Helsinki University of Technology

Systems Analysis Laboratory Research Reports E12, April 2003

THE THREAT OF WEIGHTING BIASES IN ENVIRONMENTAL DECISION ANALYSIS

Raimo P. Hämäläinen and Susanna Alaja

Working draft updated 2.6.2006

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Distribution:

Systems Analysis Laboratory

Helsinki University of Technology

P.O. Box 1100

FIN-02015 HUT, FINLAND

Tel. +358-9-451 3056

Fax +358-9-451 3096

systems.analysis@hut.fi

This report is downloadable at

www.e-reports.sal.hut.fi/pdf/E12.pdf

Series E - Electronic Reports

www.e-reports.sal.hut.fi

ISBN 951- 22-6486-2

ISSN 1456-5218

Abstract

We investigate the existence of biases, in particular the so called splitting bias, in a real environmental decision analysis case. The splitting bias refers to a situation where presenting an attribute in more detail may increase the weight it receives. We test whether the splitting bias can be eliminated or reduced through instruction and training. The case was the regulation of a lake-river system and the test group consisted of university students and local residents. The test groups carried out attribute weightings with different value tree structures.

Our results show that the splitting bias remains a real threat. In the weighting experiments most students avoided the bias. However, nearly all of the local residents showed a systematic splitting bias so that the total weight of the attributes grew in proportion to the number of sub-attributes included. We discuss ways to eliminate the biases by balanced problem structuring.

Keywords: Environmental decision making, multicriteria decision analysis, splitting bias, structuring

1. Introduction

Decision analytic value tree evaluation of goals, objectives and alternatives has become common practice in environmental decision making and policy analysis. For references and examples see e.g. Hämäläinen (1992), Marttunen and Hämäläinen (1995), Miettinen and Hämäläinen (1997), Hayashi (2000), Hobbs and Meier (2000), Gregory et al. (2001), Hämäläinen et al. (2001), Janssen (2001), Kangas et al. (2001), Schmold et al. (2001), Bell et al. (2001), and Kiker et al. (2005). The reports on decision analysis studies easily focus on the common observation that decision makers have been very satisfied with the approach. However, in practice the reliability of the results is seldom considered. Due to the growing importance of environmental decision analysis, we need to better understand the potential problems, too (see e.g. Bell et al. 2001, Rauschmayer 2001, Wenstøp and Seip 2001, Leskinen and Kangas 2005). These include the role of the analyst (Brown 2005), biases and procedural mistakes one can face in weight elicitation (Keeney 2002) as well as the interpretation of the criteria weights (Hämäläinen and Salo 1997). The effects of attribute ranges are also a related issue and a potential source of problems (von Nitzsch and Weber 1993, Fisher 1995). Moreover, there is no guarantee of the convergence of the results with different weighting methods (Stillwell et al. 1987, Pöyhönen and Hämäläinen 2001). There exists a number of cognitive biases caused by the structure of the value tree which one should try to avoid in practical applications, see Weber et al. (1988), Borcherding and von Winterfeldt (1988), Weber and Borcherding (1993), Baron (1997), Pöyhönen and Hämäläinen (1998), Pöyhönen et al. (2001). In environmental problems the effects of the model structure are most important as there often are alternative possibilities for the structuring. Naturally we would like to work with models and weighting procedures which are not sensitive to the structure. These behavioural issues are finally receiving some attention also in the introductory texts on decision analysis (e.g. Hobbs and Meier 2000; Belton and Stewart 2002). There are already some environmental studies taking the possibility of biases into account see e.g. Pöyhönen and Hämäläinen (2000), Bell et al. (2001), Bojórquez-Tapia et al. (2005), Marttunen and Hämäläinen (2006). However, so far these biases have not been studied systematically on an individual level or with a real life case. Our aim is to investigate how the structure of the value tree affects the relative importance of different objectives in an environmental decision context when the decision makers have good instructions and direct feedback about the results. Can the threat of biases be avoided in this way?

The case considered was the development of a new regulation policy for the Päijänne-Kymijoki lake-river system. An important part of the process was to introduce new methods into public participation (see Marttunen and Hämäläinen 1995, Hämäläinen et al. 2001, Mustajoki et al. 2004, Marttunen and Hämäläinen 2006). Value tree analysis and personal decision analysis interviews were among these. We felt it was also important to study whether it would be possible to avoid or reduce the possible biases through training and proper instructions.

The subjects consisted of local citizens living in the Lake Päijänne area called stakeholders in the text as well as university students. The university students acted as a reference group having taken a course in decision analysis.

Our main focus is the splitting bias as it is likely to be very easily present in environmental applications where value trees of different structure and detail are typically tested and evaluated before selecting the one to be used. The early studies on biases have been based on group averages with student subjects (Weber et al., 1988; Borcherding and von Winterfeldt 1988), while our recent analyses of (Pöyhönen et al. 2001, Pöyhönen and Hämäläinen 1998, 2000) were based on individual responses. The present study is the first one which investigates biases in a real life case and with real stakeholders too. In addition we also study whether the splitting bias can be avoided or reduced through preparatory training and instruction.

2. Biases and research hypotheses

2.1 Systematic splitting bias

The splitting bias refers to the phenomenon where adding new subattributes to a branch in a value tree produces an increase in the overall weight of that branch. In this study we shall try to find out if people produce this in a systematic way. We study the relationship between the relative magnitude of the splitting bias and the number of attributes in each branch of the value tree. If the decision maker shows no splitting bias, the weight ratios of the overall weights for different branches of the value tree should remain constant. However, if the weight ratio grows in proportion to the ratio of the numbers of attributes then the decision maker follows a systematic splitting bias (see Figure 1).

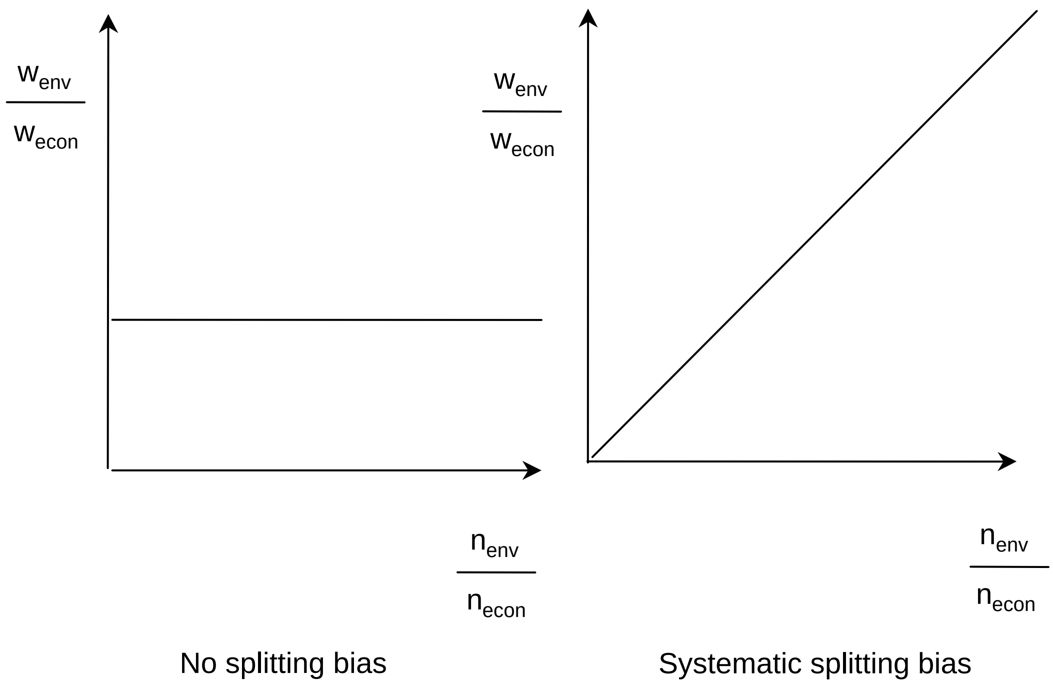

Figure 1. *The effect of the number of attributes on the ratio of branch weights.*

Consider, for example, variations of the value tree (Figure 2) with two main branches, the environment and the economy, and with a different number of subattributes in each branch. Let us assume that we elicit weights on the lowest level and obtain the upper level weights as sums of the twig level weights for each branch. If the responses are nonbiased, the ratio of the total weight of the attributes in the environment branch w_{env} and the total weight of the attributes in the economy branch w_{econ}, i.e. w_{env}/w_{econ} should remain constant. If, however, there is a systematic splitting bias, the weight ratio w_{env}/w_{econ} grows in proportion to the ratio of the number of attributes in these two branches i.e. n_{env}/n_{econ} , where n_{env} is the number of the attributes in the environment branch and n_{econ} is the number of the attributes in the economy branch. Figure 1 illustrates the effects of these two behavioral patterns.

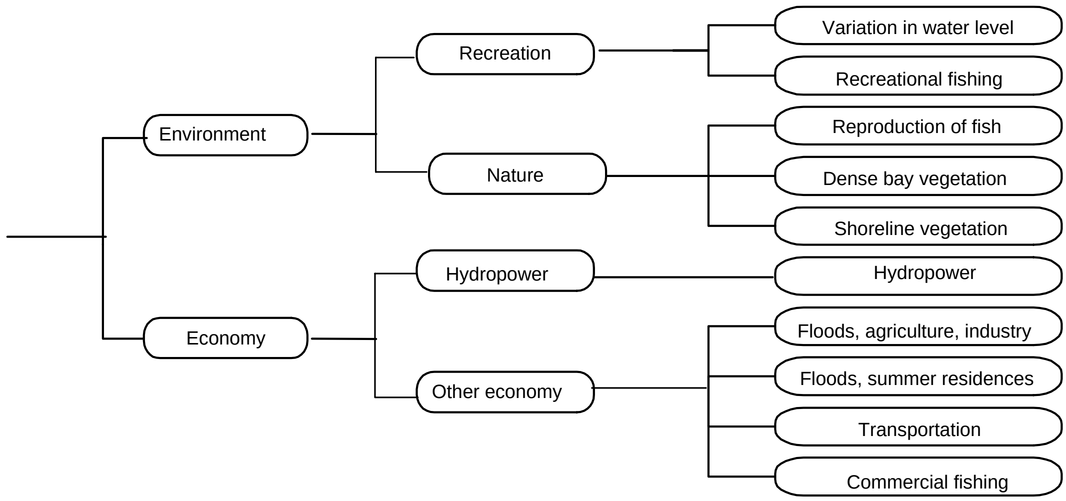

Figure 2. *The value tree used in the experiment.*

The magnitude of the splitting bias may vary between the branches of the value tree. In the Lake Päijänne decision model the value tree is first divided into two branches, the environment and the economy. Usually, the economic attributes are easier to measure. In this study, for example, they are all described in terms of money. Clear and understandable measures for the environmental attributes are often more difficult to find. Also the different measures for the environmental attributes are seldom commensurable.

2.2 Swapping of attribute levels

Another phenomenon examined is the effect of the order of the attributes. In the Lake Päijänne case one of the major questions is how to aggregate the impacts in the Lake Päijänne area and those in the downstream areas and in the River Kymijoki.

Figure 3 shows the two different ways of structuring the problem. Many environmental problems involve similar choices as the impacts are often geographically distributed and the interests of the stakeholder groups vary regionally. If the analysis is reliable the choice between these two models should not yield significant differences in the resulting weights. To our knowledge this question has not been studied before.

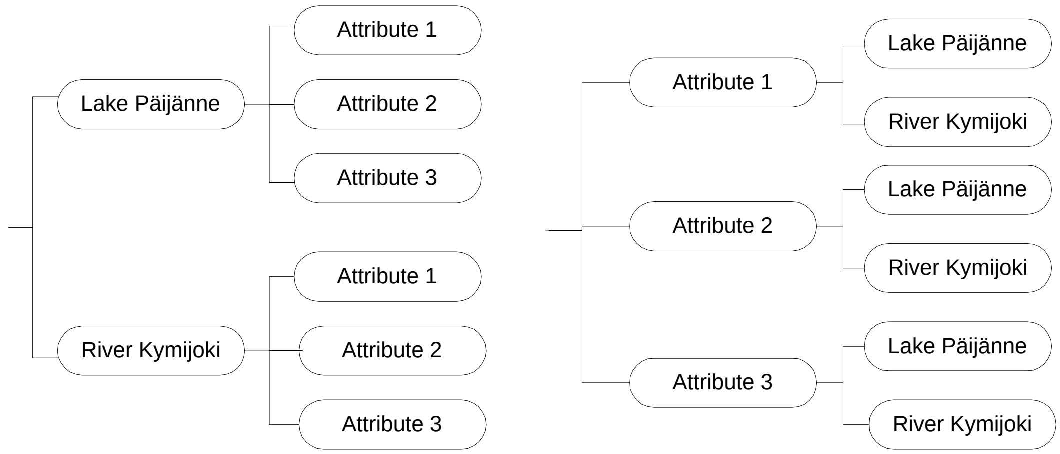

Figure 3. *Swapping attribute levels - two different ways of structuring the Lake Päijänne case.*

H1: The splitting bias does exist

H2: The systematic splitting bias does exist

H3: The splitting bias can be eliminated by taking a course on decision analysis

H4: The splitting bias can be eliminated through good instruction before the weighting

H5: Visual feedback in weighting will eliminate the splitting bias

H6: The magnitude of the splitting bias will be larger when the environmental attributes are split than when the economic attributes are split

H7: In a real problem the real stakeholders will be better able to eliminate the splitting bias

H8: The order of the attribute has an effect on the resulting weights

Table 1. *Research hypotheses of the experiment*

3. Experiment on interactive weighting

3.1 Weighting method

Our idea was to use a weighting method which practitioners would probably use themselves for value tree analysis. The trade-off method (Keeney and Raiffa 1976, Keeney 2002) was not considered as it is not typically applied with hierarchical models. The basic simple multiattribute rating technique (SMART) (Edwards 1977) is perhaps the most commonly used weighting method in practice. In this method one first selects the attribute of least importance and assigns it a certain number of points (typically 10). Then all the other attributes are given points relative to this reference. The SWING method (von Winterfeldt and Edwards 1986) is very similar to SMART. It explicitly refers to and compares the ranges of the attribute outcomes. The most important change from lowest to highest outcome is used as the reference for the comparison and it is typically given 100 points.

In this experiment the weighting was done with the SWING method where one explicitly refers to the attribute ranges. The reason for using SWING rather than SMART was that we assumed that the decision makers would find it easier to compare changes in the ranges with respect to the most important attribute rather than to the least important one as is done in SMART. Weighting was nonhierarchical except for when the so called swapping of attribute levels effect was studied.

3.2 Subjects

The subjects participating in the experiment were 30 students enrolled in a course on decision analysis at the Helsinki University of Technology (HUT) and local citizens and stakeholders from Asikkala, a small town situated in the southern end of Lake Päijänne. The students were volunteers and they were not given any extra credit for participating. In this way we expected to reduce the number of unmotivated test subjects and attract those who would be interested in the task itself.

The students were familiar with both the SWING weighting method and the splitting bias. They had had a lecture on the splitting bias and the range effect a month earlier and they had weighted a job selection problem related to these two biases. They were also familiar with the Lake Päijänne regulation case as it had been used during the course in weighting assignments. The students had been given roles of different stakeholders and they had structured the problem and weighted a value tree with various weighting methods. In the current experiment they were told either to use their personal opinions or the imagined opinions of the stakeholder roles they played. In the Asikkala community we had four citizen groups each consisting of approximately 10 persons and they worked in separate sessions. The first group consisted of students from the local high school. Two of the groups were from a vocational agricultural school and one group consisted of adult stakeholders who represented recreational fishing and boating interests. Additionally there were environmental experts from the Finnish Environment Institute as well as two summer residents. None of these subjects had previous familiarity with decision analytical modelling. In the beginning of the session the Lake Päijänne regulation problem was presented. Then the subjects were taught the basics of value tree analysis including the principles of the SWING weighting method. The last part of the instructions covered the different value trees used in the experiment. The subjects were warned of the possibilities of biases.

3.3 The weighting tasks

The value trees used in the experiment were variants of the basic value tree used for the real Lake Päijänne regulation study. All participants used the computers interactively by themselves but two assistants were present to help with technical problems. The subjects were also explained the results they got throughout the experiment. The sessions lasted approximately 60 - 90 minutes. The subjects participating in the experiment had not been involved in the structuring of the value trees. This caused some discussion as some of them did not think the value tree fully represented their personal view of the regulation problem.

At the time of this experiment the real Lake Päijänne project had not yet been completed, and thus some of the impact data was still missing. Thus the subjects were given illustrative ranges of impacts which were realistic in magnitude for the attributes (Table 2). We did not consider the uncertainties in the outcomes.

As the aim of the experiment was to find out whether biases can be reduced with training and instruction the subjects were told to pay close attention to the structure of the different value trees. They were explained how one easily allocates excessive weight to the split attributes or to impacts categories that have a detailed description compared to more aggregated ones. In each case, they were encouraged to adjust the weights according to the true importance of each attribute range. They were also shown examples how the splitting bias can occur.

Attribute	Worst	Best
Variation of water level	30 cm	0 cm
Recreational fishing	0 %	20 %
Natural reproduction of fish	0 %	30 %
Dense shallow water vegetation	20 %	0 %
Shoreline vegetation	0 %	25 %
Hydroelectricity production	-5 MFIM	0 FIM
Flood damages for agriculture and industry	0.3 MFIM	1 FIM
Flood damages for summer residences	1 MFIM	2 FIM
Transportation	-0.1 MFIM	3 FIM
Commercial fishing	-0.5 MFIM	4 FIM

Table 2. Ranges of the attributes used in the experiment.

In the beginning of the weighting task, the subjects were shown prints of the value trees and the table of the ranges of the attributes (Table 2). The interactive computer interface was built on the Excel spreadsheet program (see Appendix).

The aim was to make the customized interface very easy to use. The points for the attributes could either be entered as numbers or by using a scroll bar with the mouse. The subjects could see both the points they gave as well as the resulting weights as graphical bars.

Each weighting task was presented on an Excel sheet of its own. The experiment was divided into three stages. At the ends of these stages the subjects were shown their weights. The first part compared twig level weights to upper level weights (Figure 4). The second part studied the splitting of attributes in value trees (Figures 5 and 6). The order in which the value trees were presented was randomized. The third part studied the swapping of the attribute levels (Figure 7). The order of the two value trees was randomized.

After the two weighting tasks in the first part the subjects were asked whether the sums of the twig level weights or the upper level weights represented their opinion better.

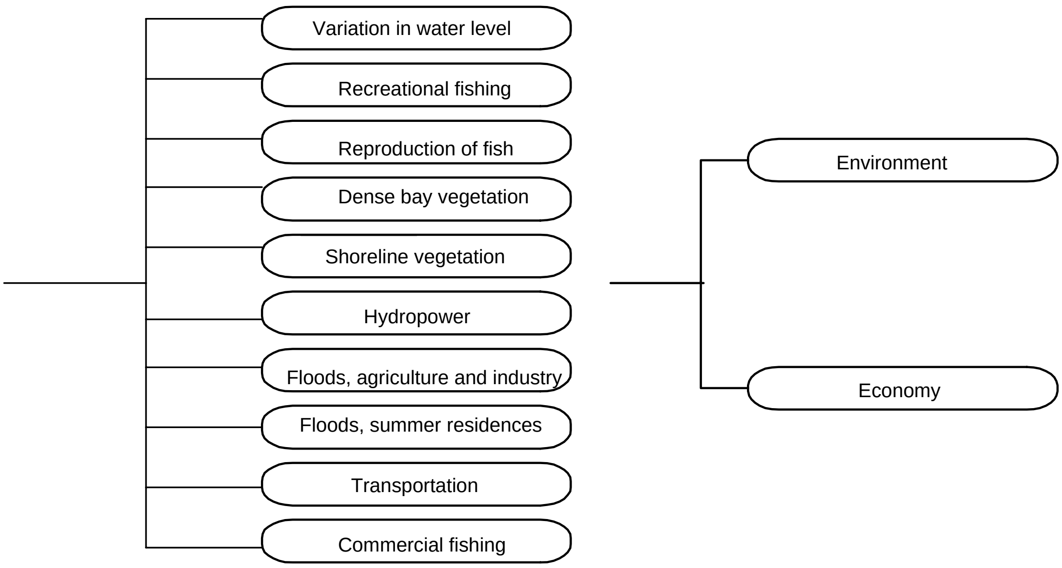

Figure 4. *The value trees in part I: twig level weights vs. upper level weights.*

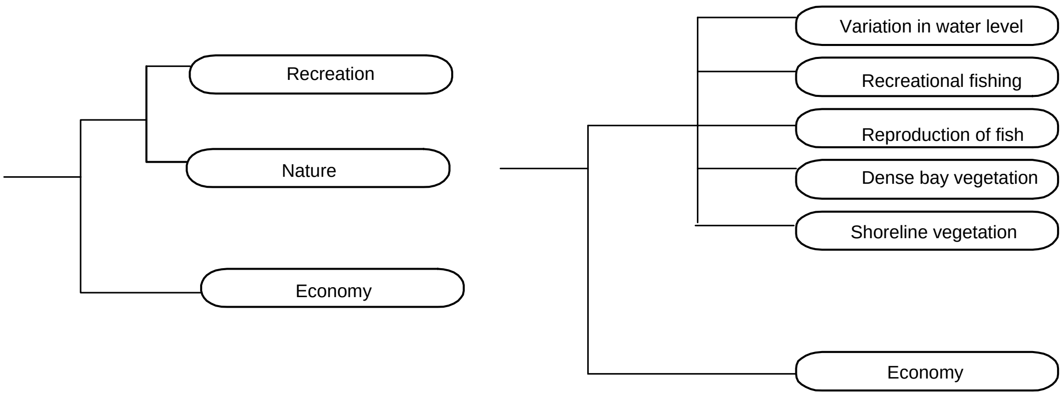

Figure 5. *The environmental attribute is split into two (ENV2) and five (ENV5) sub-attributes while the attribute economy remains aggregated.*

In the second part the subjects weighted four different value trees (ECON5, ECON2, ENV5, ENV2, see Figures 5 and 6). The trees consisted of one attribute in one branch of the value tree and either two or five attributes in the other branch. In order to minimize the effects of learning, the order in which the trees were split was randomized so that the order was either ECON5, ECON2, ENV5, ENV2 or ENV5, ENV2, ECON5, ECON2.

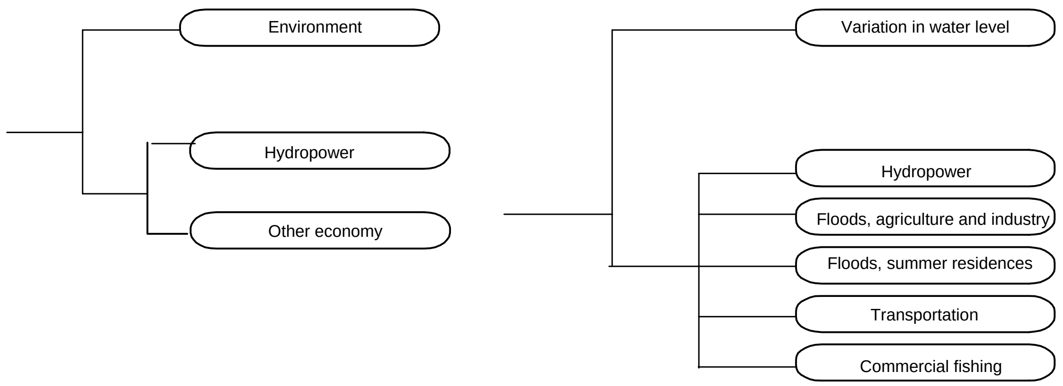

Figure 6. *The economic attribute is split into two (ECON2) and five sub-attributes (ECON5) while the attribute environment remains aggregated.*

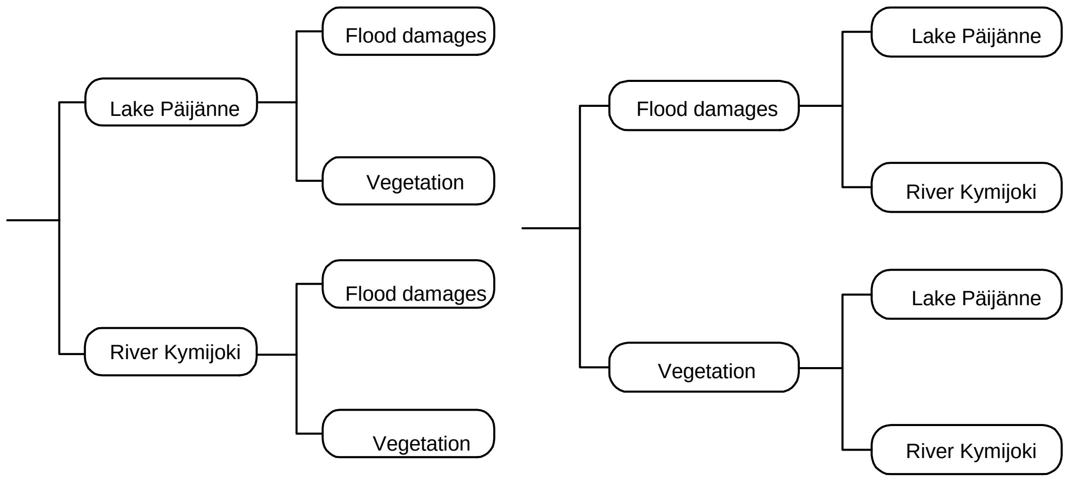

Figure 7. *Swapping the order of the attribute levels.*

3.4 General observations about the experimental setting and interviews

Some of the stakeholders had strong feelings against the regulation in general and it was blamed to be the cause of many undesired impacts on the environment, some of which were in fact not related to the regulation in any way. A common misbelief was also that the power companies are able to decide the regulation policy by themselves.

The expectations about the new regulation policy were also rather unrealistic. Many stakeholders seemed to think that ending the regulation would result in their ideal: A stable water level throughout the year and, at the same time, a significant increase in the fish catch. People also had difficulties in thinking of impacts which were currently not present. For example, many people felt that flood damages were not an important issue as they were under control with the current regulation policy.

The cognitive load of the test seemed to be quite high for many stakeholders. Yet, most of them seemed concentrated when listening to the main impacts of the regulation. Also it seemed that the basic idea of the SWING weighting method was understood quite well. However, already the description of the different structurings of the value trees seemed to confuse many stakeholders.

In the experiment, the number of different value trees was probably too high. The stakeholders had problems especially with the environmental attributes for which the measures were given in percentages. Compared to the students, the majority of the stakeholder subjects were at first also somewhat afraid of using the computers by themselves.

3.5 Summer residents and environmental experts

The experiment was repeated with two summer residents of Lake Päijänne and three environmental experts from the Finnish Environmental Institute. One of these experts also had a summer cottage at Lake Päijänne.

The summer residents were selected for the interviews because we also wanted to have stakeholder subjects with urban professions and a higher level of education. Both summer residence owners chosen for this interview are in managerial positions in their work and are experienced in making decisions in complex settings. Both interviews took place in their own offices with a similar introduction as was given to the stakeholders in the Asikkala community.

Also three employees of the Finnish Environmental Institute completed the experiment. Two of them had worked with the Lake Päijänne regulation case and were familiar with the basics of value tree analysis as well. All three subjects used the computers themselves. Two adopted the roles of some stakeholders to make the weighting easier for them.

4. Results

H1:	The splitting bias does exist: Yes, all stakeholders and some of the students showed a splitting bias in their responses
H2:	The systematic splitting bias does exist: Yes, all except for one of the stakeholders produced this. Only some of the students had a systematic splitting bias in their responses
H3:	The splitting bias can be eliminated by taking a course on decision analysis: Yes, the students who attended a course on decision analysis had clearly less biased responses than the stakeholders
H4:	The splitting bias can be eliminated through good instruction before the weighting: Not supported. Even though the stakeholders were explained the risk of the splitting bias before the weighting, they could not avoid bias.
H5:	Visual feedback in weighting will eliminate the splitting bias: Not supported. The majority of the subjects exhibited a splitting bias even though the interface used provided direct visual feedback.
H6:	The magnitude of the splitting bias will be larger when the environmental attributes are split than when the economic attributes are split: Not supported. The splitting bias was not larger when the environmental attributes were split than when the economic attributes were split.
H7:	In a real decision problem real stakeholders will be better able to eliminate the splitting bias: Not supported. The students were able to avoid the splitting bias better than the stakeholders
H8:	The order of the attribute levels has an effect on the resulting weights: Not supported. The swapping of attribute levels did not produce clear differences in the weights.

Table 3. *Summary of the results for different hypotheses.*

4.1 Splitting bias

The results clearly support hypotheses H1 according to which the splitting bias does exist. In this experiment, the division of an attribute normally increased the weight of that attribute. Only in few cases the attribute weight slightly decreased but this is probably explained by too strong of an attempt to avoid the bias.

Based on the results, the university students can be divided into four groups: (i) those who had practically no bias, (ii) those who had a moderate bias, (iii) those who had a clear bias and (iv) those who over adjusted while trying to avoid the bias. These four groups were similar in size.

Direct visual feedback provided by the computer interface did not prevent the splitting bias. Thus hypotheses H5 does not gain support. The results did not support hypotheses H6 either.

There was no difference in the magnitude of the splitting bias when the environmental attributes were split or when the economic attributes were split. This would suggest that the cause for the splitting bias is not the lack of good and clear measures for the environmental attributes, but some human tendency to give similar weights to all the attributes presented or the inability to separate the importance of attributes of closely similar type.

Splitting	**Students**	**Stakeholders**
ENV - ENV2	8 %	26 %
ENV2 - ENV5	9 %	21 %
ECON - ECON2	7 %	35 %
ECON2 - ECON5	11 %	36 %

Table 4. *Average changes in the weights after the attribute is divided. The difference between the two groups of subjects can be seen clearly.*

In this part of the experiment, the differences between the stakeholders and the students were very clear. Figure 8 shows that, in the average total weights, the splitting bias does exist in both groups and the effect of the bias is always to increase the weight of the divided attribute. For the stakeholders the magnitude of the bias is significantly larger. Thus, the real problem owners were not better in avoiding the bias and hypotheses H7 is not supported.

According to these results Hypotheses H3 is supported. Attending a course on decision analysis did help to eliminate the splitting bias. Hypotheses H4 is not supported because all of the stakeholders had instructions before the experiment but still had a significant splitting bias in their responses.

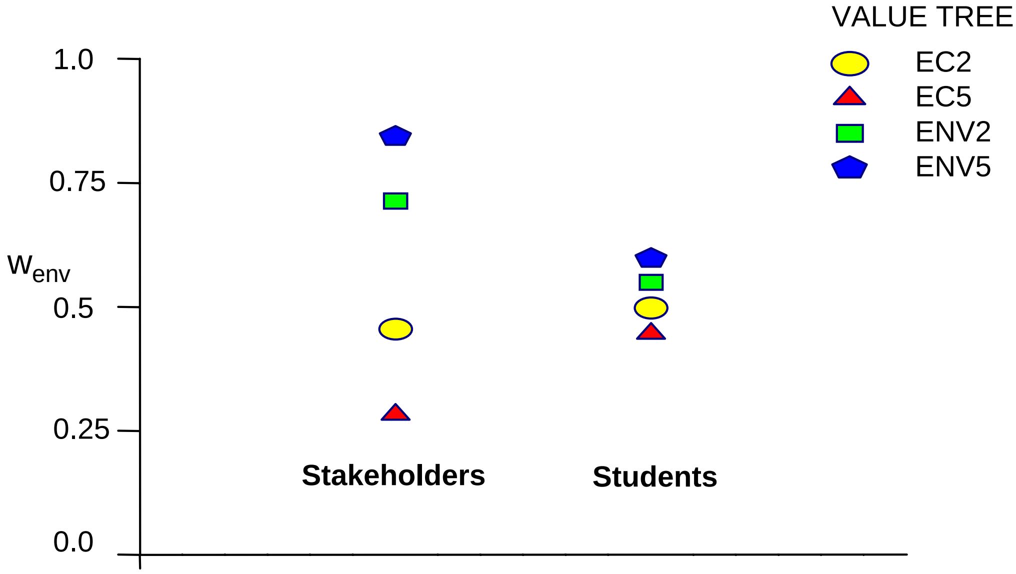

Figure 8. *Group averages of the total weight for the attribute environment (w_{env}) in different value trees. If the results would be nonbiased, the weight would be equal for all value trees. Both the stakeholders and the students exhibit a bias, but for the stakeholders the magnitude of the bias is larger.*

One explanation for this difference is that the students participating in the course on decision analysis were interested in the weighting method and the origin of the biases and could thus better adjust their responses. On the other hand, most of the students felt that they did not have any real opinions about the impacts in the Lake Päijänne regulation problem. So in the weighting they may have concentrated more on the arithmetics of their responses in the weight elicitation than on their actual subjective responses. A number of students even over-adjusted so that the combined attributes rather than the split got a higher weight. Some of the students, indeed, admitted that they were doing arithmetic calculations in order to answer consistently.

The existence of the splitting bias was also investigated with a multifactor analysis of variance. The test statistic was the total weight of the environment. The factors used were the different value trees (ECON2, ECON5, ENV2 and ENV5) and the group (students and stakeholders). Table 5 shows the results of this analysis. The weights did depend significantly on the value tree used (p-value = 0.000) but not on the group (p-value = 0.11).

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob. Level	Power (α=0.05)
A (group)	1	0.1769	0.1769	2.63	0.109	0.126
B(A)	67	4.5051	6.72E-02			
C (value tree)	3	5.1159	1.705	149.59	0.000*	1.000
AC	3	1.7882	0.596	52.29	0.000*	1.000
BC(A)	201	2.2913	1.14E-02			
S	0	0				
Total (Adjusted)	275	14.798				
Total	276					

* Term significant at alpha = 0.05

Table 5. *Analysis of variance table for the total weights of the economy in value trees ECON2, ECON5, ENV2 and ENV5. The weights obtained with various trees differed significantly (p-value = 0.000).*

4.2 Systematic splitting bias

In the first two parts of the experiment the subjects weighted six value trees of different structures. In these value trees the ratios of the numbers of the environmental and economic attributes (n_{env}/n_{econ}) were 1/5, 1/2, 1, 1, 2 and 5. If the decision maker has no splitting bias, the weight ratio of the total weight for the environment and the economy (w_{env}/w_{econ}) remains constant in all six value trees. Whereas, if the decision maker has a systematic splitting bias, then the weight ratio grows in proportion to the ratio of the number of attributes.

In our experiment, 17 out of the 30 students were able to adjust their responses and showed no bias. For them the weight ratios for the environment and the economy (w_{env}/w_{econ}) remained constant in all six value trees. However, the other students did show a systematic splitting bias. They followed a pattern in which the weight ratio grows linearly with the ratios of the number of attributes. Of the 39 stakeholders, 38 produced the systematic splitting bias. One of them had a significant bias which did not follow any systematic pattern. These results support hypotheses H2. The systematic splitting bias does exist.

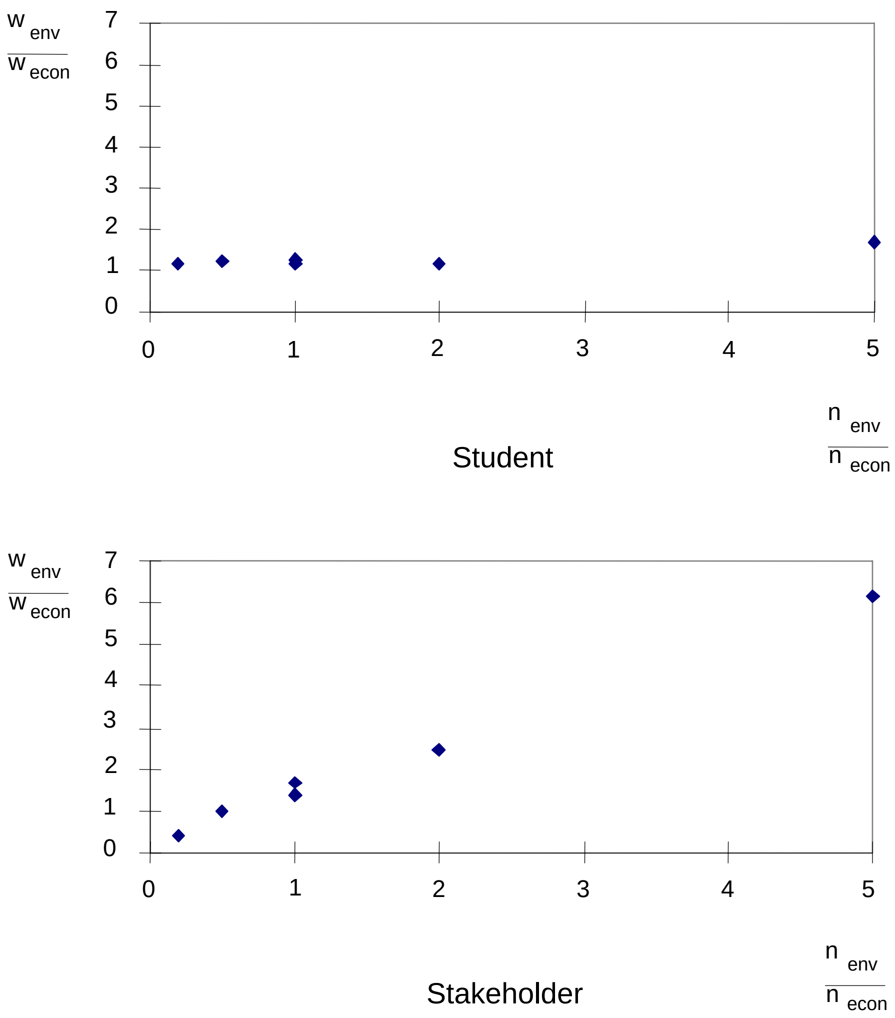

Figure 9. *Examples of the weight ratios as function of the ratios of the number of attributes. For the student the weight ratios are constant and thus there is no bias. For one of the stakeholders, however, the weight ratio increases and this represents a systematic splitting bias.*

4.3 Possible explanations for the splitting bias

In the SWING weighting method the highest number of points given is typically 100. If the decision maker for some reason decides that the smallest possible number to be used is fixed, then the splitting bias is likely to occur automatically. Pöyhönen et al. (2001) describes this in detail.

In the stakeholder group there were some subjects whose range of points given was so limited that the splitting bias could not have been avoided. Figure 10 shows frequency distributions of the smallest points used. In the student group everyone using only even tens produced a clear splitting bias. This can, however, be a sign of weak motivation and lack of interest in adjusting the responses carefully, rather than a real response scale effect.

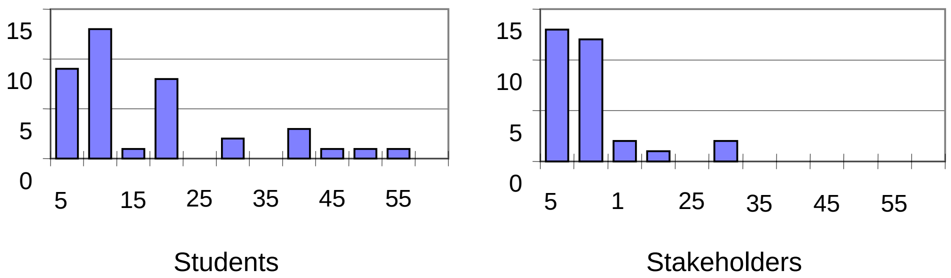

Figure 10. *Frequency distributions for the smallest points used in the weightings in flat weighting, upper level weighting and when weighting value trees ENV5, ENV2, ECON5 and ECON2.*

The limited use of the scale is seen in the fact that all elements in one branch are typically given very similar points. This could reflect the thinking pattern that "all environmental variables are (equally) important". Figure 11 shows an example of this. In the stakeholder group this kind of behavior was common. If the magnitude of the numbers given to split attributes remains the same as the number given to the whole branch then this will always lead to the splitting bias.

What can cause this kind of behavior is an interesting question. One explanation could be that the decision makers trust the problem structure too much, i.e. the value tree presented to them. They may follow some general heuristics that when different attributes are compared against each other, they are always of the same magnitude of importance and could thus be given equal weights. The interested reader could see Gigerenzer and Todd (1999) for a description of different heuristics humans often use. Our observation is similar to that found by Fischoff et al. (1978) in their fault tree experiments. Their subjects seemed to be too confident about the fault tree presented and underestimated the importance of what had been left out. Similarly, in our experiment the subjects may have thought that everything presented in the value tree deserves to get some or a fair amount of importance. In the case of extreme differences in importance, one could also think that the decision makers feel that in the elicitation procedure there is a lower limit (e.g. 10) above which all the numbers given should remain.

The points given by a university student who was able to avoid the splitting bias are shown in Figure 12. One sees that he has extended the range of numbers given so that the smallest number used was 5. When the environmental attributes are split, he is able to give more points to the attribute economy so that the weight ratio of the environment and economy has remained almost constant. It is clear that such adjustments require a lot from the decision maker. In our experiment none of the stakeholders were able to do this. The situation becomes problematic especially with graphical interfaces. Representing the number 100 together with small numbers like 5 is difficult because the difference in the size of the resulting bars becomes so large. This is an important phenomenon to be considered when using decision support software with a limited resolution on the screen.

4.4 Twig weights vs. upper level weights

The twig weights given at the lowest level of the value tree and upper level weights for the environment and the economy (see value trees in Figure12) were also compared. This comparison was done using the total weights for the environment and the economy. In the group average there was no difference at all between these sets of weights. For a large part of the subjects, especially the students, the weights obtained in flat twig level weighting and upper level weighting were very similar. Some of the stakeholders, however, had very significant differences in these weights. One observation was that in these weightings the attribute 'environment' often received more weight than the sum of the subattributes in the environment branch.

The symmetry of the value tree (see Figure 13) may explain why the two weighting methods yield similar weights. As both branches had five subattributes and as the total weights for the environment and the economy were close to 0.5, differences were not very likely to appear. This issue should be studied further with less symmetric value trees.

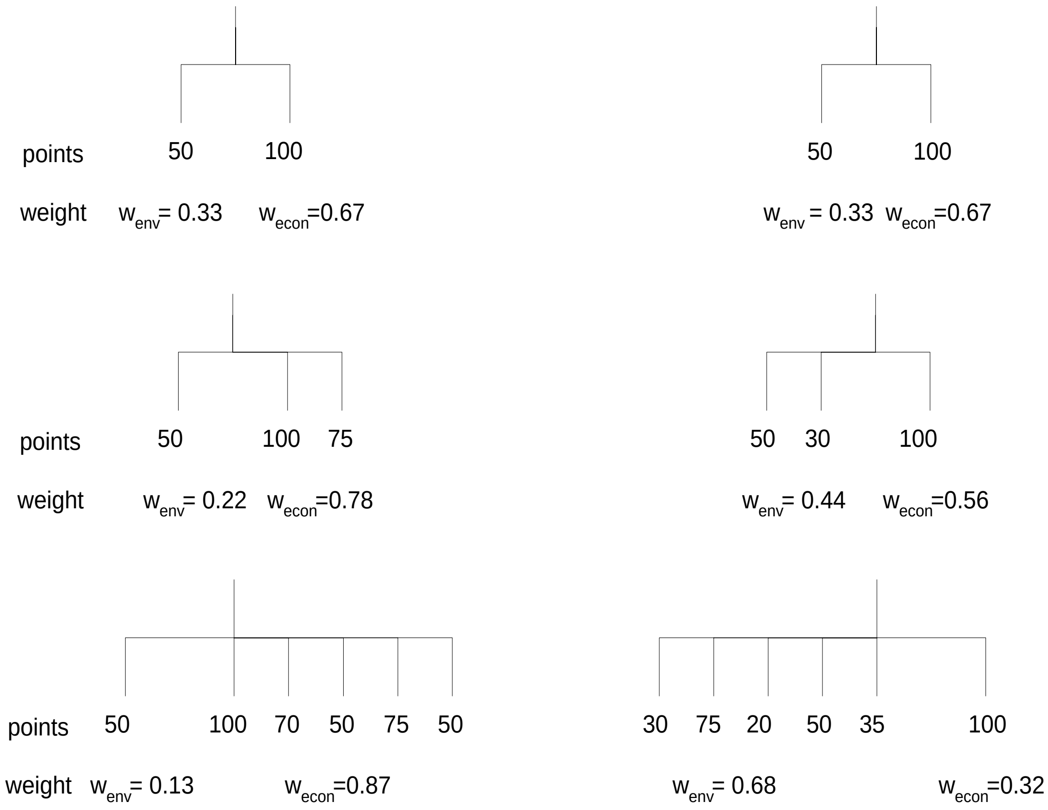

Figure 11. *Points given in SWING weighting by a stakeholder when the main attributes were split. The points given to an attribute and its split subattributes are similar in magnitude. This leads to a typical splitting bias where the weight of the split branch increases.*

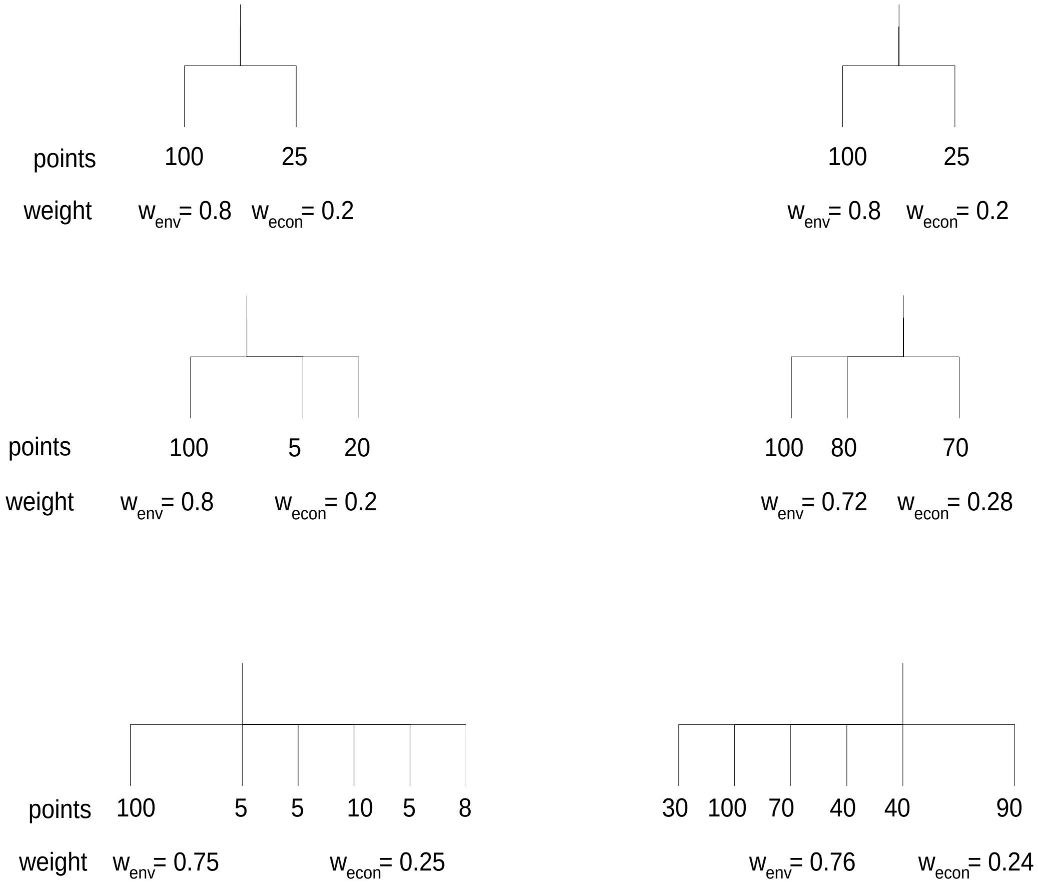

Figure 12. *Points given by one of the students in SWING weighting when the main attributes are split. The adjustment is done by using low points for the split attributes (left) and by increasing the points for the unsplit attribute (right). In this way the splitting bias is nearly avoided.*

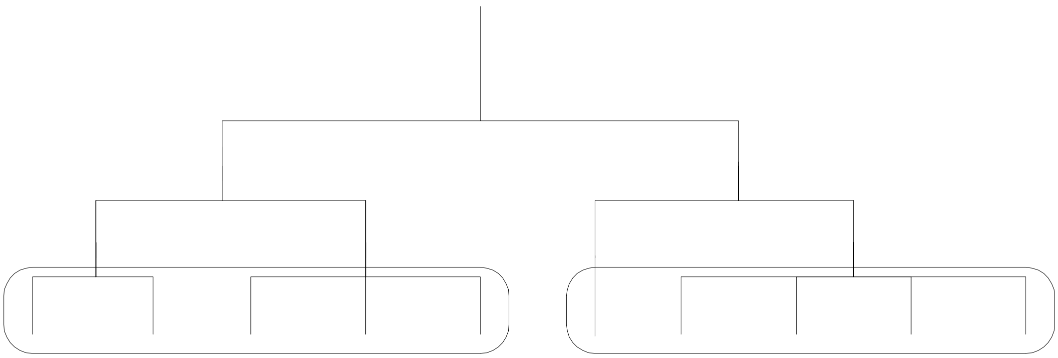

Figure 13. *Value tree used in the swapping of attribute levels experiment. The tree is symmetric in the sense that both branches consist of 5 lowest level elements.*

	Students	Stakeholders
Flat weights	11	19
Upper level weights	7	12

Table 6. *Students' and stakeholders' responses to the question which weights described their opinion better, the sums of the flat weights or the weights given directly at the upper level.*

4.5 Swapping of attribute levels

The effect of the order of the attribute levels was studied by weighting the simplified value trees shown in Figure 7. The results did not change significantly either with the students or the stakeholders. Thus hypotheses H8 cannot be supported.

The effect of the order of the attribute levels was examined through multifactor analysis of variance. The test statistic used was the maximum weight ratios obtained in the two value trees. The p-value for the effect of the value tree was high (p = 0.29) implicating no differences in the weights. There was not a significant difference between the groups (stakeholders or students), the p-value was 0.5.

Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob. Level	Power (α=0.05)
A (group)	1	9.36	9.358	0.46	0.50	0.076
B(A)	61	1237	20.29			
C (value tree)	1	5.562	5.562	1.16	0.29	0.116
AC	1	8.5E-02	8.50E-02	0.02	0.89	0.051
BC(A)	61	292.8	4.8373			
S	0	0				
Total (Adjusted)	125	1545				
Total	126					

* Term significant at alpha = 0.05

Table 7. *Analysis of Variance table for the maximum weight ratios for weights obtained for the two value trees.*

One should note that, in our experiment, the differences in the weights were easily avoided by following a very simple rule: "keep the points the same". For example, Lake Päijänne was always given 100 points and the River Kymijoki 70 points. These same scores could be used on the other level too so that the flood damages always received 100 points and the shoreline vegetation 70 points. This kind of behavior would have eliminated the differences in the weight for the two value trees. Our results were based on simple value trees. However, we cannot say what would have happened if the structure of the value tree had been more complex. This also remains a topic for future research.

After having weighted the value trees, the subjects were also asked which approach they preferred. No generally preferred approach emerged. The regional division first and second got equally much support.

4.6 General observations

The stakeholders had different kinds of difficulties in the weighting than the students. In the interviews many of the stakeholders said that they had difficulties because they felt that all the attributes were very important to them. On the other hand, many of the students did not find any attributes to be of any importance to them. Probably partly because of these two reasons, the average weights of all subjects were divided equally among the ten lowest level attributes. This can be seen in Figures 14 and 15. Compared with the stakeholders, the students had slightly more variations in their weights.

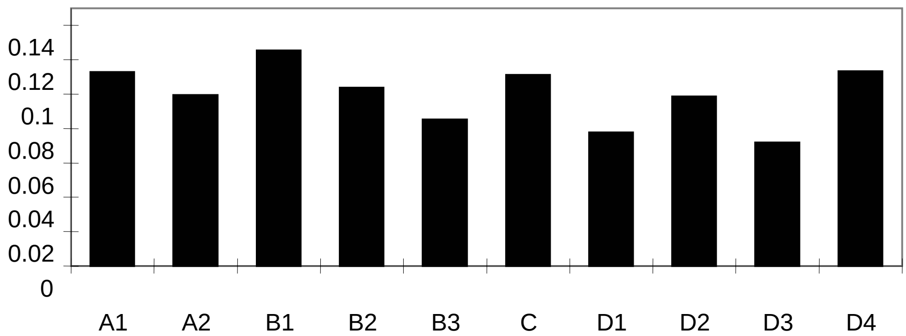

Figure 14. The *average weights of the students for the 10 lowest level attributes obtained in flat weighting.*

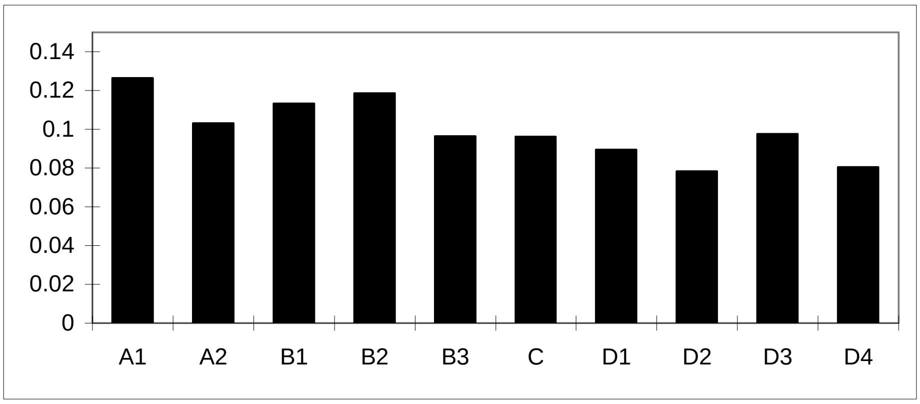

Figure 15. *The average weights of the stakeholders for all ten attributes attained in flat weighting.*

It is interesting to observe that in the economy branch, where everything is measured in money, the weights do not reflect the attribute ranges. There are significant differences in the magnitudes of the attribute ranges but still, for the majority of the subjects, all attributes attain weights that are very similar in magnitude. For example, the attribute hydroelectricity production (C) has a 50 times larger range than the attribute transportation (D3). Yet, in the stakeholders' group averages transportation attains a larger weight than hydroelectricity production.

It is unlikely that the only explanation would be that the subjects did not look at the attribute ranges when they were giving scores to the attributes. It seems to matter a lot who gets the economic benefits or carries the costs.

The results of the summer residents with managerial professions also ended up with a typical splitting bias. Their results were similar to those of the other stakeholders. The two experts of the regulation problem only produced a relatively small bias.

5. Discussion

The results show that there is a serious threat posed by the possibility of the splitting bias. In our experiment the splitting bias could not be eliminated among the stakeholders by instruction or by direct interactive feedback of the results in weights elicitation. Practically all of the stakeholders participating in the experiment demonstrated a clear splitting bias. Moreover, a new result was that their splitting bias was found to be systematic. This means that the weight of an attribute grew in proportion to the number of its subattributes. Of the university students who had participated in a course on decision analysis, only a few had this systematic splitting bias. The majority of them could either avoid the splitting bias or the bias they had was very moderate.

The effects of instruction can be seen in the difference between the results of the students and the stakeholders. However, the results may also reflect the fact that the students did not find the decision problem to be important for themselves. Thus another explanation for their good results can be that they did not try to express their real opinions but concentrated only on avoiding the bias by computing the consistent responses.

The cognitive burden of the experiment was rather heavy for the stakeholders. Thus, an open question remains how much better the stakeholders could have avoided the bias if they had had more time to familiarize themselves with the method and the origins of the cognitive biases. For example, having the instructions and the experiment in separate sessions might have improved the stakeholders' results.

One of the findings of this experiment is that there is a clear difference between results for real stakeholders and those found for university students. In the decision analysis literature most experiments have so far been carried out with university students. Our results would suggest that it is not enough to run experiments with students if one wants to find out all the behavioral phenomena related to decision modeling.

If the decision makers have difficulties in understanding the method correctly then both the role of the facilitator and the computer interface used in the weighting procedure become very important.

The results of this study would support the use of hierarchical models in the problem representation. Hierarchical weighting instead of nonhierarchical weighting allows comparison of only elements that are on a similar level of importance. This can help to eliminate the splitting bias and it also leads to comparing attributes in smaller clusters. In hierarchical weighting the open question is how difficult do the decision makers find it to compare the different branches of the value tree on the upper levels. In these comparisons the interpretation of attribute ranges within a branch becomes rather complicated. Earlier research by Stillwell et al. (1987) suggests that hierarchical weighting produces steeper weight differences. There is clearly a need to study hierarchical and nonhierarchical weighting more thoroughly in real applications.

One possible way to avoid the splitting bias could be the use of some other kind of simple preference elicitation technique such as ranking. People can be prone to giving ordinal information even if the elicitation questions are related to cardinal orders (see e.g. Pöyhönen et al. 2001). The attractiveness of ranking based methods was noted early when the SMARTER technique was proposed (Edwards and Barron 1994, Barron and Barrett 1996). Unfortunately, SMARTER produces the splitting bias automatically by itself because of the way the weights are computed as centroids. However, recently the derivation of weights in methods based on incomplete information has been improved and generalized. The techniques based on Preference Programming (Salo and Hämäläinen 2003) such as PAIRS (Salo and Hämäläinen 1992), PRIME (Salo and Hämäläinen 2001) and RICH (Salo and Punkka 2005) do not use centroid type approaches. However, the price to be paid is that one needs to be satisfied with intervals of overall scores or additional rules have to be used to get point estimates. One could also think that weighting methods which are based on averaging a redundant set of simple pairwise comparisons, such as the AHP (Saaty 1980), would be less prone to the splitting bias. All of these ideas naturally need further studies with experimental tests. Since publicly available software exists (Hämäläinen 2003), it would be natural for practitioners to carry out experiments on these alternative approaches.

Based on the results of this experiment it is difficult to say whether the order of the attribute levels affects the resulting weights. For the majority of the subjects the weights were similar with both value trees. On the other hand, the value trees used in this part of the experiment were very small and perhaps too simple. As the order of the attribute levels is an important question in environmental decision models, this is an area which also needs further research.

The results of our study do indeed raise serious concerns about the reliability of the numerical results of MCDA models especially in environmental decision making. In these problems we usually have many alternative ways to structure the value tree and it is most concerning if each could produce different attribute weights and results. In decision aiding we as the professionals should have high ethics, acknowledge the problems and develop ways of guaranteeing that the help we provide is critically evaluated and correctly understood (Brown 2005). We should not sell our tools blindly as reliable approaches.

However, we should not underestimate the value of the learning process created by the MCDA analysis and DA interviews. This always remains very important but we have to be very careful when a prescriptive approach is taken. The explicit consideration of the ranges of attributes in the decision context opens the big picture and is likely to be a major contributor to the improved communication and learning. These considerations would not usually be done without the explicit attempt to numerically elicit the attribute weights.

It is interesting to note that in the contingent evaluation literature we also have a counterpart of the splitting bias. It is the so called part-whole bias disaggregation effect or embedding effect. For related discussions see Kahneman and Knetsch (1992), Carson and Mitchell (1995) and McDaniels et al. (2003). One possible explanation for the splitting bias could be that we tend to follow a heuristic where all the factors at hand are considered relatively equally important. Then the splitting of any one attribute would naturally increase its importance. Then the heuristic would suggest that the real reason for the splitting is not an academic exercise but a real observation that the original attribute was underweighted as it carries two equally important subattributes.

If the natural tendency is to follow an equal weights heuristic then we should really keep this in mind when structuring the value trees. They should be relatively balanced and hierarchically weighted so that intermediate level aggregated attributes would be used to adjust the weights of lower level attributes.

Acknowledgements

This work was carried out in cooperation with the Finnish Environment Institute (SYKE) in the research program RESTORE 2000 of the Academy of Finland "Waterway Restoration as a Multiobjective Decision Process". We acknowledge the invaluable help of the project leader Mr. Mika Marttunen from the SYKE.

References

Baron, J., (1997), "Biases in the Quantitative Measurement of Values for Public Decisions." Psychological Bulletin, Vol. 122.

Barron, F.H. and Barrett, B.E., (1996), “Decision Quality Using Ranked Attribute Weights.” Management Science, Vol. 42, No. 11, pp. 1515–1523.

Bell, M.L., Hobbs, B.F., Elliott, E.M., Ellis, H. and Robinson Z., (2001), “An Evaluation of Multi-Criteria Methods in Intergrated Assessment of Climate Policy.” Journal of Multi-Criteria Decision Analysis, Vol. 10, pp. 229–256.

Belton, V. and Stewart, T.J., (2002), Multiple Criteria Decision Analysis – An Integrated Approach. Kluwer Academic Publishers, Massachusetts.

Bojórquez-Tapia, L.A., Sánchez-Colon, S., Florez, A. (2005), “Building Consensus in Environmental Impact Assessment Through Multicriteria Modeling and Sensitivity Analysis.” Environmental Management, Vol. 36, No. 3, pp. 469-481.

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

Brown, R., (2005), “The Operation Was a Success but the Patient Dies: Aider Priorities Influence Decision Aid Usefulness.” Interfaces, Vol. 35, No. 6, pp. 511-521.

Carson, R. and Mitchell, R., (1995), “Sequencing and Nesting in Contingent Valuation Surveys.” Journal of Environmental Economics and Management, Vol. 28 (2), pp. 155-173.

Edwards, W., (1977), "How to Use Multiattribute Utility Measurement for Social Decision Making." IEEE Transactions on Systems, Man and Cybernetics, SMC-7, pp. 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. Vol 60, pp. 306–325.

Fischoff, B., Slovic, P. and Lichtenstein, S., (1978), "Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation." Journal of

Experimental Psychology: Human Perception and Performance, Vol. 4, pp. 330–334.

Fisher, G., (1995), "Range Sensitivities of Attribute Weights in Multiattribute Value Models." Organizational Behavior and Human Decision processes. Vol. 62, No. 3.

Gigerenzer, G. and Todd, P.M., (1999), "Simple Heuristics That Make Us Smart." Oxford University Press, New York.

Gregory, R., McDaniels, T. and Fields, D., (2001), “Decision Aiding , Non Dispute Resolution: Creating Insights Through Structured Environmental Decisions.” Journal of Policy Analysis and Management. Vol. 20, pp. 415–432.

Hayashi, K., (2000), “Multi-Criteria Analysis for Agricultural Resource Management.” European Journal of Operational Research. Vol. 122, pp. 486–500.

Hobbs, B.F. and Meier, P., (2000), Energy Decisions and the Environment: A Guide to the Use of Multicriteria Methods. Kluwer Academic Publishers, Boston.

Hämäläinen, R.P., (1992), "Decision Analysis Makes its Way into Environmental Policy in Finland." OR/MS Today, Vol. 19, No. 3, pp. 40–43.

Hämäläinen, R.P., (2003), “Decisionarium – Aiding Decisions, Negotiating and Collecting Opinions on the Web.” Journal of Multicriteria Decision Analysis, Vol. 12, No. 2–3, pp. 101–110.

Hämäläinen, R.P., Kettunen, E. Marttunen, M. and Ehtamo, H., (2001), "Evaluating a framework for multi-stakeholder decision support in water resources management." Group Decision and Negotiation, Vol. 10, No. 4, pp. 331– 353.

Hämäläinen, R.P. and Salo, A., (1997), "The Issue is Understanding the Weights." Journal of Multi-Criteria Decision Analysis, Vol. 6, pp. 340–343.

Janssen, R., (2001), “On the Use of Multi-Criteria Analysis in Environmental Impact Assessment in the Netherlands.” Journal of Multi-Criteria Decision Analysis, Vol. 10, pp. 101–109.

Kahneman, D. and Knetsch, J., (1992), “Valuing Public Goods: the Purchase of Moral Satisfaction.” Journal of Environmental Economics and Management, Vol. 22, pp. 57-70.

Kangas, J., Kangas, A., Leskinen, P. and Pykäläinen, J., (2001), ”MCDM Methods in Strategic Planning of Forestry on State-Owned Lands in Finland: Applications and Experiences.” Journal of Multicriteria Decision Analysis, Vol. 10, pp. 257–271.

Keeney, R.L., (2002), "Common mistakes in making value trade-offs." Operations Research, Vol. 50, No. 6, pp. 935–945.

Keeney, R. and Raiffa, H., (1976), "Decisions with Multiple Objectives: Preferences and Value Tradeoffs." John Wiley & Sons, New York.

Kiker, G.A., Bridges, T.S., Varghese, A., Seager, T.P. and Linkov, I., (2005), ”Application of Multicriteria Decision Analysis in Environmental Decision Making.” Integrated Environmental Assessment and Management, Vol. 1, No. 2, pp. 95–108.

Leskinen, P. and Kangas, J., (2005), ”Multi-Criteria Natural Resource Management with Preferentially Dependent Decision Criteria.” Journal of Environmental Management, Vol. 77, pp. 244-251.

Marttunen, M. and Hämäläinen, R.P., (1995), "Decision Analysis Interviews in Environmental Impact Assessment." European Journal of Operational Research, Vol. 87, pp. 551–563.

Marttunen, M. and Hämäläinen, R.P., (2006), "Decision Analysis Interviews in the Collaborative Management of a Large Regulated Water Course." Manuscript.

McDaniels, T.L., Gregory, R., Arvai, J. and Chuenpagdee, R., (2003), “Decision Structuring to Alleviate Embedding in Environmental Valuation.” Ecological Economics, Vol. 46, pp. 33-46.

Miettinen, P. and Hämäläinen, R.P., (1997), "How to Benefit from Decision Analysis in Environmental Life Cycle Assessment." European Journal of Operational Research, Vol. 102, No. 2, pp. 279–294.

Mustajoki, J., Hämäläinen, R.P. and Marttunen, M., (2004), ”Participatory Multicriteria Decision Analysis with Web-HIPRE: A Case of Lake Regulation Policy.”, Environmental Modelling and Software, Vol. 19, No. 6, pp. 537–547.

von Nitzsch, R. and Weber, M., (1993), "The Effect of Attribute Ranges on Weights in Multiattribute Utility Measurement." Management Science, Vol. 39, No. 8.

Pöyhönen, M. and Hämäläinen, R.P., (1998), "Notes on the Weighting Biases in Value Trees." Journal of Behavioral Decision Making, Vol. 11, pp. 139–150.

Pöyhönen, M. and Hämäläinen, R.P., (2000) "There is hope in attribute weighting." Journal of Information Systems and Operational Research (INFOR), Vol. 38, No. 3, pp. 272-282.

Pöyhönen, M. and Hämäläinen, R.P., (2001), "On the Convergence of Multiattribute Weighting Methods." European Journal of Operational Research, Vol. 129, No. 3, pp. 569–585.

Pöyhönen, M., Vrolijk, H. C. J and Hämäläinen, R.P., (2001), "Behavioral and Procedural Consequences of Structural Variation in Value Trees." European Journal of Operational Research, Vol. 134, No. 1, pp. 218–227.

Rauschmayer, F., (2001), “Reflections on Ethics and MCA in Environmental Decisions.” Journal of Multi-Criteria Decision Analysis, Vol. 10, pp. 65-74.

Saaty, T.L., (1980), “The Analytic Hierarchy Process.” New York, NY, McGraw-Hill, Inc.

Salo, A. and Hämäläinen, R.P., (1992), “Preference assessment by imprecise ratio statements.” Operations Research, Vol. 40, No. 6, pp. 1053–1061.

Salo, A. and Hämäläinen, R.P., (2001), “Preference Ratios in Multiattribute Evaluation (PRIME) - Elicitation and Decision Procedures under Incomplete Information.” IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 31, No. 6, pp. 533–545.

Salo, A. and Hämäläinen, R.P. (2003), “Preference Programming.” Manuscript. (Downloadable at http://www.sal.hut.fi/Publications/pdf-files/msal03b.pdf)

Salo, A. and Punkka, A. (2005), “Rank Inclusion in Criteria Hierarchies.” European Journal of Operational Research, Vol. 163, No.2, pp. 338–356.

Schmold, D.L., Kangas, J., Mendoza, G.A. and Pesonen, M., (2001), ”The Analytic Hierarchy Process in Natural Resource and Environmental Decision Making.” Amsterdam, The Netherlands, Kluwer.

Stillwell, W.G., von Winterfeldt, D. and John, R.S., (1987), "Comparing Hierarchical and Nonhierarchical Weighting methods for Eliciting Multiattribute Value Models." Management Science, Vol. 33, pp. 442–450.

von Winterfeldt, D. and Edwards, W., (1986), "Decision Analysis and Behavioral Research." Cambridge University Press, New York.

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

Weber, M. and Borcherding, K., (1993), "Behavioural Influences on Weight Judgments in Multiattribute Decision Making." European Journal of Operational Research, Vol. 67, pp. 1–12.

Wenstøp, F. and Seip, K., (2001), “Legitimacy and Quality of Multi-Criteria Environmental Policy Analysis: a Metaanalysis of Five MCE Studies in Norway.” Journal of Multi-Criteria Decision Analysis, Vol. 10, pp. 53-64.

Appendix

Screenshots of the computer interface used in the experiments

Tasks: First choose the most important change. Assign it 100 points. Give points to other changes reflecting their relative importance with respect to the most important one.

COMPARISON A

	Points	Weight
Variation of water level	43	0.08
Recreational fishing	65	0.11
Reproduction of fish	54	0.09
Dense bay vegetation	76	0.13
Shoreline vegetation	100	0.17
Hydroelectricity production	45	0.08
Flood damages, industry and	23	0.04
Flood damages, summer res	56	0.1
Transportation	87	0.15
Commercial fishing	23	0.04

COMPARISON B

	Points	Weight
Environment	100	0.61
Variation of water level		
Recreational fishing		
Reproduction of fish		
Dense bay vegetation		
Shoreline vegetation		
Economy	65	0.39
Hydroelectricity production		
Flood damages, industry and		
Flood damages, summer res		
Transportation		
Commercial fishing		