ for that input. This shows to what degree the acceptability of one alternative would dominate, if that resolution were to be chosen.

Figure 4 shows six probability vectors sorted from left to right according to their entropies, which are equally spaced from 2.32 down to 0.1. The value 2.32 is the maximum possible entropy for a vector with five elements. It is achieved when all elements of the vector have the same value and are therefore least suited to identify a most-preferred alternative. On the right, the vector with entropy 0.1 is almost indistinguishable from the standard unit vector (0, 0, 0, 1, 0), which would be achieved at $h = 0$.

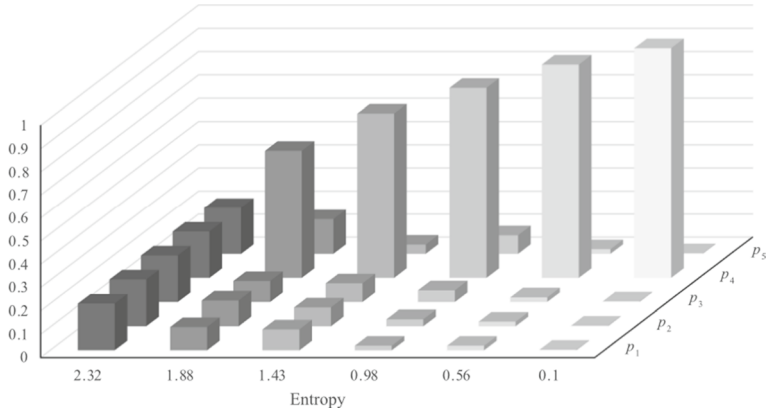

Fig. 4 Entropies of six different acceptability vectors

The entropy metric has a high sensitivity for large h values: the step from $h = 2.32$ to $h = 1.88$ marks the transition from indistinguishable acceptabilities to a clear favorite. By contrast, the same-sized entropy decrement from $h = 0.56$ to $h = 0.1$ marks only a small perceived difference in the acceptabilities. Figure 4 suggests that, if the vectors represent rank-1 acceptabilities, a threshold of $h < 1.0$ is sufficient to declare a very good soft consensus on the most-preferred alternative. At $h = 0.1$, we have $p_4 = 0.99$, so the probability of a different alternative becoming most-preferred if the remaining discrepancies were to be resolved randomly is only 1%. Since the maximum possible value for the entropy h_{max} grows slowly with the number of alternatives, a threshold of $h < 0.3 \cdot h_{max}$ is recommended for the general case.

In Fig. 2 (right), the rank-1 acceptabilities are $b_1^1 = b_2^1 = 0.5$, since both alternatives are most-preferred in eight of the 16 instances. Accordingly, the current entropy of this acceptability vector has the maximum possible value of $h = 1.0$, indicating the lowest possible degree of consensus.

Figure 2 (left) shows a more interesting situation. The rank-1 acceptabilities are $b_1^1 = 11/16$ and $b_2^1 = 5/16$, since 11 of the 16 possible combinations of inputs yield a_1 as the most-preferred alternative, and 5 yield a_2. The current entropy is $h = 0.9$ (in the range $0 \leq h \leq 1$), which suggests that these acceptabilities do not provide a reliable recommendation for the most-preferred alternative. This result may seem counter-intuitive, because both decision-makers individually prefer alternative a_1 over a_2, and a typical output metric such as Eq. (2) would report a hard consensus. However, the assumptions on which the CAE metric is based shows that this conclusion might be premature. If decision-maker DM_2 had the more convincing arguments regarding the performance of a_1 with respect to both criteria, and DM_1 had better arguments concerning a_2, then a_2 would become the most-preferred alternative. Schulz-Hardt and Mojzisch call the tendency of groups to (incorrectly) agree on an alternative before all arguments have been considered *premature consensus* (Schulz-Hardt and Mojzisch 2012). The ability to identify input combinations for which each alternative might become

most-preferred (or indeed achieve any given rank) is an important feature of the CAE consensus metric.

The entropy of the acceptabilities is an output consensus metric. However, its interpretation differs from that of classical metrics, and Eq. (2) is not applicable. Instead of 'coincidence among solutions', i.e., the degree of similarity between the decision-makers' individual rankings, the CAE metric measures the degree to which the entire set of instances generated by the inputs produces the same most-preferred alternative. In this sense, soft consensus may be interpreted as 'partial agreement'. Hard consensus implies not only that all decision-makers prefer the same alternative, but also that all combinations of their inputs do so as well. It is therefore a more stringent requirement.

The potential entropies make the following consensus-building heuristic possible (with reference to the steps in Fig. 1):

1. Compute the potential entropies for each available resolution in each discrepancy.
2. Use $h < 0.3 \cdot h_{max}$ for a 'very firm' consensus or $h = 0$ for a hard consensus.
3. Identify the discrepancy containing the lowest potential entropy. All decision-makers are involved in its resolution.
4. This step is empty: the algorithm does not prescribe an adjustment direction, and the decision-makers are free to agree on a unanimous value.

The identification rule at Step 3 is a heuristic for minimizing the overall cost of the consensus iteration. It will usually (but not always Goers and Horton (2023a)) choose the shortest path to consensus. Ideally, the shared mental model will lead to the entropy-optimal resolution at each step, because it improves the consensus metric by the greatest amount. Other resolutions will cause the entropy to decrease by a smaller amount or even to increase. If the decision-makers are unable to agree on a resolution, the discrepancy becomes unavailable for future consideration and the algorithm returns to Step 3. Each failed resolution increases the number of iterations needed to achieve consensus by one. Non-entropy-optimal resolutions may (but do not necessarily) increase the number of steps.

In the case of a cooperative decision, the agreement on a unanimous value in Step 4 is the result of the shared mental model. In the case of a competitive decision, it will require negotiation and/or compromise.

Table 1 shows a summary of the principal differences between the CAE/CMAA algorithm and conventional input distance/input-averaging methods.

4 Numerical Examples and Comparisons

4.1 Illustration of the Consensus-building Process

This section contains an illustration of the CAE-based consensus-building process, which is described in detail in (Goers and Horton 2023a). A synthetic decision

Table 1 Comparison of consensus metrics and consensus-building approaches

Attribute	CAE/CMAA	Input distance/input averaging
Target decision type	Cooperative	Competitive
Type of consensus metric	Output	Input
Interpretation of outlying judgements	Can signal important information	Misrepresent the group opinion
Decision method is applied to ...	Multiple combinations of inputs	Average of inputs
Preferredness of alternatives	Strength of support (acceptability)	According to decision model
Basis for consensus	Shared understanding	Numerical compromise
Consensus metric	Entropy of the acceptabilities	Similarity between inputs
Meaning of hard consensus	All instances yield the same most-preferred alternative	Coincidence of judgements and preferences
Consensus cost metric	Number of iterations	Various (#Iterations, #adjusted alternatives, total adjustment, ...)
Identification rule	Minimize potential entropy	Minimize cost
Direction rule	None	Towards group average

problem is used, which is constructed solely for illustrative purposes. For simplicity of presentation, a minimally sized problem with $m = n = d = 2$ is used, and the criteria have equal weights. Performance values are in the range [0, 1]. The decision model is Simple Additive Weighting (Kaliszewski and Podkopaev 2016): alternative scores are the scalar product of the criteria weight vector and the corresponding vector of performance values.

Figure 5 shows the first step of the consensus-building process. The initial decision-maker input is shown in the table 'Judgements'. Each pair of curly brackets contains the judgements by the two decision-makers for a particular (c_j, a_i) pair. This $\{DM_1, DM_2\}$ arrangement of the values in curly brackets is continued in the

Judgements

	a_1	a_2
c_1	{0.9, 0.2}	{0.5, 0.8}
c_2	{0.3, 0.7}	{0.6, 0.2}

Acc. b_1^1=0.5, b_2^1=0.5
Curr. entropy $h = 1.0$

Potential judgm. acc. a_1	
{0.83, 0.13}	{0.63, 0.38}
{0.38, 0.63}	{0.38, 0.63}

Potential judgm. acc. a_2	
{0.13, 0.88}	{0.38, 0.63}
{0.63, 0.38}	{0.63, 0.38}

Potential entropy

{0.54, 0.54}	{0.95, 0.95}
{0.95, 0.95}	{0.95, 0.95}

Identif. rule selects (c_1, a_1).
(c_1, a_1) is resolved to 0.8.

Fig. 5 First consensus-building step for the example decision

following tables. There are discrepancies at all four locations, giving a total of $2^4 = 16$ instances. Eight instances return a_1 as the preferred alternative and eight return a_2. The resulting rank-1 acceptabilities are therefore $b_1^1 = b_2^1 = 8/16 = 0.5$. Both alternatives are equally preferred, the consensus is minimal, and the current entropy is correspondingly maximal at $h = 1.0$.

The potential judgement acceptabilities show the proportion of instances for which each alternative would be preferred if each resolution were to be applied. For example, resolving the discrepancy at (c_2, a_2) to the value 0.2 proposed by DM_2 would increase the acceptability of a_1 to the potential acceptability 0.63 and reduce the acceptability of a_2 to the potential acceptability 0.38.

The potential entropies are computed from the potential acceptabilities resulting from each resolution. The lowest value of 0.54 would be obtained by resolving the discrepancy at (c_1, a_1) either to 0.9 or to 0.2, so this discrepancy is passed to the decision-makers for resolution. After sharing their mental models of the performance of alternative a_1 with respect to criterion c_1, the decision-makers agree on a compromise judgement of 0.8, which was chosen arbitrarily for this example.

Figure 6 shows the second step of the consensus-building process. The judgement discrepancy at (c_1, a_1) has been replaced by the resolved value of 0.8, improving the current entropy to $h = 0.76$. This is slightly worse than the entropy of 0.54 that would have been attained by resolution to either of the initial judgements. The entries at (c_1, a_1) in the other tables are now undefined, because the discrepancy no longer exists. Eight instances remain, of which seven return a_1 as preferred alternative, and one returns a_2. Resolving to DM_2's input at both (c_2, a_1) and (c_2, a_2) would lead to hard consensus, with potential acceptabilities of 1.0 for a_1 and 0.0 for a_2, and a corresponding potential entropy of $\hat{h}(\lambda_2(c_2, a_1)) = \hat{h}(\lambda_2(c_2, a_2)) = 0.0$. The decision-makers resolve the discrepancy at (c_2, a_2) to the entropy-minimizing choice 0.2.

Figure 7 shows the final situation after Step 2. Two discrepancies have been resolved, and two are unresolved. Four instances remain, which all yield a_1 as the preferred alternative. The two remaining discrepancies can safely be ignored, because no resolution would change the most-preferred alternative.

Judgements

	a_1	a_2
c_1	{0.8, 0.8}	{0.5, 0.8}
c_2	{0.3, 0.7}	{0.6, 0.2}

Acc. b_1^1: 0.875, b_2^1: 0.25
Curr. entropy $h = 0.76$

Potential judgm. acc. a_1

-	{0.80, 0.75}
{0.60, 1.00}	{0.60, 1.00}

Potential judgm. acc. a_2

-	{0.20, 0.25}
{0.40, 0.00}	{0.40, 0.00}

Potential entropy

-	{0.72, 0.81}
{0.97, 0.00}	{0.97, 0.00}

Identif. rule selects (c_1, a_2).
(c_1, a_2) is resolved to 0.2.

Fig. 6 Second consensus-building step for the example decision

Judgements		
	a_1	a_2
c_1	{0.8, 0.8}	{0.5, 0.8}
c_2	{0.3, 0.7}	{0.2, 0.2}

Acc. b_1^1=1.0, b_2^1=0.0
Curr. entropy $h = 0.0$

Fig. 7 Result for the example decision

4.2 Illustration of Convergence Speed

The second example is used to demonstrate the speed of convergence that the CAE-based consensus-building algorithm can achieve. Table 2 shows a group decision with $m = n = 6$ and $d = 3$. The criteria weights w_j are fixed. The decision-maker judgements yield 26 trivalent discrepancies, 8 bivalent discrepancies and two unanimous judgements, which generate $3^{26} \cdot 2^8 \cdot 1^2 > 6.51E14$ instances. The decision was analysed using Monte Carlo simulation with 10,000 random instances. The initial rank-1 acceptabilities b_i^1 are (0.39, 0.12, 0.02, 0.45, 0.02, 0.02). Their (current) entropy is 1.75 (in the range [0, 2.58]) indicating that a most-preferred alternative should not be chosen on the basis of this data.

Despite the very large number of combinations of inputs, only three resolutions are needed to achieve hard consensus; these are shown in bold typeface at locations (c_1, a_4), (c_3, a_4) and (c_4, a_4). Resolving the discrepancy at location (c_1, a_4) to 8 reduces the current entropy to $h = 1.03$. Next resolving the discrepancy at (c_3, a_4) to 7 yields $h = 0.31$ and $b_4^1 = 0.96$, which is already a very strong recommendation for a_4 as most-preferred alternative. After the third resolution at location (c_4, a_4) to 8, consensus is achieved, with $b_4^1 = 1.0$ and current entropy $h = 0.0$. More than $3.6E13$ instances remain, and it can be easily verified that these all produce alternative a_4 as most-preferred. The 31 remaining discrepancies therefore have no effect on the most-preferred alternative and can be ignored. If rank acceptabilities for $r > 1$ are needed, the consensus-building process can be continued for $r = 2, 3, \ldots$ in turn.

Table 2 Decision problem with three steps to consensus

	w_j	a_1	a_2	a_3	a_4	a_5	a_6
c_1	6	{5, 3, 6}	{9, 2, 7}	{8, 3, 3}	{2, 2, **8**}	{3, 3, 3}	{7, 7, 4}
c_2	5	{3, 7, 3}	{0, 6, 0}	{0, 5, 2}	{8, 8, 8}	{1, 5, 7}	{7, 5, 0}
c_3	4	{5, 9, 7}	{2, 7, 8}	{8, 1, 5}	{**7**, 5, 0}	{2, 8, 9}	{5, 2, 1}
c_4	3	{8, 8, 7}	{6, 3, 1}	{1, 0, 4}	{6, **8**, 1}	{7, 4, 0}	{3, 6, 2}
c_5	2	{8, 6, 4}	{5, 1, 7}	{9, 3, 2}	{7, 7, 6}	{1, 9, 4}	{2, 8, 4}
c_6	1	{4, 9, 6}	{5, 5, 3}	{6, 9, 3}	{9, 6, 5}	{9, 1, 2}	{9, 5, 2}

4.3 Comparison of Convergence Behavior

A simulation study was carried out to compare the convergence behavior of the CAE-based consensus-building algorithm with a representative input averaging method. For the latter, the MACRP 3 algorithm of Zhang et al. (2019) was chosen. This is the input averaging algorithm that most closely approximates CMAA in structure, in that in each iteration it selects one judgement task and presents it to all decision-makers for consideration. Referring to Fig. 1, MACRP 3 is specified as follows:

1. The consensus metric is the average distance between each decision-maker judgement and its corresponding group average according to Eq. (1).
2. Zhang et al. (2019) use a threshold value of $CM_I \leq 0.1$.
3. The identification rule selects the discrepancy (j^*, i^*) containing the judgement that is most distant from the group average:

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

   All decision-makers are required to adjust their judgements.
4. The direction rule requires that the adjusted judgement for decision-maker DM_k should be between $\lambda_k(j,i)$ and $\overline{\lambda}(j^*,i^*)$, i.e., between no adjustment at all and the group average for the selected judgement.

The MACRP 3 algorithm does not consider preference discrepancies, and it assumes that all judgements are in the range [0,1].

MACRP 3 and CMAA were applied to four differently sized decisions using Monte Carlo simulation with 1,000 randomly generated decisions for each problem size. The number of criteria and decision-makers were set to $n = d = 5$, and the number of alternatives m was tested for 5, 10, 15 and 20. Criteria weights were equal, and performance judgements were uniformly distributed on the interval [0, 1]. Each method was simulated until its consensus measure reached its threshold. These were $h \leq 2\%$ and $CM_I \leq 20\%$ of their maximum possible values, respectively. For MACRP 3, this corresponds to $CM_I \leq 0.1$ in each case, because the maximum value is independent of m. For CMAA, the absolute value of the threshold varies with m, but $h < 0.1$ is satisfied in all four cases. The value 0 represents hard consensus for both metrics. Optimal resolutions were applied in both simulations: the maximum adjustment to $\overline{\lambda}(j^*,i^*)$ was used in MACRP 3, and entropy-minimizing resolutions were used in CMAA.

Figure 8 shows the results obtained with CMAA. The shading of the curves represents the different problem dimensions from $m = 5$ (lightest) to $m = 20$ (darkest). Round markers represent the current entropy h (primary vertical axis), and square markers represent CM_I (secondary vertical axis). The horizontal dashed line at $CM_I = 20\%$ marks the threshold at which Zhang et al. terminate the MACRP 3 iteration (Zhang et al. 2019). The axes have been scaled so that the initial consensus metrics are approximately co-located.

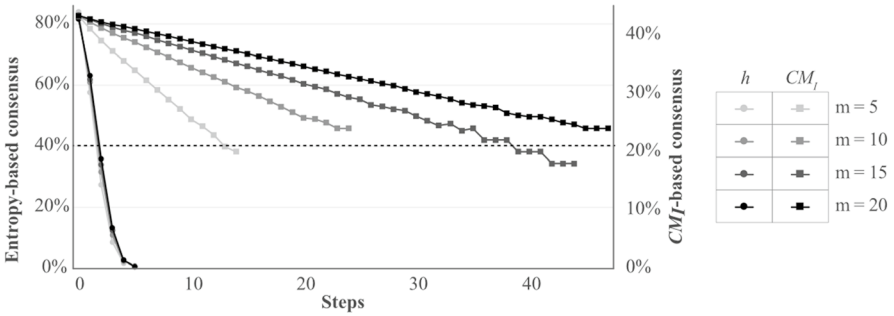

Fig. 8 Consensus processes for CMAA for different decision dimensions

On average, CMAA reaches its threshold in four steps for $m = 5$ and $m = 10$, and in five steps for $m = 15$ and $m = 20$. In all four cases, more than 90% of the instances return the same preferred alternative. The number of steps needed to reach the consensus threshold is very low and essentially independent of the problem size. By contrast, by the time CMAA had identified the most-preferred alternative at step 5, CM_I still showed a weak level of consensus and thus failed to recognize that the most-preferred alternative had been identified.

Figure 9 shows analogous results using the MACRP 3 consensus-building algorithm. The same notation is used as in Fig. 8. As m increases from 5 to 20, the number of steps required to reach a soft consensus at $CM_I \leq 20\%$ grows linearly from 11 to 40, and the CAE metric improves even more slowly. In other words, when the input distance-based iteration terminates, the agreement between the decision-makers regarding the most-preferred alternative has hardly improved at all.

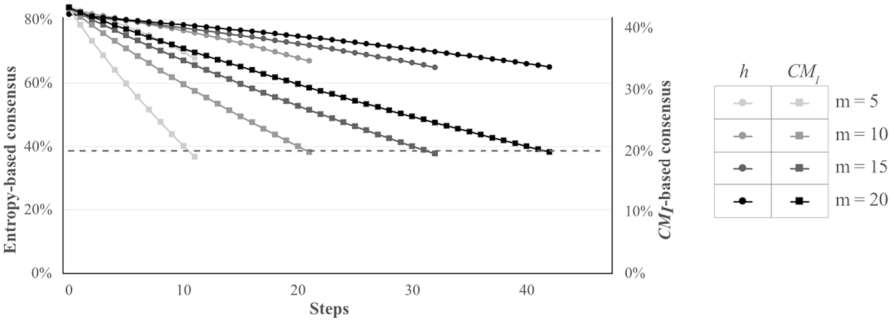

Fig. 9 Consensus processes for MACRP 3 for different decision dimensions

5 Conclusions

5.1 Summary

In this paper, we discuss the CAE consensus metric introduced by Goers and Horton (2023a), and we provide insights into its motivation and design objectives. The new consensus measure can be used in conjunction with the CMAA framework to convert any single-user multi-criteria decision method into a group method without the need for aggregation.

The new metric reveals all possible most-preferred alternatives. This feature can prevent the decision-makers from arriving at a premature consensus when a conventional output metric signals a unique most-preferred alternative.

Although it is an output measure, the CAE metric enables a cost-minimizing consensus-building heuristic. The heuristic is unbiased, in the sense that it does not prescribe the direction of input adjustments. In a cooperative decision, it achieves consensus not only in the sense that the differences between decision-maker inputs and outputs are reduced, but also in the sense that it promotes a shared understanding of the relevant issues.

In a Monte Carlo simulation study, the CAE-based consensus-building method was compared to a conventional approach based on an input distance metric and input-averaging and was found to be significantly more efficient. The combinatorial framework achieves hard consensus in a small number of steps that is almost independent of the number of alternatives. It was found that the CAE-based algorithm can correctly identify consensus about the most-preferred alternative, whereas the conventional metric fails to. When used in conjunction with conventional consensus-building using input averaging, CAE can correctly identify lack of consensus about the most-preferred alternative, even though the conventional metric has reached its termination threshold.

5.2 Outlook

This paper focused on the identification of the most-preferred alternative. The CAE metric and CMAA consensus-building approach can be easily extended, if the top-n alternatives or a complete ranking is needed instead. The consensus-building algorithm can be applied to each rank sequentially, as the established consensus of already-treated ranks is not affected by the processing of the subsequent ones.

The Identification Rule used here is optimistic, in the sense that it selects the discrepancy that contains the lowest entropy resolution, even though there is no guarantee that it will be selected by the decision-makers. The effect of alternative Identification Rules on consensus-building efficiency should be studied, for example selecting the discrepancy with the lowest expected entropy.

The CAE-based Identification Rule only looks one step ahead to the next set of potential entropies. It is also possible to look ahead several resolution steps. This

might also contribute to a reduction in the number of consensus-building steps needed.

The CMAA algorithm was designed for use in cooperative decisions. However, it is also applicable in competitive situations. In this case, a direction rule could be introduced that encourages decision-makers to adjust their inputs towards a particular value (such as the mean of the inputs). This would introduce a bias (such as with any aggregation-based approach), but it would still offer the advantages of the combinatorial approach. Furthermore, convergence speed might be improved by introducing artificial inputs to the discrepancies that represent compromise values. A facilitator could then select a discrepancy for clarification that has the lowest entropy for the compromise resolution over the one with the absolute lowest entropy, if the latter required agreement on an extreme judgement or preference.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Amenta P, Ishizaka A, Lucadamo A, Marcarelli G, Vyas V (2020) Computing a common preference vector in a complex multi-actor and multi-group decision system in analytic hierarchy process context. Annals Op Res 284:33–62

Asuquo MP, Wang J, Zhang L, 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 Eng 179:246–260

Bagoçius V, Zavadskas EK, Turskis Z (2014) Multi-person selection of the best wind turbine based on the multi-criteria integrated additive-multiplicative utility function. J Civil Eng Manag 20(4):590–599

Briggs RO, Kolfschoten GL, De Vreede GJ (2005) Toward a theoretical model of consensus building. In: Eleventh Americas conference on information systems. AMCIS, pp 101–110

Cai Y, Jin F, Liu J, Zhou L, Tao Z (2023) A survey of collaborative decision-making: bibliometrics, preliminaries, methodologies, applications and future directions. Eng Appl Artif Intell 122:106064

Chao X, Kou G, Peng Y, Viedma EH (2021) Large-scale group decision-making with non-cooperative behaviors and heterogeneous preferences: An application in financial inclusion. Eur J Op Res 288(1):271–293

Chao X, Dong Y, Kou G, Peng Y (2022) How to determine the consensus threshold in group decision making: a method based on efficiency benchmark using benefit and cost insight. Annals Op Res, pp 1–35

De Vreede T, Reiter-Palmon R, De Vreede GJ (2013) The effect of shared mental models on consensus. In: Proceedings of the annual hawaii international conference on system sciences, pp. 263–272

Del Moral MJ, Chiclana F, Tapia JM, Herrera-Viedma E (2018) A comparative study on consensus measures in group decision making. Int J Intell Syst 33(8):1624–1638

Del Moral MJ, Tapia JM, Chiclana F, Al-Hmouz A, Herrera-Viedma E (2018) An analysis of consensus approaches based on different concepts of coincidence. J Intell Fuzzy Syst 34(4):2247–2259
Dong Q, Saaty TL (2014) An analytic hierarchy process model of group consensus. J Syst Sci Syst Eng 23:362–374
Dong Y, Xu J (2016) Consensus building in group decision making. Springer, Singapore
Dong Y, Zhang H, Herrera-Viedma E (2016) Integrating experts' weights generated dynamically into the consensus reaching process and its applications in managing non-cooperative behaviors. Decis Support Syst 84:1–15
Dong Y, Zha Q, Zhang H, Kou G, Fujita H, Chiclana F, Herrera-Viedma E (2018) Consensus reaching in social network group decision making: research paradigms and challenges. Knowl Based Syst 162:3–13
Gao Y, Zhang Z (2022) Consensus reaching with non-cooperative behavior management for personalized individual semantics-based social network group decision making. J Op Res Soc 73(11):2518–2535
Gao J, Guo F, Ma Z, Huang X, 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, Horton G (2023) Combinatorial multi-criteria acceptability analysis: a decision analysis and consensus-building approach for cooperative groups. Eur J Op Res 308(1):243–254
Goers J, Horton G (2023) Selection of a product development project in a biotechnology startup using the combinatorial acceptability method. J Decis Mak Appl Manag Eng 6(2):828–852
Guo W, Gong Z, Zhang W-G, Xu Y (2023) Minimum cost consensus modeling under dynamic feedback regulation mechanism considering consensus principle and tolerance level. Eur J Op Res 306(3):1279–1295
Herrera-Viedma E, Herrera F, Chiclana F (2002) A consensus model for multiperson decision making with different preference structures. IEEE Trans Syst Man Cybern Part A Syst Humans 32(3):394–402
Herrera-Viedma E, Cabrerizo FJ, Kacprzyk J, Pedrycz W (2014) A review of soft consensus models in a fuzzy environment. Inf Fusion 17:4–13
Horton G, Goers J (2021) ABX-LEX: an argument-driven approach for the digital facilitation of efficient group decisions. Int J Inf Technol Decis Making 20(01):137–164
Jj Peng, Wang Jq WuXh, Wang J, Chen Xh (2015) Multi-valued neutrosophic sets and power aggregation operators with their applications in multi-criteria group decision-making problems. Int J Comput Intell Syst 8(2):345–363
Kacprzyk J, Fedrizzi M (1988) A -soft- measure of consensus in the setting of partial (fuzzy) preferences. Eur J Op Res 34(3):316–325
Kahneman D (2002) Maps of bounded rationality: a perspective on intuitive judgment and choice. Nobel Prize Lect 8:351–401
Kaliszewski I, Podkopaev D (2016) Simple additive weighting-a meta model for multiple criteria decision analysis methods. Expert Syst Appl 54:155–161
Keeney R (2009) The foundations of collaborative group decisions. Int J Collab Eng 1(1/2):4–18
Koksalmis E, Kabak Ö (2019) Deriving decision makers- weights in group decision making: an overview of objective methods. Inf Fusion 49:146–160
Labella A, Liua Y, Rodríguez R, Martínez L (2018) Analyzing the performance of classical consensus models in large scale group decision making: A comparative study. Appl Soft Comput 67:677–690
Lahdelma R, Salminen P (2001) SMAA-2: stochastic multicriteria acceptability analysis for group decision making. Op Res 49(3):444–454
Lin C, Kou G, Peng Y, Alsaadi F (2022) Aggregation of the nearest consistency matrices with the acceptable consensus in AHP-GDM. Annals Op Res 316:179–195
List C (2011) The theory of judgment aggregation: an introductory review. Synthese 187(1):179–207
Liu W, Li L (2019) Research on the optimal aggregation method of decision maker preference judgment matrix for group decision making. IEEE Access 7:78803–78816
Memari A, Dargi A, Akbari Jokar MR, Ahmad R, Abdul Rahim AR (2019) Sustainable supplier selection: a multi-criteria intuitionistic fuzzy TOPSIS method. J Manuf Syst 50:9–24
Moral MJ, Chiclana F, Tapia J, Herrera-Viedma E (2017) An alternative calculation of the consensus degree in group decision making problems. Procedia Comput Sci 122:735–742
Ozer I (2007) Multi-criteria group decision making methods using AHP and integrated web-based decision support systems. Master's thesis, University of Ottawa

Palomares I, Martinez L, Herrera F (2013) A consensus model to detect and manage noncooperative behaviors in large-scale group decision making. IEEE Trans Fuzzy Syst 22(3):516–530
Pang J, Liang J, Song P (2017) An adaptive consensus method for multi-attribute group decision making under uncertain linguistic environment. Appl Soft Comput 58:339–353
Pérez IJ, Cabrerizo FJ, Alonso S, Dong Y, Chiclana F, Herrera-Viedma E (2018) On dynamic consensus processes in group decision making problems. Inf Sci 459:20–35
Rouse W, Morris N (1984) On looking into the black box. prospects and limits in the search for mental models. Psychol Bull 100
Schulz-Hardt S, Mojzisch A (2012) How to achieve synergy in group decision making: lessons to be learned from the hidden profile paradigm. Eur Rev Soc Psychol 23(1):305–343
Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(4):623–666
Specchia ML, Frisicale EM, Carini E, Di Pilla A, Cappa D, Barbara A, Ricciardi W, Damiani G (2020) The impact of tumor board on cancer care: evidence from an umbrella review. BMC Health Services Res 20
Srdjevic Z, Blagdjevic B, Srdjevic B (2011) AHP based group decision making in ranking loan applicants for purchasing irrigation equipment: a case study. Bulg J Agric Sci 17(4):531–543
Susskind LE, McKearnen S, Thomas-Lamar J (1999) The consensus building handbook: a comprehensive guide to reaching agreement. Sage publications, Thousand Oaks
Tapia JM, Chiclana F, Moral MJ, Herrera–Viedma E (2023) Measuring consensus in group decision-making problems through an inequality measure. In: Dzitac, S., Dzitac, D., Filip, F.G., Kacprzyk, J., Manolescu, M.-J., Oros, H. (eds.) Intelligent Methods Systems and Applications in Computing, Communications and Control, pp. 313–319
Tapia J, Chiclana F, Moral M, Herrera–Viedma E (2022) Entropy based approach to measuring consensus in group decision-making problems. In: International conference on industrial, engineering and other applications of applied intelligent systems, pp. 409–415 . Springer
Xu Z (2009) An automatic approach to reaching consensus in multiple attribute group decision making. Comput Ind Eng 56(4):1369–1374
Xu X-J, Du Z-J, Chen X-H (2015) Consensus model for multi-criteria large-group emergency decision making considering non-cooperative behaviors and minority opinions. Decis Support Syst 79:150–160
Zhang H, Kou G, Peng Y (2019) Soft consensus cost models for group decision making and economic interpretations. Eur J Op Res 277(3):964–980
Zhang H, Dong Y, Chiclana F, Yu S (2019) Consensus efficiency in group decision making: a comprehensive comparative study and its optimal design. Eur J Op Res 275(2):580–598