

Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Multi-criteria decision analysis: Limitations, pitfalls, and practical difficulties

Permalink

<https://escholarship.org/uc/item/0cp6j7sj>

Author

Kujawski, Edouard

Publication Date

2003-02-01

Multi-Criteria Decision Analysis: Limitations, Pitfalls, and Practical Difficulties

Edouard Kujawski
Engineering Division
Lawrence Berkeley National Laboratory
Berkeley, California 94720

Abstract. The 2002 Winter Olympics women's figure skating competition is used as a case study to illustrate some of the limitations, pitfalls, and practical difficulties of Multi-Criteria Decision Analysis (MCDA). The paper compares several widely used models for synthesizing the multiple attributes into a single aggregate value. The various MCDA models can provide conflicting rankings of the alternatives for a common set of information even under states of certainty. Analysts involved in MCDA need to deal with the following challenging tasks: (1) selecting an appropriate analysis method, and (2) properly interpreting the results. An additional trap is the availability of software tools that implement specific MCDA models that can beguile the user with quantitative scores. These conclusions are independent of the decision domain and they should help foster better MCDA practices in many fields including systems engineering trade studies.

INTRODUCTION

Sound decision analysis is of critical importance in engineering and other human activities. Projects that follow a good systems engineering process typically generate or consider several alternatives at all levels of the design, ranging from the architecture down to the components. Criteria that depend on human preferences, such as ease of use, can only be subjectively evaluated; but they cannot be avoided when dealing with real-world situations. Therefore in many trade studies, engineers and analysts are faced with choosing or recommending the "best alternative" from a set of alternatives characterized by criteria that may be qualitative, quantified with different units of measure, and conflict with each other (Buede 2000). Given their tradition and fondness for quantitative methods, they naturally turn to quantitative decision analysis methods. But all quantitative methods make strong assumptions, and they should be scrutinized especially if they are used as decision-making tools in accordance with the normative point of view rather than as decision aids (Bell et al 1988).

In this paper we examine the problem of choosing the "best" alternative from a finite set of alternatives that are characterized by deterministic attributes. Formal paradigms for decision-making under certainty have existed in various guises for centuries (Hammond et al 1999). The basic steps are:

1. Define the objective(s)
2. Determine the attributes (criteria or performance measures) that relate to each objective
3. Structure the attributes into a hierarchy
4. Develop candidate alternatives or options
5. Determine the importance of the criteria and assign corresponding weights to them
6. Evaluate the alternatives with respect to the criteria
7. Combine or synthesize the multiple single-attribute values into a single aggregate multi-

attribute score

8. Step-back and evaluate the results.

But decision analysis is not simply the cut and dry step-by-step process given above. There are several long-standing and unresolved issues:

- The importance of structuring the attributes into a hierarchy
- The meaning of the weights and scores of the individual criteria
- The validity of the aggregate multi-attribute value function
- The use of a single score to adequately characterize a complex situation.

It is rare that a single alternative is as good or better than all other alternatives for all attributes. In general, it is therefore not possible to directly compare the feasible alternatives with each other as in the case of a single or scalar function. A key characteristic of most MCDA problems is that they generally do not have conclusive or unique solutions. A central idea of most MCDA approaches is that one can combine all of the criteria into a single scalar objective function and the "best" solution is the alternative with the highest score. The complex multi-dimensional decision problem is thereby reduced to a single number. This may look reasonable, but there are serious shortcomings (Kujawski 2002).

Numerous MCDA models and/or methods have been proposed for dealing with complex decision-making (Mollaghasemi and Pet-Edwards, 1997). Unfortunately, they often reflect different and conflicting perspectives. Objective guidelines for choosing an appropriate MCDA method are rare (Guitouni and Martel 1998). Although there are claims that MCDA has been successfully applied to thousands of real-world problems, very little serious ex-post analysis of these applications has been made (Corner and Kirkwood 1991). MCDA is a very active field of research and applications with enough publications to warrant a dedicated journal, *Journal of Multi-Criteria Decision Analysis*.

The choice of a MCDA model or method can play a major role in the decision and at times may lead to conflicting results. The ranking of the alternatives depends on evaluation of the attributes for each alternative, the weights of the attributes, and the synthesis or aggregate function. It is then not surprising that there have been suits with the validity of the decision-making analysis subjected to legal review (Pariseau and Oswalt 1994). In this paper we explicitly illustrate some of the limitations, pitfalls, and practical difficulties of MCDA under certainty. Uncertainty and probabilistic outcomes introduce additional complexities and require techniques not addressed in this paper (Keeney and Raiffa 1976).

A BRIEF LOOK AT A FEW MCDA MODELS

In this paper we assume that Steps 1 through 6 of the process outlined above have been carried out. Principally because of the time limitation of the presentation we focus on Step 7 and the following three popular synthesis or aggregate models: the Weighed Sum Model (WSM), the Weighed Product Model (WPM) (Daniels et al. 2001), and the Analytic Hierarchy Process (AHP) (Saaty 2000). It is important to note that all three models assume that the preference for any one criterion does not depend on the values of the other criteria. This property, known by the technical name of mutual preferential independence, depends on the specific problem and how the analysis is structured. Analysts should always assess whether independence is a reasonable assumption, and if not they need to consider more complex MCDA models that incorporate interactions (Clemen and Reilly 2001).

The Weighed Sum and Weighed Product Models. Consider n alternatives $\{A_1, \dots, A_n\}$ with m

deterministic criteria $\{C_1, \dots, C_m\}$. The alternatives are fully characterized by the decision matrix $\{S_{ij}\}$, where S_{ij} is the score that measures how well alternative A_i does on criterion C_j . The weights $\{w_1, w_2, \dots, w_m\}$ account for the relative importance of the criteria. Given our focus on synthesis models, we assume that the scores and weights have been obtained using systematic rather than ad-hoc techniques. We simply note that (1) these activities are very important and non-trivial, and (2) the weights represent trade-offs between the various criteria and therefore their values depend on the underlying scales (Bouyssou et al 2000).

In the WSM the score of alternative A_i is given by

$$S(A_i) = \sum w_j S_{ij},$$

where the sum is over $j = 1, 2, \dots, m$. For consistency, all criteria must be expressed in the same units or utiles. The WPM exhibits compensatory behavior and therefore should not be used with “musts” (Kepner and Tregoe 1965). The Grade Point Average (GPA) used in most US schools is an example of a WSM.

In the WPM the score of alternative A_i is given by

$$S(A_i) = \prod S_{ij}^{w_j},$$

where the product is over $j = 1, 2, \dots, m$. The WPM eliminates alternatives with poor attributes. For example, in an elementary school that uses a WPM for grading students, a student who fails a required subject gets to repeat the whole grade.

It is readily seen that the WSM and WPM capture different behaviors and may lead to different solutions. More generally, many different aggregate functions are possible and no model can be considered the “super model” appropriate to all decision-making situations.

The Analytic Hierarchy Process. Saaty developed the AHP in the early 1980's. Much of its popularity stems from: (1) the appearance of a scientific approach because of the use of a matrix formulation; (2) the appealing elicitation technique for the weights and scores based on pairwise comparison using a verbal scale; and (3) a readily available software tool, Expert ChoiceTM. But it has been observed that ranking irregularities can occur in the original AHP, and several variants have been formulated (Triantaphyllou 2000). Expert ChoiceTM now provides both “distributive synthesis” (original AHP) and “ideal synthesis” modes of the additive AHP. In this paper we present a new real-life example that illustrates the following potential AHP behaviors: (1) the ideal and distributive synthesis modes do not always yield the same ranking; and (2) in the distributive mode, adding an “inferior” alternative can lead to rank reversal.

Other Approaches. The above three models can be viewed as practical implementations of multi-attribute utility theory under certainty. But there are approaches, too numerous to be listed in this paper, that use rather different paradigms. Based on the author's interest and with apology to the practitioners of omitted methods, we briefly acknowledge the outranking method developed by Roy in the early 1970's (Roy 1999). The outranking method compares alternatives on the basis of two measures: concordance and discordance. Concordance reflects those cases where one pair of alternative is superior to the other. Discordance reflects the reverse case where one pair of alternative is inferior to the other. The fact that this approach is primarily used in France is an important lesson for MCDA. MCDA, by its very nature, has a strong subjective component and it reflects human nature and cultural differences. It is therefore not surprising that multiple analysts and/or decision-makers often reach different rankings. Although these psychological and cultural aspects are of great interest and are significantly important, they lie outside the scope of the present paper.

THE 2002 WINTER OLYMPICS FIGURE SKATING

In this paper we opt to analyze a real situation rather than use the ubiquitous "choosing a car" example. The 2002 winter Olympics woman's figure skating provides an interesting example of how rank reversal can occur in a scoring system. To quote (Cazeneuve 2002):

“If you don’t think skating scoring was screwy before Thursday night, you may be asking: How is it that Michelle Kwan was ahead of Sarah Hughes before Irina Slutskaya skated, only to finish behind Hughes after Slutskaya performed?”

The above problem is not limited to sports and/or political elections. It has implications for trade studies and deserves the serious consideration of systems engineers and other analysts.

The International Skating Union (ISU) Rules.

- Each judge gives marks for both the short and long programs based on technical merit and presentation. A score of 6.0 is perfect.
- The marks are used solely to decide who comes in 1st, 2nd,... in each of the two programs.
- The specific placements are assigned a point value: 1st = 1, 2nd = 2,...
- In the short program, the technical merit mark serves as the tie-breaker.
- In the long program, the presentation mark serves as the tie-breaker.
- The short-program placements are multiplied by 0.5.
- The long-program placements are multiplied by 1.0.
- The lowest total of the two “factored placements” determines the overall winner. In the case of a tie the skater with the lower long-program “factored placement” is the winner.

The ISU scoring system may be viewed as another creative MCDA model!

The Short Program. The standings and associated points in the short program were as follows:

	Ranking by judge									Total	Standing	Points
	Ger.	Rus.	Slov.	Den.	Ita.	Bel.	Fin.	Can.	USA			
<i>Others</i>												
Sarah H.	6	9	4	5	5	5	5	4	4	47	4	2.0
Michelle K.	1	2	1	1	1	2	2	2	1	13	1	0.5
Irina S.	3	1	2	2	3	1	1	1	3	17	2	1.0

Only the awarded points are of further consequence in determining the medallists.

The Long Program. To understand what took place we examine the details of the ISU scoring system and the standings before and after Irina S. skated. For the benefit of the readers interested in working through the details, we provide each judge’s combined scores in the long program:

	Combined scores by judge									
	Ger.	Rus.	Slov.	Den.	Ita.	Bel.	Fin.	Can.	USA	
<i>Others</i>										
Sarah H.	11.4	11.5	11.6	11.4	11.6	11.6	11.5	11.6	11.6	
Michelle K.	11.3	11.5	11.7	11.5	11.4	11.5	11.4	11.5	11.4	
Irina S.	11.3	11.7	11.8	11.6	11.4	11.7	11.5	11.4	11.5	

Standings before Irina S. skated. After Sarah H. and Michelle K. had skated the standings were as follows:

	Points			Standing
	Short Pgm	Long Pgm	Total	
Sarah H.	2.0	1.0	3.0	2
Michelle K.	0.5	2.0	2.5	1
Irina S.	1.0	?	?	

At this point in the competition, Michelle K. was ahead of Sarah H. But, Irina S. still had to skate.

Final results. Irina S. skates and she receives the marks (see the long program section) that result in her ranking second in the long program. The final results and standings turned out as follows:

	Points			Standing
	Short Pgm	Long Pgm	Total	
Sarah H.	2.0	1.0	3.0	1 [^]
Michelle K.	0.5	3.0	3.5	3
Irina S.	1.0	2.0	3.0	2

[^] Long program serves as tie-breaker

We have here a real-world case of how "rank reversal" can occur in a ranking system. Michelle K. was ahead of Sarah H. before Irina S. skated. After Irina S. skated, Sarah H. had won the gold medal. For most of us the ISU scoring system got it right! But we now examine this competition with the MCDA methods presented above. As already indicated, they are the methods typically used by systems engineers in the USA.

The 2002 Winter Olympics Figure Skating revisited

Now we consider scoring the 2002 winter Olympics women figure skating using the WSM, the WPM, and the AHP.

The WSM and WPM. Given that each judge grades on an interval scale of 0 to 6.0, we proceed to use this data to determine each skater's total score within the WSM and WPM. For each skater, we then compute the Arithmetic Mean (AM) and Geometric Mean (GM) of the judges' scores. To get additional insight into these scores, we compute and report the Standard Deviation (SD). The results are as follows:

	WSM				WPM			
	AM	GM	Rank	SD	AM	GM	Rank	SD
Sarah H.	17.03	17.03	3	0.20	182.69	182.59	3	6.49
Michelle K.	17.27	17.27	2	0.15	190.61	190.55	2	4.99
Irina S.	17.31	17.30	1	0.22	192.00	191.88	1	7.39

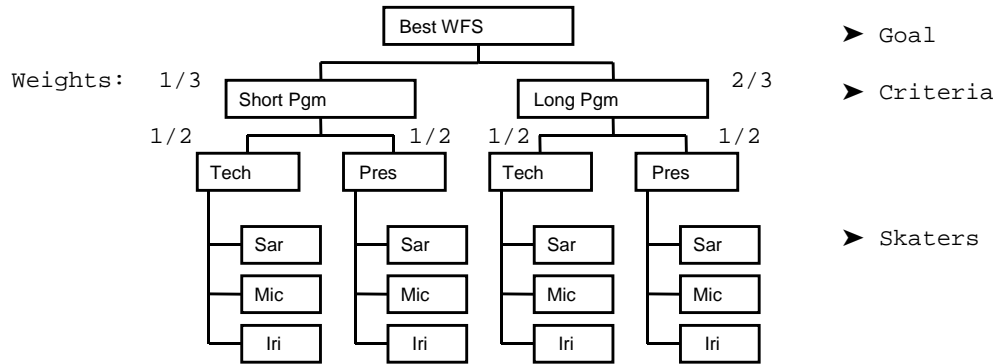
AM = arithmetic mean; GM = geometric mean; SD = standard deviation

The WSM and WPM both rank Irina S. in 1st place followed by Michelle K. and Sarah H. But statistically speaking, the scores are too close to distinguish among the three skaters. I let the reader judge whether the WSM and WPM rather than the ISU scoring system got it right! It is interesting to note that Hazelrigg (Hazelrigg 1996) has argued that Arrow's Impossibility Theorem implies that design methods based on MCDA techniques may lead to erroneous results

when dealing with a group of decision makers or multiple stakeholders.

The AHP. We consider a single hypothetical judge and the competition with and without Irina S.

Case 1: Three Skaters. We label the three skaters Sar, Mic, and Iri. We structure the problem as follows:



We assume the following perfectly consistent pairwise AHP comparison matrices for the short and long programs:

	Mic	Sar	Iri
Mic	1.00	6.30	0.77
Sar	1/6.30	1.00	0.13
Iri	1/0.77	1/0.13	1.00

	Mic	Sar	Iri
Mic	1.00	0.67	5.30
Sar	1/0.67	1.00	7.70
Iri	1/5.30	1/7.70	1.00

Expert Choice™ gives the following priority vectors and rankings in the distributive and ideal synthesis modes:

Skater	Distributive Mode		Ideal Mode	
	Priority Vector	Ranking	Priority Vector	Ranking
Sar	0.392	1	0.386	2
Mic	0.386	2	0.386	1
Iri	0.222	3	0.228	3

Thus Expert Choice™'s distributive mode selects Sar the winner while Expert Choice™'s ideal mode selects Mic the winner. The rankings of Sar and Mic in the ideal mode reflect differences in the fourth decimal places of the corresponding elements in the priority vector. The results are arithmetically valid; but they are definitely beyond the validity of the model and analysis. But for the purpose of this paper, this illustrates how the choice of an AHP synthesis model, distributive mode or ideal mode, can and does influence the AHP solution.

Case 2: Excluding Iri. Excluding Iri has no impact on the pairwise comparison between Mic and Sar. The pairwise AHP comparison matrices for the short and long programs are 2x2 subsets of the ones presented above. Expert Choice™ now ranks Mic ahead of Sara in both the distributive and ideal additive synthesis modes. This illustrates how dropping or adding an

“inferior” alternative can lead to rank reversal within the original (additive and distributive synthesis) AHP.

Concluding Remarks

We have illustrated some of the limitations, pitfalls, and practical difficulties of applying MCDA to decision-making under certainty using a real, but somewhat unconventional example, the 2002 winter Olympics woman's figure skating. But the MCDA methods considered above are independent of the decision domain and the lessons-learned are of importance in many fields including systems engineering trade studies. For example, savvy consumers seem to have an excellent appreciation of the usefulness and limitations of quantitative decision analysis techniques. They do not base their choice simply on a single criterion. They use publications such as *Consumer Reports* to help them assess products ranging from a toothbrush to a car and then choose their preferred option. In fact the preferred options are often not the highest ranked ones. It is the essence of systems thinking that the methods developed in one area often prove equally valuable in another.

There are many MCDA methods and models, but none can be considered the “best” and/or appropriate for all situations. MCDA possesses a strong craft element in addition to its technical aspects. Systems engineers and other analysts need to be cognizant of the assumptions and limitations of the different MCDA methods to use them effectively. This is necessary for the analysis to be capable of withstanding scrutiny and not to be dismissed as “juggling with meaningless and magic numbers”. Standard sensitivity analysis helps to examine the extent to which the weights and scores influence the solution and the robustness of the results within a given MCDA model. It therefore provides an indication of the limitations, pitfalls, and practical difficulties of an individual MCDA model. In addition, as illustrated in this note, different MCDA models can and do significantly influence the solution. We do not propose nor advocate that trade-offs be based on a single criterion such as cost. When properly applied, MCDA can help individuals and organizations make better decisions. But, the final decision should not be the automatic result of a MCDA; it should be made by the decision-maker(s).

REFERENCES

- Bell, D.E., Raiffa, H., and Tversky, A., *Decision Making: Descriptive, Normative, and Prescriptive Interactions*. Cambridge University Press, Cambridge, 1988.
- Bouyssou, D., Marchant, T., Pirlot, M., Perny, P., Tsoukias, A., and Vincke, P., *Evaluation and Decision Models: A Critical Perspective*. Kluwer Academic Publishers, Boston, 2000.
- Buede, D.M., *The Engineering Design of Systems*. John Wiley & Sons, New York, 2000.
- Cazeneuve, B., “Sneaking away with gold How Hughes impossibly vaulted ahead of Kwan, Slutskaya,”
http://sportsillustrated.cnn.com/olympics/2002/figure_skating/news/2002/02/21/cazeneuve_explainer/.
- Clemen, R.T. and Reilly, T., *Making Hard Decisions with DecisionTools*. Duxbury, Pacific Grove, 2001.
- Corner, J.L. and Kirkwood, C.W., “Decision Analysis Applications in the Operations Research Literature, 1970-1989,” *Operations Research*, Vol. 39, 1991, pp 206-219.
- Daniels, J., Werner, P., and Bahill, T., “Quantitative Methods for Tradeoff Analyses,” *Systems Engineering*, Vol. 4, 2001, pp 190-211.
- Guitouni, A. and Martel J.M., “Tentative Guidelines to Help Choosing an Appropriate MCDA

- Method, *European Journal of Operational Research*, Vol. 109, 1998, pp 501-521.
- Hammond, J.S., Kenney, R.L., and Raiffa H., *Smart Choices A Practical Guide to Making Better Decisions*. Harvard Business School Press, Boston, 1999.
- Hazelrigg, G.A., "The Implications of Arrow's Impossibility Theorem on Approaches to Optimal Engineering Design," *Journal of Mechanical Design*, Vol. 118, 1996, pp 161-164.
- Keeney, R.L. and Raiffa, H., *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley & Sons, New York, 1976.
- Kepner, C.H. and Tregoe, B.B., *The Rational Manager A Systematic Approach to Problem Solving and Decision Making*. Kepner-Tregoe, Princeton, 1965.
- Kujawski, E., "Selection of Technical Risk Responses for Efficient Contingencies," *Systems Engineering*, Vol. 5, No. 3, 2002, pp 194-212.
- Mollaghasemi, M. and Pet-Edwards, J., *Making Multiple-Objective Decisions*. IEEE Computer Society Press, Los Alamitos, 1997.
- Pariseau, R. and Oswalt, I., "Using Data Types and Scales for Analysis and Decision Making," *Acquisition Review Quarterly*, Vol. 1, 1994, pp.145-159.
- Roy, B., "Decision-Aiding Today: What Should We Expect?" in *Multicriteria Decision Making: Advanced in MCDM Models, Algorithms, Theory, and Applications*. Edited by T. Gal, T.J. Stewart, and T. Hanne. Kluwer Academic Publishers, Boston 1999, pp. 1.1 - 1.35.
- Saaty, T.L., *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process - Vol. VI of the AHP Series*. RWS Publications, Pittsburgh, 2000.
- Triantaphyllou, E., *Multi-Criteria Decision Making Methods: A Comparative Study*. Kluwer, Dordrecht, 2000.