Group Decis Negot
DOI 10.1007/s10726-015-9460-8

The CAR Method for Using Preference Strength in Multi-criteria Decision Making

Mats Danielson[1,2] · Love Ekenberg[1,2]

Abstract Multi-criteria decision aid (MCDA) methods have been around for quite some time. However, the elicitation of preference information in MCDA processes, and in particular the lack of practical means supporting it, is still a significant problem in real-life applications of MCDA. There is obviously a need for methods that neither require formal decision analysis knowledge, nor are too cognitively demanding by forcing people to express unrealistic precision or to state more than they are able to. We suggest a method, the CAR method, which is more accessible than our earlier approaches in the field while trying to balance between the need for simplicity and the requirement of accuracy. CAR takes primarily ordinal knowledge into account, but, still recognizing that there is sometimes a quite substantial information loss involved in ordinality, we have conservatively extended a pure ordinal scale approach with the possibility to supply more information. Thus, the main idea here is not to suggest a method or tool with a very large or complex expressibility, but rather to investigate one that should be sufficient in most situations, and in particular better, at least in some respects, than some hitherto popular ones from the SMART family as well as AHP, which we demonstrate in a set of simulation studies as well as a large end-user study.

Keywords Multi-criteria decision analysis · Ranking methods · Comparing MCDA methods

✉ Love Ekenberg
lovek@dsv.su.se

Mats Danielson
mad@dsv.su.se

1 Department of Computer and Systems Sciences, Stockholm University, Forum 100, 164 40 Kista, Sweden

2 International Institute for Applied Systems Analysis, IIASA, Schlossplatz 1, 2361 Laxenburg, Austria

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

1 Introduction

A multitude of methods for analysing and solving decision problems with multiple criteria have been suggested during the last decades. A common approach is to make preference assessments by specifying a set of attributes that represents the relevant aspects of the possible outcomes of a decision. Value functions are then defined over the alternatives for each attribute and a weight function is defined over the attribute set. One option is to simply define a weight function by fixed numbers on a normalised scale and then define value functions over the alternatives, where these are mapped onto fixed values as well, after which these values are aggregated and the overall score of each alternative is calculated. The most common form of value function used is the additive model $V(a) = \sum_{i=1}^{m} w_i v_i(a)$, where $V(a)$ is the overall value of alternative a, $v_i(a)$ is the value of the alternative under criterion i, and w_i is the weight of this criterion (cf., e.g., Keeney and Raiffa 1976). The criteria weights, i.e., the relative importance of the evaluation criteria, are thus a central concept in most of these methods and describe each criterion's significance in the specific decision context.

Despite having been around for some decades and despite having turned out to be highly useful (cf., e.g., Bisdorff et al. 2015), multi-criteria decision aids (MCDA), supporting decision making processes are still under-utilised in real-life decision problems. This situation seems to be at least partly due to a combination of lack of convergence between time constraints, and cognitive abilities of decision-makers versus the requirements of the decision aid. Several attempts have been made to solve these issues. For instance, methods allowing for less demanding ways of assessing the criteria, such as ordinal rankings or interval approaches for determining criteria weights and values of alternatives, have been suggested. The underlying idea is, as far as possible, not to force decision-makers to express unrealistic, misleading, or meaningless statements, but at the same time being able to utilise the information the decision-maker is able to supply. Similar issues are present when eliciting and assessing values for alternatives under each criterion.

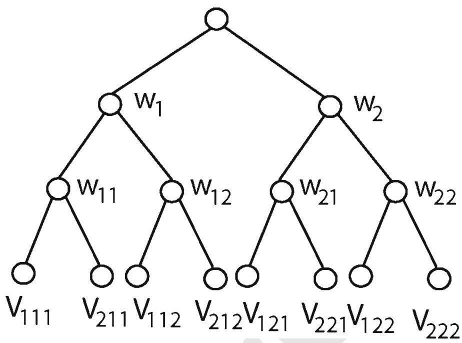
In this article, we provide a brief survey over some central and widespread MCDA methods. We then suggest a new method, the CAR (CArdinal Ranking) method, with the particular aim that weight and value functions can be reasonably elicited while preserving the comparative simplicity and correctness of the approach. Using theoretical simulations and a large user study, we investigate some properties of the method and conclude that, according to the results, it seems to be a highly competitive and applicable method for MCDA as well as group decision making when the opinions of the group members can be weighted in the same manner as the criteria.

2 MCDA Methods

There are several approaches to multi-criteria decision making, the key characteristic being that there are more than one perspective (criterion, aspect) to view the alternatives and their consequences from. For each perspective, the decision-maker must somehow assign values to each alternative on some value scale. Typically, a multi-criteria decision situation could be modelled like the tree in Fig. 1.

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

Fig. 1 A multi-criteria tree



To express the relative importance of the criteria, weights are used restricted by a normalization constraint $\sum w_j = 1$, where w_j denotes the weight of a criterion G_j and the weight of sub-criterion G_{jk} is denoted by w_{jk}. The value of alternative A_i under sub-criterion G_{jk} is denoted by v_{ijk}. Then the weighted overall value of an alternative A_i (from the example in Fig. 1) can be calculated by:

$$E(A_i) = \sum_{j=1}^{2} w_j \sum_{k=1}^{2} w_{jk} v_{ijk},$$

This is straightforwardly generalized and multi-criteria decision trees of arbitrary depth can be evaluated by the following expression:

$$E(A_i) = \sum_{i_1=1}^{n_{i_0}} x_{i i_1} \sum_{i_2=1}^{n_{i_1}} x_{i i_1 i_2} \cdots \sum_{i_{m-1}=1}^{n_{i_{m-2}}} x_{i i_1 i_2 \cdots i_{m-2} i_{m-1}}$$

$$\sum_{i_m=1}^{n_{i_{m-1}}} x_{i i_1 i_2 \cdots i_{m-2} i_{m-1} i_m} x_{i i_1 i_2 \cdots i_{m-2} i_{m-1} i_m},$$

where $x_{\ldots i_{j} \ldots}$, $j \in [1, \ldots, m]$ denote criteria weights and $x_{\ldots i_{j} \ldots 1}$ denote alternative (consequence) values.

One very important practical issue is how to realistically elicit criteria weights (and also values) from actual decision-makers, see Riabacke et al. (2012) for an overview. Considering the judgement uncertainty inherent in all decision situations, elicitation efforts can be grouped into (a) methods handling the outcome of the elicitation by precise numbers as representatives of the information elicited; and (b) methods instead handling the outcome by interval-valued variables. A vast number of methods have been suggested for assessing criteria weights using exact numbers. These range from relatively simple ones, like the commonly used direct rating and point allocation methods, to somewhat more advanced procedures. Generally in these approaches, a precise numerical weight is assigned to each criterion to represent the information extracted from the user. There exist various weighting methods that utilise questioning procedures to elicit weights, such as SMART (Edwards 1977) and SWING weighting (von

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

Winterfeldt and Edwards 1986). However, the requirement for numeric precision in elicitation is somewhat problematic. For instance, significant information is in practice always more or less imprecise in its nature. People's beliefs are not naturally represented in numerically precise terms in our minds (Barron and Barrett 1996b; von Winterfeldt and Edwards 1986). There are several versions within the SMART family of methods with seemingly small differences that have been shown to have important effects for the actual decision making. For instance, SMART and SWING were later combined into the SMARTS method. In general, trade-off methods appear to be quite reasonable for weight elicitation but can nevertheless be very demanding due to the number of required judgments by the decision-maker.

As responses to the difficulties in eliciting precise weights from decision-makers, other approaches, less reliant on high precision on the part of the decision-maker while still aiming at non-interval representations, have been suggested. Ordinal or other imprecise importance (and preference) information could be used for determining criteria weights (and values of alternatives). One approach is to use surrogate weights which are derived from ordinal importance information (cf., eg., Stewart 1993; Arbel and Vargas 1993; Barron and Barrett 1996a, b; Katsikopoulos and Fasolo 2006; Ahn and Park 2006; Sarabando and Dias 2009; Mateos et al. 2014; Aguayo et al. 2014). In such methods, the decision-maker provides information on the rank order of the criteria, i.e., supplies ordinal information on importance, and thereafter this information is converted into numerical weights consistent with the extracted ordinal information. Several proposals on how to convert the rankings into numerical weights exist, e.g., rank sum weights and rank reciprocal weights (Stillwell et al. 1981), and centroid (ROC) weights (Barron 1992). Barron and Barrett (1996b) found the latter superior to the other two on the basis of simulation experiments, but Danielson and Ekenberg (2014b) demonstrate that this holds only under special circumstances and instead suggest more robust weight functions.

In interval-valued approaches to the elicitation problem, incomplete information is handled by allowing the use of intervals (cf., e.g., Danielson and Ekenberg 1998, 2007, where ranges of possible values are represented by intervals and/or comparative statements). Such approaches also put less demands on the decision-maker and are suitable for group decision making as individual differences in importance weights and judgments can be represented by value intervals (sometimes in combination with orderings). Similarly, Mustajoki and Hämäläinen (2005) suggest an extended SMART/SWING method, where they generalize the SMART and SWING methods into a method allowing interval judgments as well. The decision-maker is allowed to enter interval assessments to state imprecision in the judgments. The extracted weight information is represented by constraints for the attributes' weight ratios, which in addition to the weight normalization constraint determine the feasible region of the weights in the interpretational step, see, e.g., Larsson et al. (2005) for a description of such techniques.

There are ways of simplifying the elicitation, e.g., the idea of assigning qualitative levels to express preference intensities in the MACBETH method (Bana e Costa et al. 2002), ranking differences using a delta-ROC approach (Sarabando and Dias 2010) or Simos's method of placing blank cards to express differences (Figueira and Roy 2002). There are also methods such as Smart Swaps with preference programming (Mustajoki

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

and Hämäläinen 2005). Other researchers mix various techniques, as in the GMAA system (Jiménez et al. 2006) which suggests two procedures for weights assessments. The extraction can either be based on trade-offs among the attributes, where decision-makers may provide intervals within which they are indifferent with respect to lotteries and certain consequences, or on directly assigned weight intervals to the respective criteria. The extracted interval values are then automatically computed into an average normalized weight (precise) or a normalized weight interval for each attribute. Such relaxations of precise importance judgments usually seem to provide a more realistic representation of the decision problem and are less demanding for users in this respect (cf., e.g., Park 2004; Larsson et al. 2005). However, there are several computational issues involved that restrict the kind of statements that can be allowed in these representations and often the final alternatives' values have a significant overlap, making the set of non-dominated alternatives too large, which must be handled, e.g., using more elaborated second order techniques (Ekenberg and Thorbiörnson 2001; Ekenberg et al. 2005; Danielson et al. 2007). There are also various approaches to modify some classical, more extreme, decision rules, e.g., the ones discussed in Milnor (1954) and absolute dominance as well as the central value rule. The latter is based on the mid-point of the range of possible performances. Ahn and Park (2008), Sarabando and Dias (2009), Aguayo et al. (2014) and Mateos et al. (2014) discuss these as well as some alternative dominance concepts. Similarly, Puerto et al. (2000) addresses an approach for utilising imprecise information and also applies it to some extreme rules as above as well as to the approach by Cook and Kress (1996). Salo, Hämäläinen, and others have suggested a set of approaches for handling imprecise information in these contexts, for instance the PRIME method for preference ratios (Salo and Hämäläinen 2001).

The handling of decision processes could be efficiently assisted by software packages. The SMART method has been implemented in computer programs (see e.g., Mustajoki et al. 2005). AHP techniques (Saaty 1980) have been implemented in, e.g., EXPERT CHOICE (Krovak 1987). There are many other software packages as well, such as M-MACBETH requiring only qualitative judgements about differences between alternatives (Bana e Costa et al. 1999) and VIP Analysis which allows imprecise scaling coefficients since the coefficients are considered variables subject to a set of constraints (Dias and Clímaco 2000). Computer support is even more necessary for computationally significantly more demanding methods, such as Danielson and Ekenberg (1998), that have to be heavily supported by the use of computer tools (Danielson et al. 2003). In conclusion, there are several approaches to elicitation in MAVT problems and one partitioning of the methods into categories is how they handle imprecision in weights (or values).

1. Weights (or values) can only be estimated as fixed numbers.
2. Weights (or values) can be estimated as comparative statements converted into fixed numbers representing the relations between the weights.
3. Weights (or values) can be estimated as comparative statements converted into inequalities between interval-valued variables.
4. Weights (or values) can be estimated as interval statements.

Needless to say, there are advantages and disadvantages with the different methods from these categories. Methods based on categories 1 and 2 yield computationally

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

simpler evaluations because of the weights and values being numbers while categories 3 and 4 yield systems of constraints in the form of equations and inequalities that need to be solved using optimisation techniques. If the expressive power of the analysis method only permits fixed numbers (category 1), we usually get a limited model that might affect the decision quality severely. If intervals are allowed (categories 3 and 4), imprecision is normally handled by allowing variables, where each y_i is interpreted as an interval such that $w_i \in [y_i - a_i, y_i + b_i]$, where $0 < a_i \leq 1$ and $0 < b_i, \leq 1$ are proportional imprecision constants. Similarly, comparative statements are represented as $w_i \geq w_j$.

In another tradition, using only ordinal information from category 2 and not numbers from category 1, comparisons replace intervals as an elicitation instrument handling imprecision and uncertainty. The inherent uncertainty is captured by surrogate weights derived from the strict ordering that a decision-maker has imposed on the importance of a set of criteria in a potential decision situation. However, we might encounter an unnecessary information loss using only an ordinal ranking. If, as a remedy, we use both intervals and ordinal information, we are faced with some rather elaborate computational problems. Despite the fact that they can be solved, when sufficiently restricting the statements involved (cf. Danielson and Ekenberg 2007), there is still a problem with user acceptance and these methods have turned out to be perceived as too difficult to accept by many decision-makers. Expressive power in the form of intervals and comparative statements lead to complex computations and loss of transparency on the part of the user.

It should also be noted that multi-attribute value theory (MAVT), despite being the main focus in this paper, is not the only suggestion for handling multi-criteria decision problems, even if it is one of the most popular approaches today. Steuer (1984) presents a variety of other methods, including outranking methods, such as ELECTRE (Roy 1968) and PROMETHEE (Brans and Vincke 1985) in various versions, where decision-makers are asked to rank information to find outranking relations between alternatives.

Validation within this field is somewhat difficult, to a large extent due to difficulties regarding elicitation. In this paper, we look at MCDM methods with less complex requirements (categories 1 and 2) but with the dual aim of achieving both high efficiency and wide user acceptance. The question of what constitutes a good method is multifaceted, but it seems reasonable that a preferred method should possess some significant qualities to a higher degree than its rivals:

- *Efficiency* The method should yield the best alternative according to some decision rule in as many situations as possible.
- *Easiness of use* The steps of the method should be perceived as relatively easy to perform.
- *Ease of communication* It should be comparatively easy to communicate the results to others.
- *Time efficiency* The amount of time and effort required to complete the decision making task should be reasonably low.
- *Cognitive correctness* The perceived correctness of the result and transparency of the process should be high.

- *Return rate* The willingness to use the method again should be high.

Below we will investigate to what extent some classes of methods from categories 1 and 2 fulfil these six qualities, where the first is measured in a simulation study (Sect. 4) and the others in a real-life user study (Sect. 5).

3 Three Classes of MCDM Methods

This section discusses three classes of value function methods that allow a relaxation of the requirement of precision, but keeping with simplicity and without resorting to interval or mixed approaches. Instead, we will here discuss if good decision quality can be obtained without significantly increasing either the elicitational or the computational efforts involved, or both, and without making it difficult for a decision-maker to understand the process. To investigate this, we will consider three main classes of methods and compare them in Sects. 4 (theoretically) and 5 (empirically). The classes are:

- Proportional scoring methods, here represented by the SMART family,
- Ratio scoring methods, here represented by the widely used AHP method, and
- Cardinal ranking methods, here represented by the CAR method proposed in this paper.

In the following, if not explicitly stated, we assume a set of criteria $\{G_1, \ldots, G_N\}$ where each criterion G_i corresponds to a weight variable w_i. We also assume additive criteria weights, i.e., $\Sigma w_i = 1$, and $0 \leq w_i$ for all $i \leq N$. We will, without loss of generality, simplify the presentation by only investigating problems with a one-level criteria hierarchy and denote the value of an alternative A_j under criterion C_i by v_{ij}.

3.1 Proportional Scoring

One of the most well-known proportional scoring methods is the SMART family. SMART as initially presented was a seven-step procedure for setting up and analysing a decision model. Edwards (1971, 1977) proposed a method to assess criteria weights. The criteria are then ranked and (for instance) ten points are assigned to w_N, i.e., the weight of the least important criterion. Then, w_{N-1} to w_1 are given points according to the decision-maker's preferences. This way, the points are representatives of the (somewhat uncertain) weights. The overall value $\mathrm{E}(\mathrm{a}_j)$ of alternative a_j is then a weighted average of the values v_{ij} associated with a_j:

$$E(a_j) = \sum_{i=1}^{N} w_i v_{ij} \Big/ \sum_{i=1}^{N} w_i .$$

In an additive model, the weights reflect the importance of one criterion relative to the others. Most commonly, the degree of importance of an attribute depends on its spread (the range of the scale of the attribute), what we call the weight/scale-dualism. This is why elicitation methods like the original SMART, which do not consider the

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

spread specifically, have been criticized (see, e.g., Edwards and Barron 1994). As a result, SMART was subsequently amended with the SWING technique (and renamed SMARTS), addressing the weight/scale-dualism by changing the weight elicitation procedure. Basically, SWING works like this:

- Select a scale, such as positive integers (or similar)
- Consider the difference between the worst and the best outcomes (the range) within each criterion, where the best level is 1
- Imagine an alternative (the zero alternative) with all the worst outcomes from each criterion, thus having value 0 (if we have defined 0 as the lowest value)
- For each criterion in turn, consider the improvement (swing) in the zero alternative by having the worst outcome in that criterion replaced by the best one
- Assign numbers (importance) to each criterion in such a way that they correspond to the assessed improvement from having the criterion changed from the worst to the best outcome

As mentioned above, one approach, which avoids some of the difficulties associated with the elicitation of exact values, is to merely provide an ordinal ranking of the criteria. It is allegedly less demanding on decision-makers and, in a sense, effort-saving. Most current methods for converting ordinal input to cardinal, i.e., convert rankings to exact surrogate weights, employ automated procedures for the conversion and result in exact numeric weights. Edwards and Barron (1994) proposed the SMARTER (SMART Exploiting Ranks) method to elicit the ordinal information on importance before being converted to numbers and thus relaxed the information input requirements from the decision-maker. An initial analysis is carried out where the weights are ordered such as $w_1 > w_2 > \cdots > w_N$ and then subsequently transformed to numerical weights using ROC weights whereafter SMARTER continues in the same manner as the ordinary SMART method.

3.2 Ratio Scoring

One of the most well-known ratio scoring methods is the Analytic Hierarchy Process (AHP). The basic idea in AHP (Saaty 1977, 1980) is to evaluate a set of alternatives under a criteria tree by pairwise comparisons. The process requires the same pairwise comparisons regardless of scale type. For each criterion, the decision-maker should first find the ordering of the alternatives from best to worst. Next, he or she should find the strength of the ordering by considering pairwise ratios (pairwise relations) between the alternatives using the integers 1, 3, 5, 7, and 9 to express their relative strengths, indicating that one alternative is equally good as another (strength = 1) or three, five, seven, or nine times as good. It is also allowed to use the even integers 2, 4, 6, and 8 as intermediate values, but using only odd integers is more common.

Much has been written about the AHP method and a detailed treatment of these is beyond the scope of this article, but we should nevertheless mention two properties that are particularly problematical. Belton and Stewart (2002) have questioned the conversion between scales, i.e., between the semantic and the numeric scale, and the employment of verbal terms within elicitation on the whole have been criticized throughout the years as their numerical meaning can differ substantially between

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

different people (cf., e.g., Kirkwood 1997). There are also particularly troublesome problems with rank reversals known since long (Belton and Gear 1983). Furthermore, the method is cognitively demanding in practice due to the large number of pairwise comparisons required as the number of attributes increases, and there are several variations of AHP, such as in Ginevicius (2009), where the method FARE (Factor Relationship) is suggested in cases when the number of attributes is large in order to reduce the number of required comparisons between pairs of attributes.

3.3 Ordinal and Cardinal Ranking Methods

As with other multi-attribute value based methods, ranking methods contain one alternative (consequence) value part and one criteria weight part. Since weights are more complicated, we will mainly discuss them in this paper. Values are handled in a completely analogous but less complex way. There is no need for values to be transformed into surrogate entities since values are not restricted by an upper sum limit.

Rankings are normally easier to provide than precise numbers and for that reason, various criteria weight techniques have been developed based on rankings. One idea mentioned above is to derive so called surrogate weights from elicitation rankings. The resulting ranking is converted into numerical weights and it is important to do this with as small an information loss as possible while still preserving the correctness of the weight assignments. Stillwell et al. (1981) discuss the weight approximation techniques rank sum and rank reciprocal weights. A decade later, Barron (1992) suggested a weight method based on vertices of the simplex of the feasible weight space. The so called ROC (rank order centroid) weights are the average of the corners in the polytope defined by the simplex $S_w = w_1 > w_2 > \cdots > w_N$, $\Sigma w_i = 1$, and $0 \leq w_i$. The weights are then simply represented by the centroid (mass point) of S_w, i.e.,[1]

$$w_i = 1/N \sum_{j=\mathrm{i}}^{N} \frac{1}{j}, \quad \text{for all i } = 1, \ldots, \text{N}.$$

For instance, in the case of four criteria and where $w_1 > w_2 > w_3 > w_4$, the centroid weight components become $w_1 = 0.5208$, $w_2 = 0.2708$, $w_3 = 0.1458$, $w_4 = 0.0625$. Despite there being a tendency that the highest ranked criterion has a strong influence on the result, as has been pointed out by, e.g., Belton and Stewart (2002), ROC weights are nevertheless representing an important idea regarding averaging the weights involved and in the aggregation of values. Of the conversion methods suggested, ROC weights have gained the most recognition among surrogate weights.

However, pure ranking is sometimes problematic. For example, Jia et al. (1998) state that due to the relative robustness of linear decision models regarding weight changes, the use of approximate weights often yields satisfactory decision quality, but that the assumption of knowing the ranking with certainty is strong. Instead, they believe that there can be uncertainty regarding both the magnitudes and ordering of weights. Thus, although some form of cardinality often exists, cardinal importance

[1] We will henceforth, unless otherwise stated, presume that decision problems are modelled as simplexes S_w generated by $w_1 > w_2 > \cdots > w_N$, $\Sigma w_i = 1$, and $0 = w_i$.

relation information is not taken into account in the transformation of rank orders into weights, thus not making use of available information.

3.4 The Delta Method

Most methods handling imprecise information try to reduce the constraint sets of feasible values, typically by delimiting the available space by linear constraints, through various elicitation procedures and a main problem in that respect is to find a balance between not forcing the decision-maker to say more than is known in terms of precision, but at the same time obtain as much information as is required for the alternatives to be discriminated from each other. Furthermore, the model must be computationally meaningful. As an example, the Delta method is a method for solving various types of decision problems when the background information is numerically imprecise. It has been developed over the years (cf., e.g., Danielson and Ekenberg 1998, 2007; Danielson et al. 2007, 2009; Ekenberg et al. 1995, 2001a, 2005, 2014). The basic idea of the method (relevant for the context in this paper) is to in one way or another construct polytopes for the feasible weights and the feasible alternative values involved and evaluate decision situations with respect to different decision rules.

The Delta method and software has successfully been used in numerous applications regarding everything from tactical hydropower management to business risks and applications for participatory democracy. However, a common factor in the applications of the method that has complicated the decision making process is the difficulties for real-life decision makers to actually understand and use the software efficiently, despite various elicitation interfaces and methods developed, such as in Riabacke et al. (2012), Danielson et al. (2014) and Larsson et al. (2014). Therefore, we have started to investigate how various subsets of the method can be simplified without losing much precision and decision power for general decision situations and can measurably perform well in comparison with the most popular decision methods available at the moment.

3.5 The CAR Method

One of the simplified methods for cardinal ranking is CAR, which extends the idea of surrogate weights as one of the main components (Danielson et al. 2014a; Danielson and Ekenberg 2014b, 2015). The idea is to first assume that there exists an ordinal ranking of N criteria, obtained by any elicitation method such as, for example, SWING.[2] To make this ordering into a cardinal ranking, information should be obtained about how much more or less important the criteria are compared to each other. Such rankings also take care of the problem with ordinal methods of handling criteria that are found to be equally important, i.e., resisting pure ordinal ranking.

We use $>_{\mathrm{i}}$ to denote the strength (cardinality) of the rankings between criteria, where $>_0$ is the equal ranking '='. Assume that we have a user induced ordering $w_1 >_{i_1} w_2 >_{i_2} \cdots >_{i_{n-1}} w_n$. Then we construct a new ordering, containing only the symbols = and >, by introducing auxiliary variables x_{ij} and substituting

[2] To be more precise, a strict ordering is not required since ties are allowed.

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

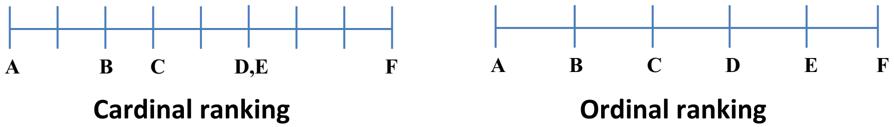


Fig. 2 Ordinal and cardinal ranking of the same information

- $w_k >_0 w_{k+1}$ with $w_k = w_{k+1}$
- $w_k >_1 w_{k+1}$ with $w_k > w_{k+1}$
- $w_k >_2 w_{k+1}$ with $w_k > x_{k_1} > w_{k+1}$ (1)
- ...
- $w_k >_i w_{k+1}$ with $w_k > x_{k_1} > \cdots > x_{k_{i-1}} > w_{k+1}$

The substitutions yield new spaces defined by the simplexes generated by the new orderings. In this way, we obtain a computationally meaningful way of representing preference strengths.

To see how the weights work, consider the cardinality expressions as distance steps on an importance scale. The number of steps corresponds straight-forwardly to the strength of the cardinalities above such that '$>_i$' means i steps. This can easily be displayed as steps on an importance ruler as suggested by Fig. 2, where the following relationships are displayed on a cardinal (left) and an ordinal (right) importance scale respectively:

- $\mathrm{w}_\mathrm{A} >_2 \mathrm{w}_\mathrm{B}$.
- $\mathrm{w}_\mathrm{B} >_1 \mathrm{w}_\mathrm{C}$.
- $\mathrm{w}_\mathrm{C} >_2 \mathrm{w}_\mathrm{D}$.
- $\mathrm{w}_\mathrm{D} >_0 \mathrm{w}_\mathrm{E}$.
- $\mathrm{w}_\mathrm{E} >_3 \mathrm{w}_\mathrm{F}$.

The decision-maker's statements are then converted into weights. One reasonable candidate for a weight function is a function that is proportional to the distances on the importance scale (Fig. 2, left). This is analogous to the equidistant criteria placed on the ordinal importance scale (Fig. 2, right). To obtain the cardinal ranking weights w_i^{CAR}, proceed as follows:

1. Assign an ordinal number to each importance scale position, starting with the most important position as number 1 (see Fig. 3).
2. Let the total number of importance scale positions be Q. Each criterion i has the position $p(i) \in \{1, \ldots, Q\}$ on this importance scale, such that for every two criteria c_i and c_j, whenever $c_i >_{s_i} c_j$, $s_i = |p(i) - p(j)|$. The position $p(i)$ then denotes the importance as stated by the decision-maker.
3. Then the cardinal ranking weights W_I^{CAR} are found by the formula[3]

[3] In Danielson et al. (2014a) and Danielson and Ekenberg (2014b), ordinal weights are introduced that are more robust than other surrogate weights, in particular. Using steps 1–3 above, cardinal weights can analogously be obtained. This is explained in detail in Danielson and Ekenberg (2015) where the performance of a set of cardinal weights are compared to ordinal weights.

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

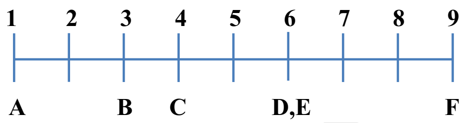


Fig. 3 Cardinal ranking with scale positions

$$w_i^{\mathrm{CAR}} = \frac{1/p(i)^{+\frac{Q+1-p(i)}{Q}}}{\sum_{j=1}^{N}\left(1/p(j)^{+\frac{Q+1-p(j)}{Q}}\right)}.$$

The CAR method follows a three-step procedure, much in analogy with the two other classes of MCDA methods. First, the values of the alternatives under each criterion are elicited in a way similar to the weights described above:

1. For each criterion in turn, rank the alternatives from the worst to the best outcome.
2. Enter the strength of the ordering. The strength indicates how strong the separation is between two ordered alternatives. Similar to weights, the strength is expressed in the notation with '$>_i$' symbols.

Second, the weights are elicited with a swing-like procedure in accordance with the discussion above.

1. For each criterion in turn, rank the importance of the criteria from the least to the most important.
2. Enter the strength of the ordering. The strength indicates how strong the separation is between two ordered criteria. The strength is expressed in the notation with '$>_i$' symbols.

Third, a weighted overall value is calculated by multiplying the centroids of the weight simplex with the centroid of the alternative value simplex. Thus, given a set of criteria in a (one-level) criteria hierarchy, $G_1, \ldots, G_n$ and a set of alternatives $a_1, \ldots, a_m$. A general value function U using additive value functions is then

$$U(a_j) = \sum_{i=1}^{n} w_i^{CAR} v_{ij}^{CAR},$$

where W_I^{CAR} is the weight representing the relative importance of attribute G_i, and $V_{IJ}^{CAR} : a_j \to [0, 1]$ is the increasing individual value function of a_j under criterion G_i obtained by the above procedure. This expression is subject to the polytopes of weights and values. This means that the feasible values are the ones in the extended polytopes defined by (1) above. Now, we define the value

$$\bar{U}(a_j) = \sum_{i=1}^{n} \bar{w}_i \bar{v}_{ij},$$

for the general value, where $\bar{w}_i$ is the centroid component of criteria weight w_i in the weight simplex and $\bar{v}_{ij}$ is the centroid component of the value of alternative a_j

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

under the criteria G_i in the simplex of values. Since we only consider non-interval valued results; the centroid is the most representative single value of a polytope. This three-step procedure contains a simple workflow that exhibits a large user acceptance, see Sect. 5.

4 Assessing the Methods

We will assess the abovementioned three classes of methods relative to our list of desired properties (qualities) at the end of Sect. 2. The first quality, efficiency, will be assessed in this section and the others in the next section. The classes will be represented by the methods SMART, AHP, and CAR respectively.

Simulation studies similar to Barron and Barrett (1996b), Ahn and Park (2008), Butler et al. (1997) and others have become a de facto standard for comparing multi-criteria weight methods. The underlying assumption of most studies is that there exist a set of 'true' weights in the decision-maker's mind which are inaccessible in its pure form by any elicitation method. We will utilise the same technique for determining the efficacy, in this sense, of the three MCDM methods suggested above. The modelling assumptions regarding decision-makers' mind-sets are mirrored in the generation of decision problem vectors by a random generator. In MCDM, different elicitation formalisms have been proposed by which a decision-maker can express preferences. Such formalisms are sometimes based on scoring points, as in point allocation (PA) or direct rating (DR) methods. In PA, the decision-maker is given a point sum, e.g., 100, to distribute among the criteria. Sometimes, it is pictured as putty with the total mass of 100 that is divided and put on the criteria. The more mass, the larger weight on a criterion, and the more important it is. In PA, there is consequently $N-1$ degrees of freedom (DoF) for N criteria. DR, on the other hand, puts no limit to the number of points to be allocated.[4] The decision-maker allocates as many points as desired to each criterion. The points are subsequently normalized by dividing by the sum of points allocated. Thus, in DR, there are N degrees of freedom for N criteria. Regardless of elicitation method, the assumption is that all elicitation is made relative to a weight distribution held by the decision-maker.[5]

The idea in both cases is to construct a set of unknowable weights that are distributed over the possible weight space. When simulating using DR the generated weights tend to cluster near the centre of the weight space. The first step in randomly generating random weights in the PA case for N attributes is to select $N-1$ random numbers from a uniform distribution on (0, 1) independently, and then rank these numbers. Assume that the ranked numbers are $1 > r_1 > r_2 \cdots > r_{n-1}$ and then let $w_1 = 1 - r_1$, $w_n = r_{n-1}$ and $w_i = r_{i+1} - r_i$ for $1 < i \leq N - 1$. These weights are uniform on the simplex (cf., e.g., Devroye 1986, Theorem 2.1, p. 207). The DR approach is then equivalent to generating N uniform [0,1] variates and setting $w_i = \frac{r_i}{\sum r_i}$. For instance, under both approaches, the expected value of w_1 is 1/3 when there are three attributes. However,

[4] Sometimes there is a limit to the individual numbers but not a limit to the sum of the numbers.

[5] For various cognitive and methodological aspects of imprecision in decision making (see, e.g., Danielson et al. 2007, 2013).

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

the resulting distributions of the weights are very different and the weights for DR are clustered in the centre of the weight space and it is much less likely that we observe a large weight on w_1.

4.1 Simulation Studies and Their Biases

In the simulations described below it is important to realize which background model we utilise. As discussed above, when following an N–1 DoF model, a vector is generated in which the components sum to 100 %. This simulation is based on a homogenous N-variate Dirichlet distribution generator. Details on this kind of simulation can be found, e.g., in Rao and Sobel (1980). On the other hand, following an N DoF model, a vector is generated without an initial joint restriction, only keeping components within [0, 100 %] yielding a process with N degrees of freedom. Subsequently, they are normalised so that their sum is 100 %. Details on this kind of simulation can be found, e.g., in Roberts and Goodwin (2002).

We will call the N–1 DoF model type of generator an *N–1-generator* and the N DoF model type an *N-generator*. Depending of the simulation model used (and consequently the background assumption of how decision-makers assess weights), the results become very different. For instance, ROC weights in N dimensions coincide with the mass point for the vectors of the N–1-generator over the polytope S_w, which is why the ROC method fares the best in simulation studies where an N–1-generator is employed (such as Barron and Barrett 1996b) and not so good in simulation studies where an N-generator is employed (such as Roberts and Goodwin 2002). In reality, we cannot know whether a specific decision-maker (or even decision-makers in general) adhere more to N–1 or N DoF representations of their knowledge. Both as individuals and as a group, they might use either or be anywhere in between. A, in a reasonable sense, *robust* rank ordering mechanism must therefore perform well under both end-points of the representation spectrum and anything in between. Thus, the evaluation of MCDM methods in this paper will use a combination of both types of generators in order to find the most efficient and robust method.

4.2 Comparing the Methods

Barron and Barrett (1996b) compared surrogate weights, where the idea was to measure the validity of the weights by simulating a large set of scenarios utilising surrogate weights and see how well different weights provided results similar to scenarios utilising true weights. The procedure is here extended with the handling of values in order to evaluate MCDM methods.

4.2.1 Generation Procedure

1. For an N-dimensional problem, generate a random weight vector with N components. This is called the TRUE weight vector. Determine the order between the weights in the vector. For each MCDM method $\mathbf{X}' \in$ {SMART,AHP,CAR}, use the order to generate a weight vector $w^{\mathbf{x}'}$.

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

2. Given M alternatives, generate $M \times N$ random values with value v_{ij} belonging to alternative j under criterion i. For each MCDM method $\mathbf{X}'$, use the order to generate a set of value vectors $v_i^{\mathbf{X}'}$.
3. Let $w_i^{\mathbf{x}}$ be the weight from the weighting function of MCDM method $\mathbf{X}$ for criterion i (where $\mathbf{X}$ is either $\mathbf{X}'$ or TRUE). For each method $\mathbf{X}$, calculate $V_j^{\mathbf{x}} = \sum_i w_i^{\mathbf{x}} v_{ij}^{\mathbf{x}}$. Each method produces a preferred alternative, i.e., the one with the highest $V_j^{\mathbf{x}}$.
4. For each method $\mathbf{X}'$, assess whether $\mathbf{X}'$ yielded the same decision (i.e., the same preferred alternative) as TRUE. If so, record a hit.

This is repeated a large number of times (simulation rounds). The hit rate (or frequency) is defined as the proportion of times an MCDM method made the same decision as TRUE.

4.3 Simulations

The simulations were carried out with a varying number of criteria and alternatives. There were four numbers of criteria $N = \{3, 6, 9, 12\}$ and four numbers of alternatives $M = \{3, 6, 9, 12\}$ in the simulation study, creating a total of 16 simulation scenarios. Each scenario was run 10 times, each time with 10,000 trials, yielding a total of 1,600,000 decision situations generated. An N-variate joint Dirichlet distribution was employed to generate the random weight vectors for the N–1 DoF simulations and a standard normalised random weight generator for the N DoF simulations. Unscaled value vectors were generated uniformly since no significant differences were observed with other value distributions. The value vectors were then used for multiplying with the obtained weights in order to form weighted values V_j^X to be compared.

The results of the simulations are shown in Table 1 below, where we show a subset of the results with a selection of pairs (N, M). The measure of success is the hit ratio as in earlier studies by others (“winner”), i.e., the number of times the highest evaluated alternative using a particular method coincides with the true highest alternative.[6] The tables below show the winner frequency utilising an equal combination of the simulation generators N–1 DoF and N DoF.

4.4 Comparing the Three MCDA Methods

Table 1 below shows the winner frequency for the three MCDA methods. SMART,[7] AHP,[8] and CAR are compared utilising an equal combination of N–1 and N DoF. The

6 A second success measure we used is the matching of the three highest ranked alternatives (“podium”), the number of times the three highest evaluated alternatives using a particular method all coincide with the true three highest alternatives. A third set generated is the matching of all ranked alternatives (“overall”), the number of times all evaluated alternatives using a particular method coincide with the true ranking of the alternatives. The two latter sets correlated strongly with the first and are not shown in this paper. Instead, we show the Kendall’s tau measure of overall performance.

7 SMART is represented by the improved SMARTER version by Edwards and Barron (1994).

8 AHP weights were derived by forming quotients w_i/w_j and rounding to the nearest odd integer. Also allowing even integers in between yielded no significantly better results.

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

Table 1 Winner frequencies in percent

N	M	SMART	AHP	CAR
3 criteria	3 alternatives	87.7	83.9	91.9
3 criteria	12 alternatives	78.2	82.5	85.8
6 criteria	6 alternatives	81.4	79.6	88.0
6 criteria	9 alternatives	79.4	80.9	86.6
9 criteria	6 alternatives	81.3	79.2	86.6
9 criteria	9 alternatives	78.9	80.2	85.1
12 criteria	3 alternatives	85.7	81.3	89.2
12 criteria	12 alternatives	77.6	81.0	82.7

Table 2 Matching of entire rankings (Kendall's *tau*)

N	M	SMART	AHP	CAR
3 criteria	3 alternatives	0.766	0.632	0.831
3 criteria	12 alternatives	0.410	0.522	0.543
6 criteria	6 alternatives	0.589	0.547	0.682
6 criteria	9 alternatives	0.474	0.505	0.585
9 criteria	6 alternatives	0.576	0.524	0.647
9 criteria	9 alternatives	0.463	0.484	0.542
12 criteria	3 alternatives	0.728	0.564	0.771
12 criteria	12 alternatives	0.376	0.428	0.437

hit ratios in the table are given in per cent and are the mean values of 10 scenario runs, i.e., 100,000 decision situations. Table 2 shows the Kendall's *tau* measure from the simulations (Winkler and Hays 1985). Kendall's *tau* is a pairwise ordering measure, measuring the number of ordered pairs of alternatives compared to the unordered ones. The *tau* lies in $[-1, 1]$ where 0 indicates no correlation between TRUE and the decision method measured and $+1$ is a perfect match.

It is clear from Table 1 that the CAR method outperforms the other methods. While CAR averages 87 %, the other two perform at around 81 %. Similarly, in Table 2 CAR displays better overall ranking compared to the other methods. The other two methods fare about equal, with SMART being somewhat stronger when fewer alternatives are involved and AHP being somewhat stronger when more alternatives are involved. This is not surprising since a very large amount of information is requested for AHP's pairwise comparisons when the number of criteria and alternatives increase. The gap up to CAR for both of the other methods is substantial considering the already high hit rate level that the methods operate at.

4.5 Noise

In the simulations above, rankings were induced from the true weights. However, the underlying assumption is that the decision-maker is able to convert beliefs into orderings almost perfectly and that the elicitation result is very accurate. The assumption

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

Table 3 The effect of noise on hit rate in percent for N = 9 criteria and M = 6 alternatives

	Noise (%)	SMART	AHP	CAR
9 criteria and 6 alternatives	0	81.3	79.2	86.6
	2	81.0	78.4	86.2
	5	79.9	75.8	84.7
	10	76.3	67.1	79.7

Table 4 The effect of noise on overall ranking (Kendall's *tau*) for N = 9 criteria and M = 6 alternatives

	Noise (%)	SMART	AHP	CAR
9 criteria and 6 alternatives	0	0.576	0.524	0.647
	2	0.557	0.519	0.637
	5	0.510	0.484	0.606
	10	0.462	0.388	0.517

of knowing the ranking with certainty is rather strong. Distortions usually affect the results, but these can to a large extent be taken into account by slightly altering the generated true weights before the order is generated. For instance, we can introduce 5 % noise by—after the generation of a true weight vector in step 1 of the generation procedure—multiplying the weights by a uniformly distributed random factor between 0.95 and 1.05 for the generation of the ranking order (not for the true test). Then the generated order simulates that the decision-maker exhibits some uncertainties regarding the true weight ordering.

Tables 3 and 4 clearly show that the behaviour of the respective methods are similar and the hit percentage naturally decreases when the amount of noise increases, especially above a couple of percent noise. The three methods are affected in much the same way and by approximately the same proportion, with AHP faring a little worse. Thus, SMART and CAR are similarly resistant to elicitation errors.

4.6 Discarding Unnatural Decision Situations

Obviously, it can be argued that the vectors generated by the simulations do not always constitute natural decision problems. For instance, the simulator could generate a weight vector with one component as high as 0.95 and the others correspondingly low. But that would probably not constitute a real-world decision problem since the decision-maker would in that case often make the decision only considering the heavily dominant criterion. Likewise, the simulator could generate a problem with a weight as low as 0.001 and such a criterion would probably not be considered at all in real life. Therefore, two filters were designed to discard weight vectors deemed unnatural. The weak filter discarded all generated true vectors with a component larger than $0.7 + 0.3/N$ or smaller than $0.05/N$. The strong filter discarded all generated true vectors with a component larger than $0.6 + 0.25/N$ or smaller than $0.1/N$. If a vector

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

Table 5 The effect of filtering on hit rate in percent for N = 9 criteria and M = 6 alternatives

	Cut-off	SMART	AHP	CAR
9 criteria and 6 alternatives	None	81.3	79.2	86.6
	Weak	81.3	79.2	87.2
	Strong	81.4	79.2	87.6

was discarded, a new vector was generated assuring that the total number of trials remained constant in each simulation.

While the exact choices of cut-off limits may seem arbitrary, the tendencies displayed are general in their nature. Table 5 shows the results from applying the cut-off filters to the selected decision simulation.

The effect of cut-off filters on the simulation results were that while SMART and AHP were to a large extent unaffected, CAR improved 1–2 % when the strong filter was applied. In particular, the ratio based AHP method seems not to improve by the filtering of generated extreme decision situations. Thus, the CAR method may be even more superior if faced only with reasonable decision situations.

5 Empirical Study

While the simulation study clearly points to CAR being theoretically preferable, a useful method must nevertheless be accepted by users in real-life decision situations. To find out how the three methods are perceived in real-life decision making, we made a study involving 100 people[9] that made one large real-life decision each. The decisions ranged from selecting country or area to live in, choosing a university program, or buying an apartment to acquiring goods like cars, motorcycles, computers, or smart phones. A requirement was that it was an important decision for that individual that he or she would be making in the near future. They were asked to consider problems with around 4 criteria and 6 alternatives. Furthermore, the report should contain only real facts and data together with the decision made. Each individual was given 2–3 weeks to complete the task and made the decision using all three methods available and was subsequently asked to reflect on their respective traits and characteristics. The methods were assisted by very similar and equally functional computer tools ensuring that all three methods were applied correctly. Adequate help with the methods was available throughout the processes.

Their reports contained decision data and results from all three methods and a comparison between the methods. In particular, the decision-makers ranked the methods on five attributes (qualities): (A) easiness of use; (B) communicating the results to others; (C) amount of time and effort required; (D) perceived correctness and transparency; and (E) willingness to use the method again. For each attribute, each decision-maker ranked the methods as 1, 2, or 3 with 1 being the foremost in each attribute, e.g., the easiest to use. The Avg. column shows the average position each method obtained for this attribute.

[9] The subjects had 2–4 years of university studies with no or little mathematical background. Thus, their level of education corresponds to an average decision making manager in many organisations.

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

Table 6 Easiness of use

A	1	2	3	Avg.
SMART	24	69	7	1.83
AHP	1	9	90	2.89
CAR	75	22	3	1.28

Table 7 Communicating the results to others

B	1	2	3	Avg.
SMART	48	35	16	1.68
AHP	4	17	78	2.75
CAR	47	47	5	1.58

Table 8 Amount of time and effort required

C	1	2	3	Avg.
SMART	31	61	7	1.76
AHP	10	8	81	2.72
CAR	58	30	11	1.53

Table 9 Perceived correctness and transparency

D	1	2	3	Avg.
SMART	26	50	23	1.97
AHP	25	13	61	2.36
CAR	48	36	15	1.67

In Table 6, the results of the attribute easiness of use can be seen. For example, 75 respondents found CAR to be the easiest to use while 90 found AHP to be the hardest to use. It is notable that only three respondents considered the CAR method to be the hardest to use.

Similarly, Table 7 shows the results for ease of communicating the results to others. In this case, CAR and SMART were almost equal, followed by AHP far behind.

In the same manner, the remaining tables show the results for the attributes amount of time and effort required to complete the decision making task (Table 8), perceived correctness of the result and transparency of the process (Table 9), and the decision-maker's willingness to use the method again (Table 10). CAR turned out to be the least time-consuming method, followed by SMART and with AHP far behind.

The perceived correctness is in conformity with the simulation results. CAR is the preferred method followed by SMART and with AHP last.

Regarding the willingness to use the method again, CAR clearly outperforms the others

For attributes B, C, and D, there were 99 valid responses and for E there were 97 out of 100 respondents. From the tables, it can be seen that CAR clearly is the preferred method while AHP is the least preferred in all five attributes. The largest difference

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

Table 10 Willingness to use the method again

E	1	2	3	Avg.
SMART	20	52	25	2.05
AHP	10	20	67	2.59
CAR	67	25	5	1.36

between CAR and the other methods was found in willingness to use the method again, while the smallest was found in communicating the results, where SMART was almost equally favoured. These results were not contradicted by the free text parts of the reports. The results of the user study in conjunction with the simulation study indicate the usefulness of the CAR method.

6 Conclusion

There is a need of methods striking a balance between formal decision analysis and reasonable cognitive demands. We have suggested a method that seems to constitute such a reasonable balance between the need for simplicity and the requirement of accuracy in MCDA and the weighting of group member opinions in group decision making. We also compared this approach (the CAR method) to methods from the popular SMART family as well as AHP. The CAR method takes ordinal knowledge into account, but recognizing that there is sometimes quite substantial information loss involved with this, we have quite conservatively extended a pure ordinal scale approach with the possibility to supply cardinal information as well. We found that the CAR method outperforms the others, both in terms of simulation results as well as in user studies, pointing to CAR as a very competitive candidate to the other hitherto more widespread methods.

Its efficiency was measured by simulation results for various numbers of alternatives and criteria, along the classical lines for assessing surrogate weights. These results show that CAR is superior regarding correctness. We also conducted a real-life user study. We studied 100 individuals previously not particularly familiar with MCDA methods, where each individual was given 2–3 weeks to complete an important decision making task. They made the decision using all three methods available and were subsequently asked to reflect on the methods' respective traits and characteristics. The study clearly showed that the CAR method generally and significantly was top-of-the-form for all the criteria above.

In conclusion, the goal was to find a more useful MCDA method with a reasonable elicitation component, which would reduce some of the applicability issues with existing more elaborate methods that we and others have developed over the years, but at the same time being able to capture more information than pure ordinal approaches. The CAR method extends rank-order weighting procedures, by taking both ordinal information as well as some cardinal relation information of the importance of the attributes into account. By this, we can sometimes avoid employing methods we and others have previously suggested for handling imprecision in decision situations, and which have turned out to be difficult to understand for normal decision-makers. The suggested method nevertheless gives significantly better simulation results than commonly used

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

competitors, such as SMART and AHP, while still seemingly being reasonably easy to understand. It was perceived not to require too much time nor be very demanding. Thus, a method utilising cardinal rankings such as CAR seems to be a serious candidate to consider. This said, it is always difficult to estimate the correctness of various methods. There is further need for empirical testing in real-life cases to determine how suitable this method is for a wider spectrum of domains and this method should be benchmarked against several others. But this article clearly demonstrates a potential advantage over some prevailing methods, but there exist a large amount of MCDA methods suggested and all of these have not been compared systematically against each other and in the future we will compare the CAR method with other approaches suggested over the years, in particular the promising dominance rules suggested in Sarabando and Dias (2009), Aguayo et al. (2014) and Mateos et al. 2014. Still, so far it seems that the CAR method has some very interesting features and provides decent decision quality.

Acknowledgments This research was funded by the Swedish Research Council FORMAS, Project Number 2011-3313-20412-31, as well as by Strategic funds from the Swedish government within ICT—The Next Generation.

References

Aguayo EA, Mateos A, Jiménez-Martín A (2014) A new dominance intensity method to deal with ordinal information about a DM's preferences within MAVT. Knowl Based Syst 69:159–169

Ahn BS, Park KS (2008) Comparing methods for multiattribute decision making with ordinal weights. Comput Oper Res 35(5):1660–1670

Arbel A, Vargas LG (1993) Preference simulation and preference programming: robustness issues in priority derivation. Eur J Oper Res 69:200–209

Bana e Costa CA, Vansnick J-C (1999) The MACBETH approach: basic ideas, software, and an application. In: Meskens N, Roubens M (eds) Advances in decision analysis, vol 4., Mathematical modelling: theory and applicationsKluwer Academic Publishers, Dordrecht, pp 131–157 7

Bana e Costa CA, Correa EC, De Corte JM, Vansnick JC (2002) Facilitating bid evaluation in public call for tenders: a socio-technical approach. Omega 30:227–242

Barron FH (1992) Selecting a best multiattribute alternative with partial information about attribute weights. Acta Psychol 80(1–3):91–103

Barron F, Barrett B (1996a) The efficacy of SMARTER: simple multi-attribute rating technique extended to ranking. Acta Psychol 93(1–3):23–36

Barron F, Barrett B (1996b) Decision quality using ranked attribute weights. Manag Sci 42(11):1515–1523

Belton V, Gear T (1983) On a short-coming of Saaty's method of analytic hierarchies. Omega 11(3):228–230

Belton V, Stewart T (2002) Multiple criteria decision analysis: an integrated approach. Kluwer Academic Publishers, Dordrecht

Bisdorff R, Dias LC, Meyer P, Mousseau V, Pirlot M (eds) (2015) Evaluation and decision models with multiple criteria: case studies. Springer, Berlin

Brans JP, Vincke PH (1985) A preference ranking organization method: the PROMETHEE method. Manag Sci 31:647–656

Butler J, Jia J, Dyer J (1997) Simulation techniques for the sensitivity analysis of multi-criteria decision models. Eur J Oper Res 103:531–546

Cook W, Kress M (1996) An extreme-point approach for obtaining weighted ratings in qualitative multicriteria decision making. Naval Res Logist 43:519–531

Danielson M, Ekenberg L (1998) A framework for analysing decisions under risk. Eur J Oper Res 104(3):474–484

Danielson M, Ekenberg L (2007) Computing upper and lower bounds in interval decision trees. Eur J Oper Res 181(2):808–816

Journal: **10726-GRUP** Article No.: **9460** TYPESET DISK LE CP Disp.:**2015/11/25** Pages: **23** Layout: **Small-X**

Danielson M, Ekenberg L, He Y (2014a) Augmenting ordinal methods for attribute weight approximations. Decis Anal 11(1):21–26

Danielson M, Ekenberg L (2014b) Rank ordering methods for multi-criteria decisions. In: Proceedings of 14th group decision and negotiation—GDN 2014, Springer

Danielson M, Ekenberg L (2015) Using surrogate weights for handling preference strength in multi-criteria decisions. In: Proceedings of 15th group decision and negotiation—GDN 2015, Springer

Danielson M, Ekenberg L, Johansson J, Larsson A (2003) The DecideIT decision tool. In: Bernard J-M, Seidenfeld T, Zaffalon M (eds) Proceedings of ISIPTA'03, pp 204–217, Carleton Scientific

Danielson M, Ekenberg L, Larsson A (2007) Distribution of belief in decision trees. Int J Approx Reason
[8] 46(2):387–407

Danielson M, Ekenberg L, Larsson A, Riabacke M (2013) Weighting under ambiguous preferences and
[9] imprecise differences in a cardinal rank ordering process. Int J Comput Intell Syst

Danielson M, Ekenberg L, Riabacke A (2009) A prescriptive approach to elicitation of decision data. J Stat Theory Pract 3(1):157–168

Danielson M, Ekenberg L, Larsson A, Riabacke M (2014) Weighting under ambiguous preferences and imprecise differences in a cardinal rank ordering process. Int J Comput Intell Syst 7(1):105–112

Devroye L (1986) Non-uniform random variate generation. Springer, Berlin

Dias LC, Clímaco JN (2000) Additive aggregation with variable interdependent parameters: the VIP analysis software. J Oper Res Soc 51(9):1070–1082

Ekenberg L, Boman M, Danielson M (1995) A tool for coordinating autonomous agents with conflicting goals. In: Proceedings of the 1st international conference on multi-agent systems ICMAS '95, pp 89–93, AAAI/MIT Press

Ekenberg L, Boman M, Linneroth-Bayer J (2001a) General risk constraints. J Risk Res 4(1):31–47

Ekenberg L, Danielson M, Larsson A, Sundgren D (2014) Second-order risk constraints in decision analysis. Axioms 3:31–45. doi:10.3390/axioms3010031

Ekenberg L, Thorbiörnson J (2001) Second-order decision analysis. Int J Uncertain Fuzziness Knowl Based Syst 9(1):13–38

Ekenberg L, Thorbiörnson J, Baidya T (2005) Value differences using second order distributions. Int J Approx Reason 38(1):81–97

Edwards W (1971) Social utilities. In: Engineering economist, summer symposium series, vol 6,
[10] pp 119–129

Edwards W (1977) How to use multiattribute utility measurement for social decisionmaking. IEEE Trans Syst Man Cybern 7(5):326–340

Edwards W, Barron F (1994) SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. Organ Behav Hum Decis Process 60:306–325

Figueira J, Roy B (2002) Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. Eur J Oper Res 139:317–326

Ginevicius R (2009) A new determining method for the criteria weights in multicriteria evaluation. Int J Inf Technol Decis Making 10(6):1067–1095

Jia J, Fischer GW, Dyer J (1998) Attribute weighting methods and decision quality in the presence of response error: a simulation study. J Behav Decis Making 11(2):85–105

Jiménez A, Ríos-Insua S, Mateos A (2006) A generic multi-attribute analysis system. Comput Oper Res 33:1081–1101

Katsikopoulos K, Fasolo B (2006) New tools for decision analysis. IEEE Trans Syst Man Cybern A Syst Hum 36(5):960–967

Keeney R, Raiffa H (1976) Decisions with multiple objectives: preferences and value tradeoffs. Wiley, New York

Kirkwood CW (1997) strategic decision making: multiobjective decision making with spreadsheets. Wadsworth Publishing, Belmont

Krovak J (1987) Ranking alternatives-comparison of different methods based on binary comparison matrices. Eur J Oper Res 32:86–95

Larsson A, Johansson J, Ekenberg L, Danielson M (2005) Decision analysis with multiple objectives in a framework for evaluating imprecision. Int J Uncertain Fuzziness Knowl Based Syst 13(5):495–509

Larsson A, Riabacke M, Danielson M, Ekenberg L (2014) Cardinal and rank ordering of criteria—addressing prescription within weight elicitation. Int J Inf Technol Decis Making 13

Mateos A, Jiménez-Martín A, Aguayo EA, Sabio P (2014) Dominance intensity measuring methods in MCDM with ordinal relations regarding weights. Knowl Based Syst 70:26–32

Journal: 10726-GRUP Article No.: 9460 TYPESET DISK LE CP Disp.:2015/11/25 Pages: 23 Layout: Small-X

Milnor, J (1954) Games against nature. In: Decision processes. Wiley
Mustajoki J, Hämäläinen R (2005) A preference programming approach to make the even swaps method even easier. Decis Anal 2:110–123
Mustajoki J, Hämäläinen R, Salo A (2005) Decision support by interval SMART/SWING—incorporating imprecision in the SMART and SWING methods. Decis Sci 36(2):317–339
Park KS (2004) Mathematical programming models for characterizing dominance and potential optimality when multicriteria alternative values and weights are simultaneously incomplete. IEEE Trans Syst Man Cybern A Syst Hum 34(5):601–614
Puerto J, Mármol AM, Monroy L, Fernández FR (2000) Decision criteria with partial information. Int Trans Oper Res 7:51–65
Rao JS, Sobel M (1980) Incomplete Dirichlet integrals with applications to ordered uniform spacing. J Multivar Anal 10:603–610
Riabacke M, Danielson M, Ekenberg L (2012) State-of-the-art in prescriptive weight elicitation. Adv Decis Sci. doi:10.1155/2012/276584
Roberts R, Goodwin P (2002) Weight approximations in multi-attribute decision models. J Multi Criteria Decis Anal 11:291–303
Roy B (1968) Classement et choix en présence de points de vue multiples (la méthode ELECTRE). La Revue d'Informatique et de Recherche Opérationelle 8:57–75
Saaty TL (1977) A scaling method for priorities in hierarchical structures. J Math Psychol 15:234–281
Saaty TL (1980) The analytic hierarchy process. McGraw-Hill, New York
Salo AA, Hämäläinen RP (2001) Preference ratios in multiattribute evaluation (PRIME)–elicitation and decision procedures under incomplete information. IEEE Trans Syst Man Cybern A Syst Hum 31:533–545
Sarabando P, Dias L (2009) Multi-attribute choice with ordinal information: a comparison of different decision rules. IEEE Trans Syst Man Cybern A 39:545–554
Sarabando P, Dias L (2010) Simple procedures of choice in multicriteria problems without precise information about the alternatives' values. Comput Oper Res 37:2239–2247
Steuer RE (1984) Sausage blending using multiple objective linear programming. Manag Sci 30(11):1376–1384
Stewart TJ (1993) Use of piecewise linear value functions in interactive multicriteria decision support: a Monte Carlo study. Manag Sci 39(11):1369–1381
Stillwell W, Seaver D, Edwards W (1981) A comparison of weight approximation techniques in multiattribute utility decision making. Organ Behav Hum Perform 28(1):62–77
Winkler RL, Hays WL (1985) Statistics: probability, inference and decision, Holt. Rinehart & Winston, New York
von Winterfeldt D, Edwards W (1986) Decision analysis and behavioural research. Cambridge University Press, Cambridge