On Attribute Weighting in Value Trees

Mari Pöyhönen

Helsinki University of Technology
Systems Analysis Laboratory
P.O. Box 1100
FIN-02015 HUT, Finland
http://www.hut.fi/Units/Systems.Analysis

Dissertation for the degree of Doctor of Technology to be presented with due permission for public examination and debate in Auditorium E at Helsinki University of Technology (Espoo, Finland) on the 17th of April, 1998, at 12 o'clock noon.

ISBN 951-22-4003-3
ISSN 0782-2030
Libella
Otaniemi 1998

Helsinki University of Technology
Systems Analysis Laboratory
Research Reports
A73, March 1998

Preface

The work for this thesis has been carried out at Systems Analysis Laboratory at Helsinki University of Technology. I want to express my warmest thanks to my thesis supervisor, professor Raimo P. Hämäläinen, for his dynamic way to supervise me. I learned a lot during these years. It was a great pleasure to collaborate with Dr. Ahti Salo and Dr. Hans C. Vrolijk. Ahti provided me a good start in this work by telling me what I should read and learn. With Hans it was fun to study mysteries of human behavior.

I am greatly indebted to preliminary examinators, professor Jyrki Wallenius and professor Valerie Belton, for their efforts to go through my thesis and for their invaluable comments. Jyrki together with professor Pekka Korhonen have kindly encouraged me to follow the research activities at Helsinki School of Economics. Our inter-university graduate school led by Raimo, Pekka, and Jyrki has provided me countless opportunities for interesting discussions with people such as Dr. Johanna Bragge and Ms. Tarja Joro.

Systems Analysis Laboratory is a great place to study and work. All the courses there with their excellent teachers, such as Dr. Harri Ehtamo, Dr. Pertti Laininen, and Dr. Jukka Ruusunen, gave me a set of tools and a way to think that will follow me the rest of my life. All the people who worked there during these years created an atmosphere where it was fun to work and air was full of brilliant thoughts. As well as inspired by the colleagues, I was inspired by students who I had a privilege to teach, discuss with, and who helped me in my work. I feel that often I learned more from their questions compared to what they learned from my answers.

The list of people who have contributed to this thesis would be too long to be presented here. I want to express my deepest gratitude to all friends and family who have supported me during these years. It is a privilege to have friends like Eetu, Minna, Pauli, and Tuula. Without them I would have completed the thesis but I would have been very unhappy person. My mother has taught me to appreciate knowledge in all forms and I greatly appreciate this lesson. Ritva and Heikki have kindly supported me through my study years in all possible ways and they gave me their most precious property, their son.

Kari has been the sun, the moon, stars, wind, and rain of my life.

Delft, March 18, 1998

Mari Pöyhönen

Academic dissertation

Systems Analysis Laboratory
Helsinki University of Technology

On attribute weighting in value trees

Author:	Mari Pöyhönen
Supervisor:	Professor Raimo P. Hämäläinen
Preliminary examiners:	Professor Valerie Belton, University of Strathclyde, United Kingdom Professor Jyrki Wallenius, Helsinki School of Economics
Official opponent:	Professor Jim Corner, University of Waikato, New Zealand

Publications

The dissertation consists of the present summary and the following papers:

[I] Pöyhönen, M., Hämäläinen, R.P., and Salo, A. (1997), "An experiment on the numerical modeling of verbal ratio statements," *Journal of Multi-Criteria Decision Analysis*, 6, 1-10.

[II] Pöyhönen, M. and Hämäläinen, R.P. (1997), *On the convergence of multiattribute weighting methods*, Helsinki University of Technology, Systems Analysis Laboratory, Research Report A66.

[III] Hämäläinen, R.P. and Pöyhönen, M. (1996), "On-line group decision support by preference Programming in Traffic Planning," *Group Decision and Negotiation*, 5, 485 - 500.

[IV] Pöyhönen, M. and Hämäläinen, R.P. (1998), "Notes on the weighting biases in value trees," forthcoming in *Journal of Behavioral Decision Making*

[V] Pöyhönen, M., Vrolijk, H.C., and Hämäläinen, R.P. (1997), *Behavioral and procedural consequences of structural variation in value trees*, Helsinki University of Technology, Systems Analysis Laboratory, Research Report A69.

[VI] Pöyhönen, M. and Hämäläinen, R.P. (1998), *There is hope in attribute weighting*, Helsinki University of Technology, Systems Analysis Laboratory, Research Report A74.

The research problem settings in publications [I], [II], and [III] are based on the ideas of professor Raimo P. Hämäläinen and Dr. Ahti Salo. In these three publications the experimental design has been done in close co-operation with the co-authors. The author has had the main responsibility over the implementation of the experiments and analysis in publications [I], [II], and [III]. The author has written these three articles in close co-operation with professor Hämäläinen.

The initial research hypotheses and the main contributing ideas in publications [IV], [V], and [VI] are author's own work. The author has done the experimental design for experiments in publications [V] and [VI]. The experimental data was collected by the author for publication [VI] and by Dr. Vrolijk for publication [V]. The statistical analysis was done by the author in publications [IV] and [VI] and in co-operation with Dr. Vrolijk in publication [V]. The author has written publications [IV] and [VI] in close co-operation with professor Hämäläinen and publication [V] in co-operation with Dr. Vrolijk and professor Hämäläinen.

Abstract

The thesis focuses on the biases appearing when decision makers are asked to give numerical preference statements on attribute weights in multiattribute value tree analysis (MAVT). The thesis focuses on two problems. First, different attribute weight elicitation methods yield different weights although they are based on the same theoretical principles. Second, the attribute weights change when the structure of a value tree is varied. We run two experiments to study the differences between five different weight elicitation methods (Analytic Hierarchy Process, direct weighting, Simple Multi Attribute Rating Technique, SWING weighting, and tradeoff weighting) and to study how verbal expressions are used in preference elicitation. The weighting methods do yield different weights. These differences originate from the way decision makers restrict their responses depending on the numbers that the methods explicitly or implicitly propose. With the biases related to the structural variation of value trees we first point out that earlier experiments are insufficient because they have drawn conclusions of individual behavior based on averages over large group of subjects. In an experiment we then show that the weights change when the structure of a value tree changes because the subjects again give numbers to describe their preferences so that they clearly favor some response scales. We propose, based on the results from our experiments and earlier observations, that two same origins for many different problems in attribute weighting are that the decision makers give numbers that reflect ordinal information on preferences only and that the weights are normalized to sum up to one. The decision makers' interpretation of the numbers that they use differs from the assumptions of value theory. One way to avoid these problems is to develop more interactive weighting procedures. Different weighting methods should be combined and decision makers should interactively evaluate the results during the weighting. The role of the interactiveness was tested in two experiments. The other focused on the so called preference programming techniques that were tested in a group decision making situation. The other experiment showed that the increased awareness of the methodology can decrease the biases.

1 Introduction

In making decisions involving multiple objectives we have to consider non-commensurable and conflicting goals and weigh them against each other. To support our decision making and to justify our choices we make statements on the importance of these goals. For example, in selecting a new job I could state that "my most important goal is to maximize my salary." Yet, I would not select a job where salary is only 10 FIM higher compared to other jobs if that job is otherwise worse than the other ones. It is not clear how the statements on the importance of goals should be interpreted. Multiattribute value tree analysis (MAVT) attempts to give these statements a specific meaning and uses this type of preference statements to support the decision making process.

In MAVT decision makers are asked to give numerical preference statements which are used to calculate the attribute weights describing the trade off between objectives (Keeney and Raiffa, 1976; Dyer and Sarin, 1979). Theoretically sound attribute weighting is, however, subject to many biases (see e.g. Hogarth, 1980; Weber and Borcherding, 1993; Borcherding et al., 1994). An example is the splitting bias referring to a phenomenon that an attribute receives more weight if it is split into subattributes (Weber et al., 1988; Borcherding and von Winterfeldt, 1988). Nevertheless, it has been unclear where many of these phenomena originate from. They may originate from the behavior of the decision makers or from the procedures used in the weighting. Furthermore, there does not exist guidelines for practitioners how to avoid these traps. This thesis deals with these issues and presents new experimental results related to the weighting biases in the MAVT framework.

The motivation to do experimental research related to attribute weighting biases is the increased use of the decision analysis methods in aiding public decision making (for a review of applications see Corner and Kirkwood, 1992). Value tree analysis has been found to be a useful tool especially in energy policy and environmental decision making problems (Hämäläinen, 1988; 1990; Keeney et al., 1986; Keeney and McDaniels,1992; Marttunen and Hämäläinen, 1995; Miettinen and Hämäläinen, 1997). Attribute weights play a key role in identifying the opinions of stakeholders. Thus it should be made sure that stakeholders interpret attribute weights in the same way and that other factors, like the choice of a particular weight elicitation method, do not affect the results. The experiments run in the controlled laboratory conditions are needed because in that way the different factors affecting the weights can be separated and studied. One should note, however, that laboratory experiments are not the same as real applications (see also Hobbs, 1986; Fraser et al., 1992). The student subjects do not have commitment towards the final results and they are not involved in the problem structuring. Although this thesis presents results from experiments that are mainly run with students we want to stress that the ideas presented in this thesis should be tested with real applications. This should be the main direction for future research.

The thesis consists of two main topics. The first topic is the comparison of different attribute weight elicitation methods. Paper [I] presents an experiment where the use of words in preference elicitation is studied. Five different weight elicitation methods are compared in paper [II] that presents results from an experiment run through the Internet. As far as we know, this is the first time when an experimental study related to attribute weighting used the Internet in data collection. In this thesis, however, the special issues related to modern information technology are discussed only briefly (see Appendix A of publication [II]). Paper [III] focuses particularly on group decision making and interactive weighting. Paper [III] presents through an experiment how a new weighting method, Preference Programming (Salo and Hämäläinen, 1995), and a related software (Hämäläinen and Kettunen, 1994) work in a realistic decision making situation.

The second topic of the thesis is the structural variation of value trees. Attribute weights should not change if the way to present attributes is slightly varied. Attribute weights should remain the same even if attributes are divided or if the objectives are organized in a value tree in a new way. In paper [IV] we describe some problems appearing in the earlier studies on the biases related to the structural variation of value trees. In the earlier studies the averages over large group of subjects are used to describe the behavior of individual decision makers. We suggest that based on those results one cannot conclude that biases appear at the individual level. The ideas presented in paper [IV] were tested in an experiment reported in paper [V]. Structural variation of value trees indeed affects the weights of individual subjects. We also present some possible explanations for the phenomena reported in paper [V]. Paper [V] also summarizes how many of the phenomena related to the attribute weighting have the same origins.

Although we know that weighting biases appear in experiments we do not have any reports whether they appear in real situations or not. Neither there exist any guidelines for practitioners how to avoid the biases. The attribute weighting experiments so far differ from practical situations so that weighting have not been interactive. This is a clear shortage in our experiments too. Paper [VI] discusses how much the interactiveness in the weight elicitation and the learning of the methodology alone affect the weighting biases appearing when the structure of a value tree is changed. The rest of this summary briefly presents the main results in papers [I] -[VI].

2 Multiattribute Value Tree Analysis

Throughout this work I deal with multiattibute value tree analysis only. Value tree analysis is one methodology that is today used in practice to support decision making and that is implemented on many user-friendly softwares. MAVT is not, however, the only methodology developed for multi objective decision making situations. There is a wide range of methods based on multiobjective optimization. Details of many of these methods can be found for example from Steuer (1986). So called outranking methods, such as ELECTRE and PROMETHEE, ask decision makers only ranking information and aim at finding outranking relations between alternatives that finally lead to one dominating alternative (see e.g. Roy, 1990). Psychological research attempts to find models that describe as accurately as possible the way people do make decisions involving multiple objectives. These ideas are used, for example, in marketing research where methods such as conjoint analysis (see e.g. Green and Srinivasan, 1990) are widely used. It is actually shown in experiments that models like prospect theory better describe the human behavior than utility or value theory (Kahneman and Tversky, 1979; Korhonen et al., 1990). However, it is not yet proved which methodology is superior to *support* decision making process.

Value tree analysis is not necessarily the winner among the methods but it is at this moment used in practice. Especially the Analytic Hierarchy Process (Saaty, 1984; 1990), which is in practice used in the same way than other value tree methods, has made its way to many practical applications. It is not anymore a tool of academic interest only. These methods are traditionally used in situations where the decision making problems are extremely complex and involve a lot of qualitative and subjective elements. Typically the decisions are long term strategic decisions when it is justified to go through extensive and expensive analysis. Examples of such problems are strategic energy decisions (e.g. Hämäläinen, 1988; 1990; Keeney et al., 1986; Keeney and McDaniels, 1992;), decisions on research project selection (e.g. Islei et al., 1991; Gear and Read, 1993), and organizational planning problems (Belton et al., 1997). However, especially the AHP is used also with smaller scale problems where it is used to capture expert opinions and values. In these cases the value tree methodology does not provide the whole framework for the analysis but is used to solve some smaller subproblems where there is a need to take many criteria and subjective opinions explicitly into account. These applications do not necessarily end up to be reported in scientific articles that usually focus on complex, public policy applications. Furthermore, these analyses are usually done by people who are not very familiar with the methodological details.

One advantage of value tree methodology is indeed that it is basically simple method for decision makers to understand and for everybody to use. Simplicity does not mean, however, that the method would always be used and understood correctly. Theoretical details of value theory can be found from Krantz et al. (1971), Keeney and Raiffa (1976), and Dyer and Sarin (1979). Value tree analysis is suitable for decision making problems where there exists a discrete set of alternatives that are evaluated with respect to a set of quantitative or qualitative attributes. The alternative $\mathbf{x}$ is described as a set of outcomes, called also consequences, $(x_1, ..., x_n)$, where n is the number of attributes. This thesis does not consider utility theory, i.e. probability and utility function, related issues. Thus, the analysis is restricted to the cases where the consequences of alternatives do not include uncertainty. The component value of each consequence x_i, $v_i(x_i)$, is described either with a value function or the decision maker gives values directly (rating). These values are normalized to be between 0 and 1. With additive value model the overall value of an alternative is

$$v(x) = \sum_{i=1}^{n} w_i v_i(x_i),$$

where w_i is the weight of an attribute i. The attribute weight reflects the relative importance of the change in the attribute from the worst attribute level to the best attribute level. The attribute weights are usually normalized to sum up to one. This additive value model can be used only if the attributes are mutually preferentially independent (see details in Keeney and Raiffa (1976)). This is often not the case and most often the assumption is not even validated. However, additive structure for preferences is assumed because it is simple and with that model the attribute weights have a more intuitive interpretation. In this thesis we assume that the decision making problem is structured in a way that we can assume the value function to be additive.

There exist many weighting methods that use different questioning procedures to elicit weights w_i, such as SMART (Edwards, 1977) and SWING weighting (von Winterfeldt and Edwards, 1986; Edwards and Barron, 1994). This thesis studies especially the Analytic Hierarchy Process (AHP) (Saaty, 1980; 1994) that is in principle a separate methodology based on different theoretical assumptions. The discussion on the relation between the AHP and multiattribute value tree analysis, however, is going on. Many researchers think that the AHP and value tree analysis are both based on value theory (see Salo and Hämäläinen, 1997, and the rejoinders). In this thesis the AHP is treated as one multiattribute value tree weighting method because the weight elicitation questions presented to the decision maker are almost the same with these two groups of methods and thus they should lead to the same attribute weights.

The attributes are organized as a value tree (called also as an objectives hierarchy or a decision hierarchy) where the groups of attributes are called as objectives (Keeney and Raiffa, 1976; Keeney, 1992). In the AHP vocabulary the attributes and also the objectives at the upper levels of a value tree are called criteria. In the papers of this thesis both the terms attribute and criterion are used. Figure 1 shows an example of a value tree that was used in paper [III] to evaluate traffic plans in Helsinki metropolitan area.

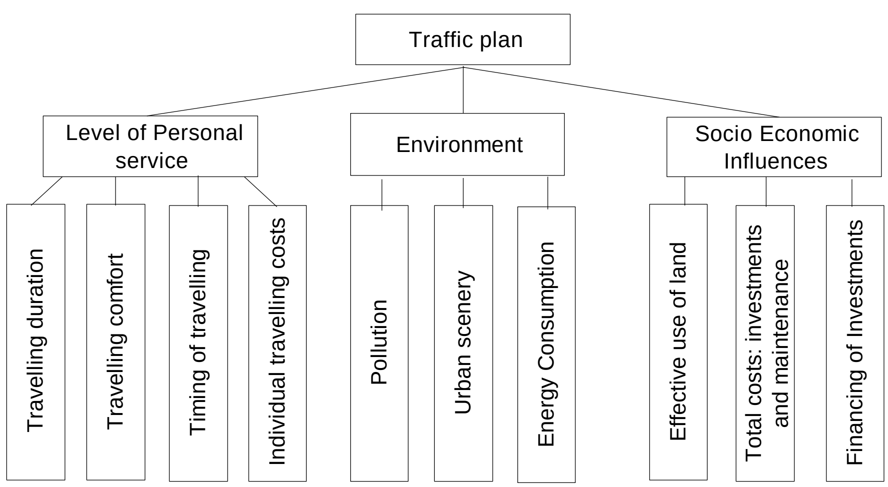

Figure 1: *An example of a value tree.*

The formulation of the value tree is not an easy task and takes most of the time in real applications. Value trees should include all relevant aspects but still be as small as possible, the attributes included into a value tree should be preferentially independent of each other if the additive value model is used, and a value tree should represent the opinions of all the stakeholders. Phillips (1984), Gregory and Keeney (1994), and

Belton et al (1997), among others, give good overviews of problems and solutions in problem formulation. The value tree formulation should be a process involving all the stakeholder groups and thus is a group decision making situation. The role of the facilitator is crucial in this process.

In a value tree the weights are elicited either by weighting all the lowest level attributes simultaneously (non-hierarchical weighting) or by weighting attributes and objectives within one branch at a time and multiplying the local weights through the value tree (hierarchical weighting). Figure 2 shows examples of these two ways to weight attributes. One should note that in the analysis only the final attribute weights are needed. With non-hierarchical weighting the decision maker is not asked to weight objectives and these weights are actually not needed at all. The weight of an objective at the upper level is by definition the sum of the attribute weights below. A problem with non-hierarchical weighting is that in real applications the total number of attributes presented to a decision maker at one time may be very large. The details of value tree analysis and weighting methods can be found from (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986; Weber and Borcherding, 1993) and from papers [IV], [II], and [VI].

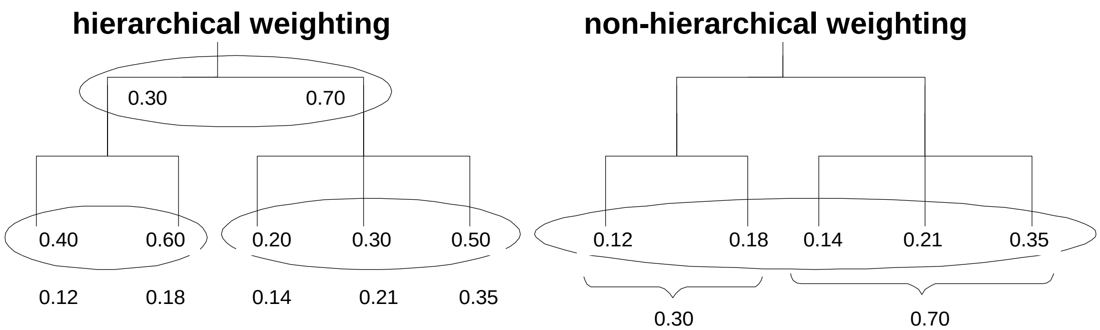

***Figure 2**: In hierarchical weighting weights are elicited and normalized within each level and branch at the time, the final attribute weights are calculated by multiplying the local weights down through the tree. In non-hierarchical weighting all the lowest level weights are elicited simultaneously.*

3 Comparison between attribute weight elicitation methods

Differences between weighting methods

The interpretation of an attribute weight within value theory is the same regardless of the weighting method. The attribute weight reflects the relative importance of the change in the attribute from the worst attribute level to the best attribute level. This interpretation of a weight is not followed within all multi criteria decision making methods. For example, outranking methods try to avoid this interpretation for weights.

Most of the weighting methods used in value trees elicit the weights by focusing on the comparisons between two attributes at the time (the AHP, SMART, SWING, Tradeoff weighting). The decision makers are asked to give weight ratios. In SMART, for example, the decision maker is asked to identify the least important attribute and compare other attributes against this least important attribute. It is surprising that the methods yield different weights although they are based on similar statements on weight ratios (see e.g. Hobbs, 1980; Schoemaker and Waid, 1982; Belton, 1986; Borcherding et al., 1991; Srivastava et al. 1995; Olson et al. 1995). One clear difference between the methods is the way how the ranges of attributes are presented in the elicitation questions. For example, in the AHP the attribute range is not presented at all, while in SWING a longer questioning form tries to ensure that the decision maker is really thinking the presented attribute ranges. The range effect presented by von Nitzsch and Weber (1993) and Fischer (1995) proposes that decision makers ignore the attribute ranges with the weighting methods that do not stress it.

The decision makers are, however, willing to answer the questions asking very generic "importance" of attributes (Korhonen and Wallenius, 1995; Belton and Gear, 1997; Salo and Hämäläinen, 1997). This means that decision makers have some intuitive interpretation for weights but the weight is not necessarily reflecting the attribute range of that particular problem (see also paper [VI]). Yet, the differences between the weighting methods may partly originate from the way the elicitation questions introduce the meaning of an attribute weight to decision makers. This hypothesis was tested in the experiment of paper [II]. With the AHP the attribute ranges were either presented or not while otherwise the questions were kept the same. There were no clear differences between these modes, but in that experiment this result is partly explained by the experimental design.

In addition to the formulation of the elicitation question another difference between the weighting methods is the numerical scale that they explicitly or implicitly use. An extreme case is the AHP that restricts the decision makers responses so that they can use only few numbers, integers from one to nine (Saaty, 1980, 1994). This scale restricts the resulting weights. Furthermore, the integers lead to an uneven distribution of weights (Salo and Hämäläinen, 1997). For example, the change of the weight ratio from 1 to 2 leads to big changes in weights whereas the change in weights is almost invisible when the weight ratio is changed from from 8 to 9. Salo and Hämäläinen (1997) proposed new balanced scales that allow to reach weights that cover the range from 0 to 1 evenly. One such scale is used in papers [I] and [II] to see how the change of the scale affects weights. With this scale the integers from 1 to 9 are replaced with numbers 1, 1.22, 1.50, 1.86, 2.33, 3.00, 4.00, 5.67, and 9.00. The main conclusion from both papers [I] and [II] is that the change of the evaluation scale used in the AHP affects the weights considerably.

The AHP also uses verbal expressions to help the decision makers to describe the strength of their preferences. The results of paper [I] suggest that the use of verbal expressions in preference elicitation includes many risks. The words are highly context dependent so that the numerical interpretation of the word varies from one decision making situtation to another depending on the set of alternatives or attributes compared at the time. There are also interpersonal differences. Generally the decision makers fail to understand the relationship between the verbal exressions and numbers. The same result appears also in paper [II] where subjects used the words equally often although the corresponding numbers varied (see Table 1). For example, with the nine point integer scale the verbal statement “strongly more important“ is connected with the weight ratio 5 whereas with the balanced scale the same ratio is 2.33. With both scales, however, the decision makers use this verbal statement equally often. There is a clear analogy between the use of words in preference elicitation and the use of words in probablility assessment (for a review see Budescu and Wallsten, 1995). The context dependency of the preferences is an interesting research question and the consequences and possible sources of biases in this area have not yet been thoroughly studied in connection with value trees (Goldstein, 1990; Tversky and Kahneman, 1992; Tversky and Simonson, 1993).

	Scale	
Verbal statement	1-to-9	Balanced
Equally important	7%	6%
-	7%	6%
Slightly more important	19%	14%
-	15%	11%
Strongly more important	22%	18%
-	8%	14%
Very strongly more important	13%	19%
-	3%	5%
Extremely more important	6%	7%

Table 1: *The percentages show how often each verbal expression was used with different evaluation scales (data from paper [II]).*

Other weight elicitation techniques do not give evaluation scales explicitly, but they implicitly propose numbers that the decision makers tend to use. The weighting methods compared in paper [II] do yield different weights, but we propose that the differences originate from the way decision makers select numbers to describe their preferences. The decision makers increase or decrease the given starting numbers by almost even steps and restrict their responses according to the number that they start with. The SMART, for example, asks the decision maker to give 10 to the least important attribute and evaluate other attributes against this starting point. With this method decision makers tend to use numbers that are multiples of ten. A decision maker uses, for example, numbers 10, 20 , 40 and 70 in case of four attributes. The same decision makers, however, may use numbers 70, 80, 90, and 100 for the same attributes while weighting with SWING that asks the decision maker to start by giving 100 point for the most important attribute. This kind of restrictions of the evaluation scales affect the weights. Our results suggest that this is the main origin leading to the differences in the weights elicited with different weighting methods. The decision makers use numbers that reflect the rank of attributes only against the underlying theoretical assumptions.

Weighting methods converge

Are the numbers used in connection with a weighting method an essential feature that defines that method? The answer is no because with every weighting method it is allowed to use any numbers. Other scales can be used with the AHP and one can start the SMART procedure from whatever number. Are there any characteristics that make the weighting methods different? The weighting methods are approaching each other in formulating the elicitation question (Edwards and Barron, 1994; Salo and Hämäläinen, 1997) and new softwares combine different methods efficiently (e.g. Hämäläinen and Lauri, 1993). For example, from the point of view of a user there is no difference in the AHP and SMART questions if the same evaluation scale is used with both methods. The methods do have attractive special features like the consistency checking in the AHP and the incorporation of attribute ranges into the SWING questions. These features of the weighting methods, however, can easily be combined and the weighting procedures can de adapted for a particular application. There is no superior weighting method. The researchers should focus more on the procedures how the methods are used in practice in interaction with decision makers and how the decision makers interpret the results instead of searching for superior theoretical base for the methods.

One promising direction to develop weighting methods to be more practical is to make them more interactive and use more computer support. So called preference programming techniques are value tree methods where the interactive weight elicitation is an essential part of the methods (Salo and Hämäläinen, 1992; 1995). With preference programming methods the weight elicitation questions can basically follow whatever value tree method such as AHP or SMART. The key idea is that decision makers do not give single numbers to describe their preferences but instead they give intervals of numbers for weights and values. These preference intervals are interactively revised until dominances between the alternatives are found (Salo and Hämäläinen, 1992; Salo and Hämäläinen, 1995). These methods are suggested to be especially useful in group decision making as the intervals of preference statements can be used to describe the diversity of opinions within a group (Hämäläinen et al., 1991; Hämäläinen and Leikola, 1994). The new method and the related information technology are tested in paper [III] in a group decision making situation. The paper showed that the use of this kind of a method can stimulate and structure the discussion and helps the groups to proceed in a controlled way.

Although only one weighting method was used in paper [III] there are still many ways to proceed in a group. The experiment in paper [III] studied differences between two procedures used to guide the interaction between the decision makers. In this case clear differences between the procedures did not show up but it was clear that each group was progressing on its own way. Most probably the results of the paper would look the same even if the weight elicitation questions would have followed another weighting method procedure. The weighting method itself is not the essential ingredient in stimulating the discussion. The weighting methods are only tools used in the analysis and the main thing is to focus on the process how they are used.

4 Consequences of structural variation of value trees

In addition to different weighting methods another source affecting the attribute weights against theoretical expectations is the structure of a value tree. First papers reporting that the structure of the value tree unexpectedly affects the attribute weights are Stillwell et al. (1987), Weber et al., (1988), and Borcherding and von Winterfeldt (1988). Since then, very few new studies have appeared. All these experiments are run with students and the conclusions on the changes in weights are mainly based on the averages of weights over large groups of subjects. In paper [IV] we show that by averaging the weights one looses information on the actual extent and origin of the biases. In the worst case the averages are nobody's opinions. Thus, one cannot conclude how the biases appear at the individual level (see Figure 3). In paper [IV] we also show that weighting methods that use the rank of attributes only in weight calculation, such as SMARTER (Edwards and Barron, 1994; Barron and Barrett, 1996), lead to biased weights if the attributes are divided into subattributes although the decision maker is fully consistent in his evaluations.

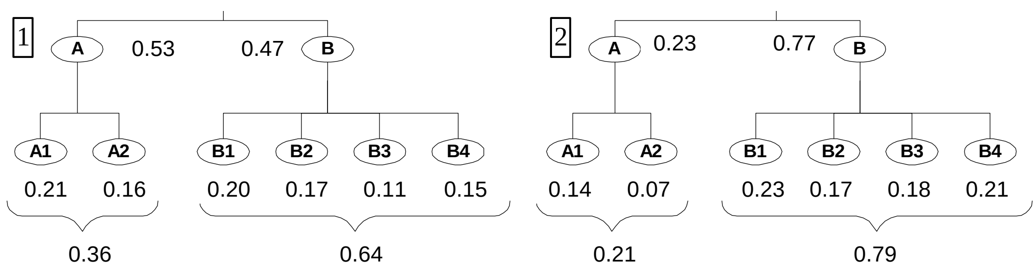

Figure 3: *The value tree 1 shows the splitting bias for averages of weights and the value tree 2 shows consistent weights of one subject that was included in the averages of the tree 1 (data from paper [V]).*

Paper [IV] is a note that criticizes the earlier experiments. The main conclusion of paper [IV] is that we do not know whether the attribute weights change for an individual decision maker depending on the structure of a value tree and why this may happen. The answers for these questions were looked for in Paper [V] that presents results from a new experiment studying the division of attributes in a value tree. First result of the paper is that the weighting biases indeed appear differently if the conclusions are based on the averages of weights instead of weights from individual subjects (see Figure 3). Second result of the paper is that also at the individual level there are biases in weights. The changes in weights depend, however, on the subject and on the structure of the value tree. The division of an attribute, for example, can decrease as well as increase the weight of an attribute. The main result of the paper is that an origin for the biases in weights is that the subjects give numbers to describe their preferences so that they clearly favor some response scales. They behave in the same way than in the experiment of paper [II]. Basically subjects give numbers that reflect ordinal information on preferences only. In this way they interpret the numbers differently than what the theory of value measurement assumes. The use of restricted response scales and the normalization of weights together lead to biases in weights if the structure of a value tree is changed.

It is important to know the origins for biases so that they can be avoided in practice. Real world applications do not report on any problems in weight elicitation in value trees. It has been suggested that the interactive weight elicitation eliminates the biases (Weber and Borcherding, 1993; Marttunen and Hämäläinen, 1995). This has not, however, been systematically tested. In paper [VI] we present preliminary thoughts how the learning of the methodology can reduce biases in weights when the structure of a value tree is changed. The experiments were run in the same way as the original studies on the splitting bias and the range effect (Weber et al., 1988; von Nitzsch and Weber, 1993). The groups are, however, small and thus these experiences are very preliminary.

5 Connections between weighting biases

Together the results in papers [II], [IV], and [V] show that two origins for many weighting biases are that decision makers give numbers that reflect the rank of attributes only and that weights are normalized to sum up to one so that they depend on each other. In paper [V] it is proposed that a group of weighting biases originates from these two sources (see Figure 4). We suggest that the differences between the weighting methods and the changes in weights when the structure of the value tree is changed are results from decision makers' tendency to give ranking information only of attributes. In the AHP there appears the rank reversal phenomenon that originates from the normalization of component values (Belton and Gear, 1983; Salo and Hämäläinen, 1997). The so called number-of-attribute-levels effect in conjoint analysis has been shown to originate from the use of ordinal information: decision makers are asked or they anyway give only ranks of alternatives (Wittink et al., 1982; 1989; Steenkamp and Wittink, 1994). The summary of these two sources for the biases is not yet fully tested and the framework presented in Figure 4 is meant to be a proposal to stimulate future research.

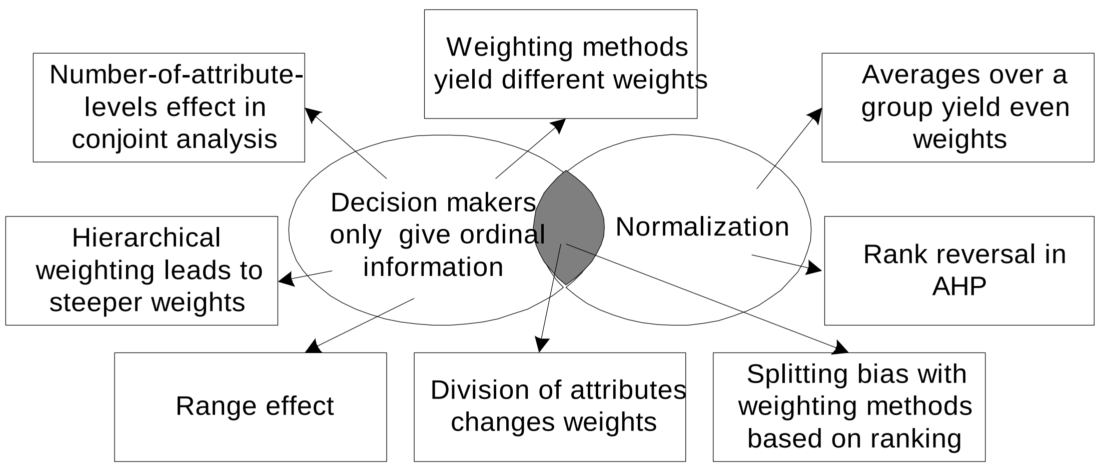

Figure 4: *Different biases and theoretical problems in attribute weighting originate from the decision makers' tendency to give ordinal information on preferences only and from the normalization of weights.*

6 Conclusions

To conclude this thesis work I want to raise up three main conclusions that are further elaborated in this chapter:

- In practice there are no differences between the weighting methods and any value tree weighting method is equally acceptable.

- The biases in weights are real. The structure of a value tree and the selection of weighting methods affect results. The biases originate from the used weighting procedures as well as from the behavior of decision makers.

- The origins for many of these phenomena are the same. When the origins are known, it is possible to develop the analysis procedures to be bias-free.

This far the results found in experiments have had too little impact on the practice of weight elicitation. The main message for practitioners is that all the different methods are equally applicable. None of the methods, however, should be used without an analyst who is able to guide the decision maker through consistency checkings and knows the details and weak points of each method. All the possible sources of biases cannot be controlled in practice. Thus, it is a challenge to develop the analysis procedures so that only the most relevant consistency checkings are made.

Still, we have not seen any reports on biases in connection with applications. One simple reason for this is that in practice the existence of the biases is not checked. All the phenomena discussed in this thesis are based on more or less controlled laboratory experiments. The laboratory experiments are not the same as real situations and there are several sources of errors in experiments itselves. The results on biases also vary from one experiment to another. Generally the link from controlled experiments to practical situations is not straightforward. However, there exists at least one exception. The experimental setup resembles the situation of survey studies and in this sense the laboratory results already warn that the weighting biases appear when people are asked the attribute weighting type of questions on paper without further interaction with them.

The interactive weight elicitation has the key role in practical weighting and should be the main direction for future studies on attribute weighting. With many multicriteria decision making techniques based on optimization the interactive evaluation of alternatives at each step is the essential part of the methodology. Also in that field the behavioral aspects of the interactiveness start to get more attention (Buchanan, 1994; Korhonen and Wallenius, 1996). A normal part of computerized weighting is the interactive evaluation of the results during the weighting. Yet, the computer implementation of the weighting methods has not received as much attention as it should receive.

The comparison of individual weighting techniques with individual decision makers without interactive weighting process in a laboratory experiment is a relatively easy task. With interactiveness and, especially, with a group of decision makers the experimental research on attribute weighting is far more complicated. Group decision making has its own special features. The interactiveness is often combined with computer support. The role of computer interface and the interactiveness should be separated in research setups. The facilitator has a major role in real life decision making when these types of methods are used. These several factors make it much more difficult to study interactiveness. The more affecting factors we have the more difficult it is to test them in real life situations in a controlled way. In future the main new results should emerge from carefully designed case studies.

In my opinion these case studies should start by raising up the role of a facilitator. The basic assumption is that an analyst, whose responsibility is to elicit the decision maker’s preferences, is skillful and aware of the biases. However, we do not know yet whether a facilitator is enough to prevent the biases in weights. In the case studies the biases should be checked throughout the weighting process with extra questions that, for example, check what happens if an attribute is divided. The analyst makes comparisons between different weights and ask a decision maker elaborate on the inconsistencies. In this way we should end up with a set of bias-free weights. The inconsistencies in weights that have appeared tell us about the extent of this type of problems in practice. After we have some more knowledge how the weighting in a real situation proceeds with a facilitator, we should add research questions related to, for example, computer aid. The big question is whether decision support methods, such as value tree analysis or the AHP, ever describe reliably the preferences of decision makers without the help of an analyst.

References

Barron, F.H. and Barrett, B.E., (1996), "Decision Quality Using Ranked Attribute Weights," *Management Science*, 42, 1515-1523.

Belton, V. and Gear, T., (1983), "On a Short-coming of Saaty's Method of Analytic Hierarchies," *OMEGA*, 3, 228-230.

Belton, V., (1986), "A Comparison of the Analytic Hierarchy Process and a Simple Multi - Attribute Value Function," *European Journal of Operational Research*, 26, 7-21.

Belton, V. and Gear, T., (1997), "On the Meaning of Relative Importance," forthcoming in *Journal of Multi-Criteria Decision Analysis*.

Belton, V., Ackermann, F., and Shepherd, I., (1997), "Integrated Support from Problem Structuring through to Alternative Evaluation Using COPE and VISA," *Journal of Multi-Criteria Decision Analysis*, 6, 115-130.

Borcherding, K. and von Winterfeldt, D., (1988), "The Effect of Varying Value Trees on Multiattribute Evaluations," *Acta Psychologica*, 68, 153-170.

Borcherding, K., Eppel, T. and von Winterfeldt, D., (1991), "Comparison of Weighting Judgments in Multiattribute Utility Measurement," *Management Science*, 37, 12, 1603-1619.

Borcherding, K., Schmeer, S. and Weber, M., (1994), "Biases in Multiattribute Weight Elicitation," in Caverni, J.-P., Bar-Hillel, M., Barron, H. and Jungermann, H. (Eds.) *Contribution to decision research*, Amsterdam, North-Holland.

Buchanan, J., (1994), "An Experimental Evaluation of Interactive MCDM Methods and the Decision Making Process," *Journal of Operational Research Society*, 45, 1050-1059.

Budescu, D. and Wallsten, T.S., (1995), "Processing Linguistic Probabilities: General Principles and Empirical Evidence," *The Psychology of Learning and Motivation*, 32, 275-318.

Corner, J.L. and Kirkwood, C.W., (1992), "Decision Analysis Applications in the Operations Research Literature 1970-1989," *Operations Research*, 2, 206-219.

Dyer, J.A. and Sarin, R.K., (1979), "Measurable multi-attribute value functions," *Operations Research*, 27, 810-822.

Edwards, W., (1977), "How to use multiattribute utility measurement for social decision making," *IEEE Transactions on Systems, Man and Cybernetics*, SMC-7, 326-340.

Edwards, W. and Barron, F.H., (1994), "SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement", *Organizational Behavior and Human Decision Processes*, 60, 306-325.

Fischer, G.W., (1995), "Range Sensitivity of Attribute Weights in Multiattribute Value Models," *Organizational Behavior and Human Decision Processes*, 62, 252-266.

Fraser J.M., Smith P.J., and Smith, J.W. (1992), "A catalog of errors," *International Journal of Man-Machine Studies*, 37, 265-307.

Gear, A.E. and Read, M.J., (1993), "On-line group process support," OMEGA, 21, 261-274.

Goldstein, W.M., (1990), "Judgments of Relative Importance in Decision Making: Global vs. Local Interpretations of Subjective Weight," *Organizational Behavior and Human Decision Processes*, 47, 313-336.

Green, P.E. and Srinivasan, V. (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," Journal of Marketing, 54, 3-19.

Gregory, R. and Keeney R.L. (1994), "Creating policy alternatives using stakeholder values," *Management Science*, 40, 1035-1048.

Hobbs, B.F., (1980), "A Comparison of Weighting Methods in Power Plant Siting, " *Decision Science*, 11, 725-737.

Hobbs, B.F., (1986), "What Can We Learn From Experiments in Multiobjective Decision Analysis?" *IEEE Transactions on Systems, Man and Cybernetics*, 16, 3, 384-394.

Hogarth, R.M., (1980), *Judgement and Choice*, New York, John Wiley & Sons.

Hämäläinen, R.P., (1988), "Computer Assisted Energy Policy Analysis in the Parliament of Finland," *Interfaces*, 18, 12-23.

Hämäläinen, R.P., (1990), "A decision aid in the public debate on nuclear power," *European Journal of Operational Research*, 48, 66-76.

Hämäläinen, R.P., Salo, A.A. and Pöysti K. (1991). "Observations about consensus seeking in a multiple criteria environment," *Proceedings of the Twenty-Fifth Annual Hawaii International Conference on System Sciences*, Vol. IV, IEEE Computer Society Press, 190-198.

Hämäläinen, R.P. and Lauri, H., (1993), *HIPRE 3+ Decision Support Software vs. 3.13*, Helsinki University of Technology, Systems Analysis Laboratory.

Hämäläinen, R.P. and Kettunen, E., (1994), *HIPRE 3+ Group Link, User's Guide*, Helsinki University of Technology, Systems Analysis Laboratory.

Hämäläinen , R.P. and Leikola, O., (1994), "Spontaneous Decision Conferencing in Parliamentary Negotiations," *Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press.

Islei, G., Lockett, G., Cox, B., and Stratford, M. (1991), “A decision support system using judgmental modeling: A case of R&D in the pharmaceutial industry,“ *IEEE Transactions on Engineering Management*, 38, 202-209.

Kahneman, D. and Tversky A., (1979), “Prospect Theory: An Analysis of Decisions Under Risk,“ *Econometrica*, 47, 263-292.

Keeney, R.L. and Raiffa, H., (1976), *Decisions with Multiple Objectives : Preferences and Value Trade-offs*, Wiley, New York.

Keeney, R.L., Lathrop, J.F. and Sicherman, A., (1986), "An Analysis of Baltimore Gas and Electric Company's Technology Choice," *Operations Research*, 34, 18-39.

Keeney, R.L. and McDaniels, T.L., (1992), "Value-Focused Thinking about Strategic Decisions at BC Hydro," *Interfaces*, 22, 94-109.

Korhonen, P., Moskowitz H., and Wallenius J., (1990), “Choice Behavior in Interactive Multiple-Criteria Decision Making,“ *Annals of Operations Research*, 23, 161-179.

Korhonen, P. and Wallenius, J., (1996), "Behavioral Issues in MCDM: Neglected Research Questions," *Journal of Multi-Criteria Decision Analysis*, 5, 178-182..

Krantz, D.H., Luce, R.D., Suppes, P. and Tversky, A., (1971), *Foundations of Measurement*, Vol I, Academic Press, New York.

Marttunen, M. and Hämäläinen, R.P., (1995), "Decision analysis interviews in environmental impact assessment," *European Journal of Operational Rsearch*, 87, 551-563.

Miettinen, P. And Hämäläinen, R.P., (1997), "How to Benefit from Decision Analysis in Life Cycle Assessment," to appear in *European Journal of Operational Rsearch.*

Olson, D.L., Moshkovich, H.M, Schellenberg, R., and Mechitov, A.I., (1996), "Consistency and Accuracy in Decision Aids: Experiments with Four Multiattribute Systems," *Decision Sciences*, 26, 723-748.

Phillips, L.D. (1984), “A theory of requisite models,“ *Acta Psychologica*, 56, 29-48.

Phillips, L.D. and Phillips, M.C., (1993), "Facilitated Work Groups: Theory and Practice," *Journal of Operational Research Society*, 44, 533-549.

Roy, B., (1990), “The outranking approach and the foundations of ELECTRE methods,“ in *Readings in Multiple Criteria Decision Aid*, *ed. Bana e Costa*, Springer, Berlin.

Saaty,T.L., (1980), *The Analytic Hierarchy Process*, McGraw-Hill, New York.

Saaty, T.L., (1994), "Highlights and critical points in the theory and application of the Analytic Hierarchy Process," *European Journal of Operational Research*, 74, 426-447.

Salo, A.A. and Hämäläinen, R.P., (1992), "Preference Assessment by Imprecise Ratio Statements," *Operations Research*, 40, 6, 1053-1061.

Salo, A.A. and Hämäläinen, R.P., (1997), "On the measurement of preferences in the analytic hierarchy process," (and comments by V. Belton, E. Choo, T. Donegan, T. Gear, T. Saaty, B. Schoner, A.Stam, M. Weber, B. Wedley) *Journal of Multi-Criteria Decision Analysis*, 1997, 6, 309-343..

Salo, A.A. and Hämäläinen, R.P., (1995), "Preference programming through approximate ratio comparisons," *European Journal of Operational Research*, 82, 458 - 475.

Schoemaker, P.J. and Waid, C.C., (1982), "An Experimental Comparison of Different Approaches to Determining Weights in Additive Value Models," *Management Science*, 28, 2, 182-196.

Srivastava, J., Connolly, T., and Beach, L. R., (1995), "Do Ranks Suffice? A Comparison of Alternative Weighting Approaches in Value Elicitation," *Organizational Behavior and Human Decision Processes*, 63, 112-116.

Steenkamp, J.E.M. and Wittink, D.R. (1994), "The Metric Quality of Full-Profile Judgments and the Number-of-Attribute-Levels Effect in Conjoint Analysis," *International Journal of Research in Marketing*, 11, 275-286.

Steuer, R., (1986), *Multiple Criteria Optimization: Theory, Computation, and Application*, Wiley, New Yorl, 1986.

Stewart, T.J., (1992), "A Critical Survey on the Status of Multiple Criteria Decision Making Theory and Practice," *OMEGA*, 20, 5-6, 569-586.

Stillwell, W.G., von Winterfeldt, D. and John, R.S., (1987), "Comparing Hierarchical and Nonhierarchical Weighting Methods for Eliciting Multiattribute Value Models," *Management Science*, 33, 4, 442-450.

Tversky, A. and Kahneman, D., (1992), "Loss aversion in riskless choice: A reference dependent model," *Quarterly Journal of Economics*, 107, 1039-1041.

Tversky, A. and Simonson I., (1993), "Context-dependent Preferences," *Management Science*, 39, 10, 1179-1189.

Weber, M., Eisenfuhr, F. and von Winterfeldt, D., (1988), "The Effects of Splitting Attributes on Weights in Multiattribute Utility Measurement," *Management Science*, 34, 431-445.

Weber, M. and Borcherding, K., (1993), "Behavioral influences on weight judgments in multiattribute decision making," *European journal of Operational Research*, 67, 1-12.

Wittink, D.R., Krishnamurthi, L. and Nutter, J.B., (1982), "Comparing Derived Importance Weights Across Attributes," *Journal of Consumer Research*, 8, 471-474.

Wittink, D.R., Krishnamurthi, L. and Rebstein, D.J. (1989), "The Effect of Differences in the Number of Attribute Levels on Conjoint Results," *Marketing Letters*, 1, 113-123.

von Nitzsch, R. and Weber, M., (1993), "The Effect of Attribute Ranges on Weights in Multiattribute Utility Measurements," *Management Science*, 39, 8, 937-943.

von Winterfeldt, D. and Edwards, W., (1986), *Decision Analysis and Behavioral Research*, Cambridge University press, Cambridge, MA.