PRIORITIZING STAKEHOLDER CONCERNS IN ENVIRONMENTAL RISK MANAGEMENT

by

Roberto Accorsi

Dottore in Ingegneria Nucleare
Politecnico di Milano, 1996

Submitted to the Department of Nuclear Engineering
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN NUCLEAR ENGINEERING

AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 1998

© Massachusetts Institute of Technology, 1998. All rights reserved.

Signature of Author

Department of Nuclear Engineering
April 13, 1997

Certified by
Professor George E. Apostolakis
Department of Nuclear Engineering
Thesis Advisor

Certified by
Professor Kent F. Hansen
Department of Nuclear Engineering
Thesis Reader

Accepted by
Professor Lawrence M. Lidsky
Chairman, Department Committee on Graduate Students
ABSTRACT

The management of environmental programs affects a great variety of issues and involves several stakeholders with diverse, and often conflicting, concerns. Any methodology proposed for environmental decision making should, then, be capable of integrating technical and value-laden objectives in a framework which ensures process transparency. Furthermore, only a methodology simple and easy to implement can achieve public involvement at every level, and, with it, achieve a thoroughly informed decision.

Classic decision making techniques have provided the tools to organize decision problems in a sound logical structure. Unfortunately, actual applications have shown that, when it comes to the practical quantification of stakeholder values, these tools may be too burdensome to apply. This is particularly so in cases where non-technical public is a fundamental part of the decision process.

In this thesis, the focus is on a methodology developed to assess stakeholder preferences regarding a number of objectives of environmental restoration activities, such as the minimization of costs and the impact on human health and safety, and on its combination with the structural approach provided by classic methods. Stakeholder input is used to produce a first set of relative weights using the Analytic Hierarchy Process in combination with value trees. The set of preferences so obtained, as well as possible inconsistencies in their assessments, are discussed with the stakeholders to revise the weights and reduce inconsistency. The stakeholders always have the final word regarding the relative weights.

The prioritization procedure is presented through a case study. Insights gained from this real life application are also presented.
BIographical Note

Roberto Accorsi was born in Milano, Italy, on October 2, 1971.

In September 1989 he was invited by the "Scuola Normale di Pisa" for one week as one of 100 students selected all over Italy on the basis of their achievements. In July 1990 he graduated from the Liceo Scientifico Statale, San Donato, Milano with the score of 60/60.

In July 1996 he received with honors his Laurea degree in Nuclear Engineering from the Politecnico di Milano. His thesis topic was Control of an Industrial Plant with Neural Networks Trained during Real Operation.

Roberto Accorsi has joined the Nuclear Engineering Department of the Massachusetts Institute of Technology as research staff member in September 1996. Since February 1997 he is a graduate student in the Nuclear Engineering Department.
ACKNOWLEDGMENTS

It is not for formal courtesy only that I would like to thank first my thesis advisor, Professor George Apostolakis. Beyond his scientific guidance he offered me much more than I expected from a human perspective. He has gone out of his way to make possible my MIT adventure and has always been interested in making it part of a wider American experience. It is difficult for me to show my gratitude for the trust he had in me. I will always remember our discussions in “the office”, where this project has seen its actual breakthroughs.

I would like to thank the thesis reader, Professor Kent Hansen, for his counsel and his careful review of this work.

It is impossible to find adequate words for Enrico, who has shared with me almost every day of my stay at MIT despite the distance. From a scientific point of view it is impossible to distinguish our contributions to this project. From a personal point of view, I am still surprised by how close we are to each other. Probably we have, so far, spent more time on the email than we have actually spent together! This is not the first thesis in which he features. But this time I can thank him for not being the “creative author” of its soundtrack. It was a real improvement. By the way, I have talked to that Saint-Etienne guy. He said for that Michel it’s ok. Deal?

I would like to thank in the person of Clare Egan all the NED personnel for the patience, care and competence with which they treat bugging students with perennial questions and most disparate needs. You and your outstanding daily dedication do make the difference.

I wish to thank my colleagues in this project for making this work possible. They are: Stephie Jennings, Scott Ploger, Stephanie Weisband and Erika Zachman of Advanced Sciences, Inc.; Tito Bonano, Kelly Peil, David Smith of Beta Corporation International; Susan Pickett of MIT; Abbas Ghassemi of New Mexico State University; Pat Salter and Matt Kozak of QuantiSci, Inc. I really enjoyed the atmosphere of the crazy weeks before the workshops.

Many thanks to Scott, my first American landlord, who was the first to care about my well being in a foreign country.

Chris, you didn’t want to write your own acknowledgment. You know I will never be able to thank you the way you deserve in English. Here are some blank lines for you.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>3</td>
</tr>
<tr>
<td>BIOGRAPHICAL NOTE</td>
<td>4</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>5</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>6</td>
</tr>
<tr>
<td>1 INTRODUCTION AND PROBLEM IDENTIFICATION</td>
<td>9</td>
</tr>
<tr>
<td>1.1 The advantages of a quantitative assessment of preferences</td>
<td>10</td>
</tr>
<tr>
<td>1.2 An analytic-deliberative methodology</td>
<td>13</td>
</tr>
<tr>
<td>1.3 The implementation</td>
<td>14</td>
</tr>
<tr>
<td>1.4 Thesis focus and outline</td>
<td>16</td>
</tr>
<tr>
<td>2 THE DECISION CONTEXT</td>
<td>19</td>
</tr>
<tr>
<td>2.1 Site description</td>
<td>19</td>
</tr>
<tr>
<td>2.2 The stakeholders</td>
<td>20</td>
</tr>
<tr>
<td>2.3 Remedial action alternatives</td>
<td>21</td>
</tr>
<tr>
<td>2.4 Structuring the decision: The Value Trees</td>
<td>22</td>
</tr>
<tr>
<td>2.5 The risk assessments</td>
<td>28</td>
</tr>
<tr>
<td>3 PRIORITIZING OBJECTIVES FOR INTEGRATION</td>
<td>31</td>
</tr>
<tr>
<td>3.1 Mathematical formulation</td>
<td>32</td>
</tr>
<tr>
<td>3.2 Different prioritization tools</td>
<td>34</td>
</tr>
<tr>
<td>3.2.1 Direct assessment</td>
<td>36</td>
</tr>
</tbody>
</table>
Like all real world decisions, Environmental Management (EM) problems involve a great variety of concerns, including cost, health and programmatic risks and impacts on cultural resources. A major problem with these decision dimensions is that they have different units and are difficult to define. Yet, all these aspects must be accounted for in an efficient decision making process, since failure to address any of them is likely to result in controversial decisions.

All is not lost, however, since decision theory provides a logical way of combining non-commensurate consequences of decisions by introducing utilities that serve as a common unit of measure. First, utility functions allow the decision maker to express different degrees of preference toward the impacts, taken one at a time. Second, these single-attribute utilities can be integrated in an indicator of the overall desirability of the different decision options under consideration.

A second problem makes the straightforward use of decision theory impossible. In fact, there is usually a number of different individuals or groups that are very interested in influencing the decision, even though they are not the decision makers. These individuals or groups are generally referred to as stakeholders, since they have a "stake" in the decision. These might include various government agencies and citizen groups. It is widely accepted that the stakeholders must be involved in the decision-making process from the beginning (NRC, 1996). Consequently, the decision maker must deal with a diverse audience of technical and non-technical people, as well as with the multitude of their technical and value-laden quantitative and qualitative concerns. Each stakeholder has his/her own set of objectives that s/he wishes to achieve. Although most of these objectives are shared by other stakeholders, they may not be valued the same way. For example, the costs of the proposed decision alternatives are usually viewed differently by
the site owner and the other stakeholders. It is evident, therefore, that quantitative
methods are needed to elicit stakeholder priorities and preferences.

These methodologies cannot restrict themselves to identifying and organizing the
concerns and objectives in a simple, yet all-encompassing, logic framework. Rather, they
must recognize and quantify the bargaining intrinsically present in a multi-objective
decision. In other words, the analysis can not stop at a qualitative ranking, but must reach
a quantitative assessment of the priorities of the identified objectives. In the context of
wide public involvement outlined above, it is furthermore clear that the procedure used to
elicit priorities and preferences must be accessible to all kinds of stakeholders involved
and flexible enough to account for the imprecision and inconsistencies inherent in human
subjective judgments.

1.1 The advantages of a quantitative assessment of preferences

A complete methodology for interactive decision making involving multiple
stakeholders tackling the problem on these grounds has been described by Zio (1997).
The first step is problem definition. The task is to clearly specify the ultimate goal, the
proposed alternatives, the potential stakeholders, their influence on the decision, and the
actual decision makers. We can assume that even if it can be a very delicate matter, it is
always possible to compile comprehensive lists of stakeholders and alternatives. For the
time being, one need not be concerned as to the actual plausibility of the alternatives nor
the accountability of the stakeholders. It will be sufficient to include all the individuals or
parties interested and the alternatives suggested. Since the rejection of alternatives is
often a controversial issue, unrealistic or otherwise non practicable possibilities have to
be screened with the consensus of all stakeholders. To this end, preliminary risk
assessments providing a rough estimate can be sufficient

1 This phase is extremely important, as it is inconceivable to carry out complete risk assessments for
a great number of alternatives. The criteria of screening of the alternatives can be found in Zio (1997).
The objective of every stakeholder is to see eventually implemented the alternative they prefer. It is not uncommon to find that most stakeholders express preferences towards the various alternatives at once. These preferences may be based on a deep knowledge of sound evidence as well as on deceiving arbitrary assumptions. In either case, it is difficult to believe that these “instinctive” rankings of the alternatives are the result of a thorough decision process which included a careful and quantitative evaluation of all the factors involved. It is widely recognized that both members of the public and experts do not usually make rational decisions under risk especially if the decision making process does not expressly adopt risk management concepts. Moreover, even if the decision were not affected by any uncertainty, it would still be very difficult to assess the total impact of both its costs and its benefits. And, finally, even if this were not the case, it still could be hard to tell if the benefits are worth the costs.

As obvious, it is of the uppermost importance that all the stakeholders be exhaustively informed about all the factors and the risks relevant to the decision problem. If this information is made quantitative rather than qualitative, it can be coupled with a rational and methodical quantification of the preferences. These two results can be combined to obtain, for each stakeholder, a ranking of the alternatives on a more objective basis. A quantitative analysis can show why and to what extent some alternatives may not be as desirable as they seem, or that the second or third preferred choices may be very good substitutes of the best alternative. Moreover, it could turn out that the differences between the alternatives are overshadowed by the uncertainties involved and, then, it makes little sense to define a marked standing of closely ranked alternatives. Insights such as this have a great potential in avoiding the typical stakeholder attitude of strenuous defense of the first ranked alternative.

The quantification of overall preferences has the further advantage of decomposing the problem in its fundamental elements. These, in their turn, can be ranked, leading to a quantitative indication of the decision drivers. In concrete, one can see from the contribution of the single elements to the total desirability of an alternative what are the major strengths of a high-ranked alternative or the main weaknesses of a low-ranked one.
The advantages of quantifying preferences are clear even from a behavioral point of view. This is particularly true if the task is carried out in the course of a collective debate. The attempt to quantify the impacts is expected to keep the discussion focused on the real problem rather than on abstract principles. For instance, a typical problem is to evaluate if action should be taken at a contaminated site. If the problem is posed in these loose terms the answer is obvious. The real problem starts when, say, it is found that the contamination is not likely, under all no action scenarios, to have any health impacts. It may seem a paradox, but in our case study it was found that the only likely health hazard posed by the contaminant was worker exposure if action was to be taken. This may seem a trivial result, but the reader may rest assured that in the emotionally charged public arena it is not, especially if probabilities are not made quantitative and communicated with simple tools\(^2\) to all stakeholders. Being concrete and case specific casts a completely different light on the debate and fosters a different attitude, more directed towards a real solution for the problem at hand: in this context emotional responses tend to leave room to more rational and practical considerations.

Once that every stakeholder has reached a more or less definite ranking, but, above all, s/he feels comfortable with the understanding of each alternative’s implications, the floor is open for a forum in which all the stakeholders debate their recommendation to the decision maker. In this phase a leading role is that of the mediator. Besides ensuring that the discussion is fair, the mediator can drive the group towards a unanimous recommendation. The observation that a unanimous recommendation is more likely to be a strong signal to the decision maker can be important in giving the group a common goal well worth of sacrifices. This stimulates trading among the stakeholders and the typical winner / loser attitude turns into an atmosphere of cooperation. Moreover, the group creativity is encouraged. For instance, as alternatives were decomposed to their fundamental elements during their prioritization, the stakeholders can give up some aspects of an alternative (the ones resulting in a modest utility contribution) to focus on

\[^2\) This communication aspect was a relevant part of the project, as it can be inferred by the final report title (Advanced Sciences, Inc., et al., 1997), to which we refer the interested reader.
the elements that really make it desirable to them. New alternatives can be combined as a mix of parts of different alternatives, each addressing a particular need: the optimal solution could very well be an hybrid of alternatives. A way of achieving this could be a step by step process in which unanimous conclusions are being sought. These can be very obvious ("Something has to be done") or in negative form ("That is not an option") but should attempt to reach little by little a definitive conclusion.

Of course, the introduction of quantitative methodologies is not smooth and easy. From a theoretical point of view, quantitative models can certainly capture all the aspects of a decision. The question is if this is practically achievable. A first problem is, then, if what is inevitably neglected was really negligible. A second problem concerns the fine tuning properties of the model employed and its capability of capturing the perceptions of different stakeholders. Finally, from a very radical perspective one may argue that any quantification is far from fostering agreement, since it can be itself the source of bitter controversy about the data.

All these are open questions to which it can only be replied that every case, with all its peculiarities, is a different case. Perhaps one of the most amazing surprises we found in this project were the words of one of the stakeholders, a late pensioner. Despite showing a certain distrust in the mediator, she candidly admitted that, when it came to technical issues, she had no problem in leaving the quantification not only of contaminant concentrations and health indexes, but even of her personal preferences to the analysts: not only she trusted their data, but their neutrality as well. After her words, many stakeholders, especially the non-technical ones, went along and adopted the analysts' weighting of some technical performance measures.

1.2 An analytic-deliberative methodology

The process leading to the final decision can be divided in two parts. The first, culminating in the assessment of the rankings and their substantiation, is the so called "analytic" part: it is analytic in that the problem is decomposed to its elementary
components which are then integrated in the rankings. The second, concerned with the
debate on the recommendation, is known as the “deliberative” part. It must be stressed,
however, that these two parts are not purely sequential. The entire process is, rather, an
iterative feedback process in which analysis, with its results, directs deliberation, and
deliberation, in its turn, guides analysis, focusing its efforts where needed. This opens a
scenario of long time scales. In this perspective, it is imperative to look for tools with a
potential of optimizing analysis effort.

The concept of “analytic-deliberative process” is becoming increasingly important
on the decision making scene. In 1993 the Department of Energy (DoE) invited the
National Research Council (NRC) of the National Academy of Sciences to investigate the
potential of risk assessment and risk management tools and their relevance to the decision
making process in DoE’s EM program. The NRC concluded that the use of such
approaches is both feasible and very pertinent (NRC, 1994). In a second study (NRC,
1996) the development of these recommendations lead to the conclusion that it is
appropriate to implement risk-informed decision making in the more general framework
of an analytic-deliberative process.

1.3 The implementation

The challenges outlined above clearly show that the completion of a decision
making methodology applicable in the context of EM is an extremely complex task that
must address a broad variety of concerns. Moreover, in a field where the subjectivity
involved in the decision making process is by itself a fundamental issue, it is impossible
to assume a priori that the application of decision making theories will be both
straightforward and able to capture all the aspects relevant to the decision. These
considerations make it clear that the potential of the methodology can be proved only by
implementing it in a real case.

As the principles of the methodology outlined above seemed to meet the
prerequisites of the DoE Notice of Program Interest for improvements in risk assessment
and communication, Advanced Sciences, Inc., in association with Beta Corporation International, QuantiSci, Inc., the New Mexico State University, and the Massachusetts Institute of Technology (the Team of Analysts) submitted a proposal to DoE. The methodology development and implementation program was funded under Cooperative Agreement DE-FC01-95EW55088. The project consisted of several stages and covered a time span of about two years. Its main steps were:

1) Site specific preparations
   - Site selection and preliminary information
   - Stakeholder identification and preliminary involvement
   - Identification of site-specific concerns and goals
   - Formulation of the alternatives (Remedial Action Alternatives, RAAs)

2) Alternative evaluation
   - Identification of the analysis tools (e.g. analytical and computer models) needed to complete the risk assessments for the evaluation of the impacts of each RAA
   - Development of influence diagrams to ensure consistency and organize risk assessment results
   - Risk assessments

3) Develop decision structures
   - Development of a Decision Hierarchy in close cooperation with the stakeholders to integrate risk assessments with stakeholder values
   - Elicitation of preferences
   - Calculation of rankings
4) Group deliberation for consensus

- Discussion of individual rankings to identify drivers and relevant uncertainties
- Identify group agreements and disagreements
- Formulation and discussion of tentative conclusions
- Final recommendation

These tasks were carried out by the Team in close cooperation with the stakeholders. Their input was obtained both in the form of personal interviews and in the course of workshops attended by all the stakeholders as well as by Team members. A complete description of methodology implementation and results can be found in the Final Report submitted to DoE (Advanced Sciences, Inc., et al., 1997).

1.4 Thesis focus and outline

The focus of this thesis is on the elicitation of preferences and the prioritization of the objectives present in the decision hierarchy. The main concern was to build an elicitation procedure both reliable and practical in all its applicative details. Accordingly, its test in a real life case was of the greatest importance. Fundamental part of the thesis are the practical insights derived from workshop experience. These insights are presented to show how the use of assessment tools such as AHP can dissect stakeholder input, leading the analyst beyond the bare value of the preference weights and focusing further iterations on the problem areas.

Chapter 2 opens with some background information, such as the site description, the list of the stakeholders and of the alternatives proposed. The value trees developed to structure the decision problem are then presented and commented upon.

Chapter 3 presents the basis of integration. Starting from a general mathematical overview, the problem of prioritization is posed. Several possible prioritization tools are
considered. The attention is then restricted to AHP and SMART. Since the former is the alternative of choice, it is explored in depth in Chapter 4, which opens with an exposition of the mathematical foundations of AHP. The major concerns of the literature on AHP are the issues of consistency and rank reversal: these are discussed at length in dedicated sections.

Chapter 5 presents the elicitation results and debates the use of tools such as the inconsistency ratio in interpreting stakeholder input. After the Conclusions, an Epilogue will close the presentation of the case study.
2 THE DECISION CONTEXT

This Chapter provides some details of the case study, such as the site description, the stakeholders, the RAAs and the value trees developed to structure the decision problem. The purpose is to provide the reader with some background on the project and achieve some familiarity with the elements of the decision making problem. With this goal in mind, only the essential details are given. A complete description can be found in the final report submitted to DoE (Advanced Science et al., 1997).

2.1 Site description

The site of this case study was chosen among several candidates. The selection criteria were accessibility, existence of a comprehensive database, typical environmental problems and availability of a portfolio of remediation technologies. After consideration of a few eligible sites, the Chemical Waste Landfill at the Sandia National Laboratories (SNL) in Albuquerque, New Mexico was selected.

In a no action scenario, the greatest concern is groundwater contamination. The water table is approximately 490 feet below the CWL. It was estimated that groundwater flow is headed northwest at a rate of one foot per year. The 1.9-acre CWL is approximately 4 miles south of the nearest drinking water supply and at least 3 miles from any natural groundwater discharge point. The site was used for disposal from 1962 to 1985. Wastes were disposed of in different trenches and pits, typically 8 to 12 feet deep and at least 2 feet wide. Waste disposed at the CWL include chlorinated organics, chromic acid, a spent cyanide tank, acetone, petroleum hydrocarbons, aluminum hydrides, ammonium hydroxides, mineral acids, laboratory trash and beryllium-
contaminated material. Unfortunately, most records concerning the exact location, extension and type of contamination have been lost.

An agreement with the state of New Mexico Environment Department has been signed for some preliminary remedial activities, the Voluntary Corrective Measures (VCMs). It is expected that after the VCMs the only contaminants of concern will be chromium (Cr) and organics, mainly trichloroethylene (TCE), both present at depth greater than 50 feet below ground surface. Since the decision problem was concerned with post VCMs activities, these contaminants were the only relevant and, thus, the only considered in the risk assessments.

2.2 The stakeholders

In this project, the stakeholders are defined to be individuals or organizations who are interested in, involved with or potentially affected by decisions regarding Environmental Restoration and Waste Management activities at DoE sites. For the CWL stakeholders included DoE, SNL and private contractors; workers at the site; federal, State and local authorities; environmental and citizen groups; and Native American Nations. A total of 27 stakeholders were involved in the project. Of these, 11 (see Table 2.1) took some part in the prioritization process and 5 of them were present at the final deliberation meeting.

It is worth underlining the different backgrounds of the stakeholders. Beyond the different interests they represent, it should be mentioned that some of them can be considered as “technical” stakeholders. By this we mean that these stakeholders have some technical knowledge of the issues involved and are familiar with the hazards posed by the contamination and the possible remedial alternatives, with their advantages and disadvantages. In a couple of cases this knowledge was considerable and helped the risk assessment on a very technical ground. In some other cases the stakeholders showed limited technical preparation; nonetheless their contribute was precious in keeping alive the discussion on value-laden objectives. In a real situation who the stakeholders actually
Table 2.1: Stakeholder affiliation

<table>
<thead>
<tr>
<th>STAKEHOLDER</th>
<th>ORGANIZATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Real Estate Agent / General Public</td>
</tr>
<tr>
<td>2</td>
<td>Sandia National Laboratories</td>
</tr>
<tr>
<td>3</td>
<td>County Environmental Health Department</td>
</tr>
<tr>
<td>4</td>
<td>Council of Government</td>
</tr>
<tr>
<td>5</td>
<td>US DoE</td>
</tr>
<tr>
<td>6</td>
<td>Citizen Advisory Board</td>
</tr>
<tr>
<td>7</td>
<td>Sandia Community Involvement and Issues Management</td>
</tr>
<tr>
<td>8</td>
<td>Native American Nations</td>
</tr>
<tr>
<td>9</td>
<td>County Environmental Health Department</td>
</tr>
<tr>
<td>10</td>
<td>State Environmental Department</td>
</tr>
<tr>
<td>11</td>
<td>City</td>
</tr>
</tbody>
</table>

Technology options were identified by reviewing the concerned literature and relying on past experience for similar contaminants and environments. A first list of alternatives was prepared by the Team. The aim was to assemble a comprehensive list in compliance with the Office of Solid Waste and Emergency Response directive “Guidance for Conducting Remedial Investigations and Feasibility Studies under CERCLA”. This list was discussed with the stakeholders at the first workshop and a complete set of RAAs was generated.

The RAAs identified were a mix of specific technologies each addressing a different contaminant. They are summarized in Table 2.2. RAA F, no action, was initially
<table>
<thead>
<tr>
<th>RAA</th>
<th>METAL TREATMENT (CR)</th>
<th>ORGANICS TREATMENT (TCE)</th>
<th>OTHER TECHNOLOGIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>In situ vetrification</td>
<td>Soil vapor extraction</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>In situ stabilization</td>
<td>In situ bioremediation</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>Stabilization/solidification</td>
<td>Thermal desorption</td>
<td>Excavation, onsite disposal of treatment residuals</td>
</tr>
<tr>
<td>D</td>
<td>Stabilization/solidification</td>
<td>Thermal desorption</td>
<td>Excavation, offsite disposal of treatment residuals</td>
</tr>
<tr>
<td>E</td>
<td>Offsite treatment</td>
<td>Offsite treatment</td>
<td>Excavation, complete offsite disposal</td>
</tr>
<tr>
<td>F</td>
<td>No action</td>
<td>No action</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2.2: Description of Remedial Action Alternative

inserted for comparison purposes. Given the nature of the site and the relatively low contamination involved, it was not surprising to find, after the risk assessment were performed, that it was a competitive alternative.

2.4 Structuring the decision: the Value Trees

Once the site description is complete and the RAAs are defined, it is possible, in principle, to carry out the risk assessments. However, this is not advisable because it is still not clear what are the indexes that represent how well each RAA performs. Moreover, it has not been figured out how these many indexes can be combined in a single figure to characterize the overall performance of an RAA. This task is strictly connected to the more general issue of structuring the decision problem.

The first step in the development of a structure capable of organizing the decision problem is the construction of a comprehensive list of the goals that the implementation of the decision is expected to achieve. The top goal was defined by the stakeholders with the broad expression “Maximize the Benefits of Remediation”. For the CWL, the analysts identified, always with the cooperation of the stakeholders, six broad categories of objectives functional to this goal: “Programmatic Assumptions”, “Life Cycle Cost”, “Socioeconomic Issues”, “Cultural, Archaeological, and Historic Resources”,

22
"Environment" and "Human Health and Safety". This, at least for the particular case under investigation, seemed to form an exhaustive set of objective categories that any RAA should be concerned with (Figure 2.1). An alternate way of looking at these impacts is to say that the minimization of, say, the impact on human health and safety is an objective category that contributes to the maximization of remediation benefits, i.e., the overall goal.

It should be noted that, even if these categories may not apply to every problem, the procedure is absolutely general. The objective categories are very broad, and, as such, inevitably belong to the domain of qualitative concepts and do not lend themselves to a direct quantitative evaluation. At this point, thus, it is impossible to measure quantitatively the performance of the various RAAs according to these categories. Each objective category needs to be specified in further detail. The categories, then, are decomposed in their more elementary constituents; in other words, one should define which particular objectives make up each objective category. The objectives represent the specific action goals that the decision should achieve. Within the "Socioeconomic Impacts" category, for example, "Promote Community Quality of Life" and "Promote Environmental Justice" are two objectives that the implementation of an RAA should achieve (Figure 2.2). Similarly, all the objective categories can be expanded in terms of the related objectives. The result is reported in Figure 2.3.

The organization of the objectives in this hierarchy reflects the idea of means and fundamental objectives. This important distinction is made by Gregory and Keeney...
(1994). While fundamental objectives are those issues which are of direct concern to the stakeholder, means objectives are issues which are not a concern themselves, but rather matter only through their effect on the fundamental objectives. For instance, in Figure 2.2 it can be seen that "Promote Community Quality of Life" is not a concern by itself; rather, it has an influence on the decision only as long as it affects the "Socioeconomic Impacts", i.e. the related fundamental objective.

Similarly, the objectives can be defined in terms of performance measures (PMs). These indexes are the measurable quantities needed to evaluate how well an RAA meets the objectives. For example, the impact that a remediation technology has on local economy and the changes it introduces in the ambient conditions can be considered directly quantifiable measures of how a given RAA promotes community quality of life (Figure 2.2). The transition from a qualitative, general level (the objective categories) to a more specific, and quantitative, level (the PMs) is now complete.

Explicative examples (e.g. Edwards and Newman, 1986), as well as full-scale applications are common in literature. For instance, Merkhofer and Keeney (1987) have developed a hierarchy to evaluate the performance of five sites proposed for the final geologic disposal of civil nuclear waste.
The result of the structuring process is expected to be non-unique in the sense that different stakeholders are expected to organize differently their objectives. It would make a perfect sense even if every stakeholder had a personal hierarchic structure, as long as it captures his/her perception of the problem. Even in the case of a unique tree, the priority weights assigned to the branches of the tree, being a measure of the relative importance of the corresponding issues, are expected to be different for different stakeholders. The facilitator team can try to lead the stakeholders towards a consensus on the structure of the decision problem, for ease of the analysis that follows. This would have the welcome side effect of stimulating the discussion, helping the group to better focus the details of the issue. Of course, this must not result in biasing the stakeholders and, if consensus on a structure is not reached, separate analyses must follow for each different value tree.

This was indeed our case: while some stakeholders felt that “Minimize Long Term Risk to Public Health and Safety” should be accounted for in the objective category “Human Health and Safety”, most of them wished to rename this objective category “Worker Health and Safety” and move the mentioned objective under the objective category “Environment”. Accordingly, two value trees, V1 (Figure 2.3) and V2 (Figure 2.4), were developed and their differences were predicted to affect the quantitative analysis that follows. For example, for the stakeholders following the latter point of view, we anticipate a relatively higher concern for the category “Environment”, since it also incorporates the concern for public risk.

The importance of value trees is that they provide an effective and pictorial way of organizing both quantitative and qualitative, technical and value-laden objectives in a unitary logic structure which paves the way to quantitative evaluation. The task of prioritizing the objectives is, in fact, carried out on the basis of the value tree structure.
Figure 2.3: Example of complete prioritization with value tree V1 (SH2)

Objective Categories

Maximize Benefits of Remediation

Programmatic Assumptions
  0.022

Life Cycle Cost
  0.147

Socioeconomic Impacts
  0.021

Cultural, Archaeological & Historic (CAH) Resources
  0.059

Environment
  0.376

Human Health & Safety
  0.376

Minimize Waste
  0.022
  (1)

Minimize Direct Costs
  0.147
  (1)

Promote Community Quality of Life
  0.004
  (0.167)

Promote Environmental Justice
  0.017
  (0.833)

Protect CAH Resources
  0.059
  (1)

Protect Environmental Resources
  0.376
  (1)

Minimize Long Term Risk to Public Health and Safety
  0.188
  (0.5)

Minimize Short-Term Risk to Public & Worker Health & Safety
  0.188
  (0.5)

Objectives

Quantity of Waste Transported
  0.015
  (0.685)

(0.25)

Implementation Costs
  0.037

Quantity of Process Waste Generated
  0.005
  (0.234)

Completion Cost
  0.110
  (0.75)

Changes in Ambient Condition
  0.0033
  (0.833)

Changes in Resources
  0.313
  (0.833)

Compare Total Population Health Effects
  0.017
  (1)

Number Impacted/Severity Impacted
  0.059
  (1)

Contaminant Concentration
  0.063
  (0.167)

Individual Health Risk
  0.188
  (1)

Individual Worker Health Risk
  0.094
  (0.5)

Individual Public Health Risk
  0.094
  (0.5)

Performance Measures
Figure 2.4: Example of complete prioritization with value tree V2 (SH6)
2.5 The risk assessments

The performance measures were defined to be directly quantifiable quantities. As one can verify by inspection in Figure 2.3, all these quantities can indeed be evaluated once the site and the alternatives are known.

In principle, risk assessments can be carried out to determine all kinds of impacts of all remedial actions proposed by any stakeholder on any basis. This would result in a monumental effort, which, beyond the obvious waste of time and resources, could even lose the focus of the problem in a swarm of bare numbers. It is fundamental, then, to restrict the number of alternatives in preliminary stakeholder meetings. Moreover, the value tree structure should be thoroughly exploited to better define the performance measures and limit their number to the elements that are driving the decision.

As seen in the development of the value trees, risk assessments can target issues ranging from cost to contaminant concentrations and health impacts. All these aspects are strictly related: for instance, a contaminant concentration directly affects the health impacts and the cost of a remedial alternative. In fact, cheaper alternatives are expected to result in higher concentrations. This implies that all performance measures have to be evaluated using a consistent database. To ensure consistency, Influence Diagrams (IDs) were constructed (Hong and Apostolakis, 1993). Since different analysts were in charge of different parts of the risk assessment, this need was particularly felt in this project. Of course, this is strongly advised in the implementation of projects larger than this prototype.

The risk assessment output is, for each alternative and performance measure, a probabilistic distribution of the outcomes. In the course of this project, the analysts provided, for each alternative and each performance measure, a vector of twenty numbers. These were the result of twenty different Monte Carlo samplings of the parameters involved. Since the sampling was carried out through Latin Hypercube Sampling, each realization, that is each vector, was equiprobable. For what follows, it will be sufficient to remember that for each alternative and each performance measure we can define a mean
value, representing the expected impact of that alternative on that performance measure. This will be often referred to as the performance measure value. Note that this is different from the preference weight that each stakeholder assigns to that performance measure. Indeed, while the former is independent of the stakeholder, but depends on the alternative, the opposite is true for the latter, which depends on the stakeholder but does not depend on the alternative.
PRIORITIZING OBJECTIVES FOR INTEGRATION

The value tree for this decision problem was introduced in Chapter 2. As outlined there, the value tree offers a pictorial representation of the decision dimensions. This goes well beyond its obvious communication aspect: the same tree can in fact be exploited to combine the results of the risk assessments of the different alternatives with the personal values (weights) that the stakeholders give to performance measures, objectives and objective categories in an overall parameter, the Performance Index (PI). This index is the total (or multi-attribute) utility associated with the alternative. The higher it is, the more desirable the alternative. This process is called integration.

When, for a given stakeholder, the integration is carried out for all the alternatives, the stakeholder’s ranking is generated. The PIs are then calculated for all stakeholders.

As for the variability of the results, note that, for a given stakeholder, the same weights are used for all the alternatives. Therefore, the PI will vary across the alternatives because each alternative results in different values for the performance measures. For example, two different alternatives may result in different contaminant concentrations or different costs. On the other hand, the PI of the same alternative varies across stakeholders because each stakeholder will have different weights for the performance measures. In fact, for a given alternative, the difference in the ranking across the stakeholders is certainly not to be found in the performance measure values, since the risk assessments result can not, and do not, depend on the stakeholder.

This Chapter provides an overview of the integration procedure, starting from the conversion of risk assessment results into single attribute utilities to end with the synthesis of the Performance Index.
3.1 Mathematical formulation

The task of combining the performance measure values $x$ with the stakeholder weights $w$ is a mathematically formidable one. The matter is extensively treated in literature and it is the key point of Multi-Attribute Utility Theory (MAUT, Keeney and Raiffa, 1976). In principle, the objective is simply to find a function of $x$ and $w$ yielding the PI:

$$\text{PI} = \mathfrak{I} (x, w)$$

(1)

This approach is certainly very general, but it is still far from solving the problem. Since the function $\mathfrak{I}$ is not known, eq. (1) is not practical tool. Many mathematical forms for $\mathfrak{I}$ can be found in literature. In this project the most straightforward and common one was used. A quick derivation provides us with an opportunity to introduce its major assumptions and to show how the elements contributing to the PI enter its calculation.

The first step is to realize that the performance measure values do not enter the problem as such, but have to be converted into single-attribute utilities first. In fact, in multi-objective problems, the performance measures are not commensurate: while some of them are measured in terms of dollars (for costs), others can expressed as volumes (for waste generation) or a number of expected fatalities (for health and safety concerns) or even in terms of constructed scales (for social, cultural and archaeological impacts). In our case this creates a problem because a choice must be made, say, between an alternative which is less demanding in terms of cost, but impacts significantly the cultural resources, and an alternative which preserves such resources, but is more expensive. The problem is that these performance measures are not directly comparable.

MAUT starts from the observation that it is easier to attack the problem after converting the performance measures in a common unit of measure, the utility. This is the meaning of the single-attribute utility function $u(x)$: it makes commensurate quantities, like the performance measure values, that are non-commensurate by birth. The problem, then, can be restated in the form:
\[ PI = \Im (u_1(x_1), u_2(x_2), ..., u_N(x_N), w) \]  

(2)

A further step is to note that each weight is associated to a particular performance measure and it can be used in connection with it. The PI formula, then, becomes

\[ PI = \Im (w_1u_1(x_1), w_2u_2(x_2), ..., w_Nu_N(x_N)) \]  

(3)

This has a simple interpretation in terms of the value tree. Each tree twig represents a performance measure. A performance measure can be important in a decision problem because it is highly weighted by the stakeholder or because an alternative, despite the weight, results in high\(^3\) performance measure values. Therefore, all that matters is the product of these two quantities, which is called the weighted single-attribute utility. This is, thus, what should enter the PI calculation in eq. (3).

The last step is to assume that the functional form of \( \Im \) is the sum of its variables, i.e. the weighted single-attribute utilities. The final result is that the PI is a weighted average of the single-attribute utilities, being the performance measure weights the weights of this average. In mathematical symbols:

\[ PI^j_k = \sum_{i=1}^{N} w_{i,k} u_{i,k}(x_i^j) \]  

(4)

where

- \( PI^j_k \) is the Performance Index of the j-th RAA, with j = 1, ..., 6, for the k-th stakeholder.
- \( w_{i,k} \) is the absolute weight of the i-th PM. It is assessed by the k-th stakeholder. The sum of the weights is normalized to 1. \( i = 1, 2, ..., N_{PM}. \)\(^4\)
- \( x_i^j \) is the outcome of the i-th PM for the j-th RAA, obtained from risk assessments.
- \( u_{i,k} \) is the single-attribute utility of the impact of \( x_i^j \). It is assessed by the k-th stakeholder.

---

3 This is a comparative matter. "High" should be read "high as compared to other alternatives".

4 As it can be seen from Figure 2.3 and Figure 2.4, the number of performance measures is 14 for V1 and 13 for V2.
Of these variables, PI, \{u\}, and \(w\) depend on the stakeholder. In the following this dependence is not explicitly shown for notation simplicity.

The importance of this formulation is that it reduces the assessment task to the assessment of the single-attribute utility functions and of the performance measure weights. At the same time, eq. (4) is a major assumption, requiring that several hypotheses, which have to be verified, be made.

The mentioned applicability conditions are basically independence conditions. A classic MAUT example of the independence issue is the decision problem of a kid's Saturday night out. Our decision maker can be undecided between spending the night with an aging grandmother or at a wild fraternity party. The first solution would maximize the objective "preserve family peace", the second solution the objective "self enjoyment". In a linear model, both objectives are maximized by an hybrid solution, i.e. "going to the party with grandma", which is closer to the worse solution, rather than to the best.

This example shows how a linear model can fail to capture objective interactions. The issue has been given a rigorous mathematical form: the difference among different types of dependencies was highlighted and a rigorous definition of the concepts of preferential and utility independence was reached. By using these, decision analyst can assess if an additive approach is appropriate for the case at hand. The interested reader is referred to (Clemen, 1991).

In our particular case the issue was investigated in depth by Zio (1997). The conclusion was that the advantages of using a linear model overwhelmed the disadvantages, especially in light of the robustness of linear models. PIs were, then, calculated through eq. (4).

3.2 Different prioritization tools

As the risk assessment results \((x_i^f)\) are known, two pieces are needed for PI calculations: the single attribute utility function \(u_i\) and the weights \(w_i\). The procedure to
obtain the \( u_i \) is out of the limits of this thesis. However, it is interesting to note that a prioritization procedure is needed in this calculation as well. For this reason and, mainly, because the single attribute functions are fundamental in the calculation of the PI, the description of this task is treated at length in Appendix A. There it is shown how the normalization of the \( u_i \)'s can introduce an additional proportionality factor and thus affect the actual value of \( w \).

Then, from eq. (3) it is evident that all is needed are the weights for the performance measures. The assessment of the weights is called prioritization. In principle, one need not assess a weight for any impact category nor any objective, as long as the weights for the performance measures are known. In the following the point will be made that a direct prioritization of the performance measure is not easy nor time-efficient. The tree structure can be exploited to break the task in subsequent steps. In fact, comparing elements at the same level in the tree, e.g., the impact categories first and then the objectives within each category, has been found to be much easier for people to handle. It also reduces the total number of comparisons involved. Consequently, one has a prioritization at the impact category level, a second prioritization at the objective level and a third one at the performance measure level and one can speak of prioritization at different levels, even if the procedure is the same at each stage.

For a quantitative prioritization, several procedures, such as a direct assessment, the ratio method or the Simple Multi-Attribute Rating Technique (SMART) can be used (Edwards, 1977 and Edwards, von Winterfeldt and Moody, 1988). Other approaches consist in identifying a monetary equivalent for the performance measures or resort to the concept of lotteries and certainty equivalents. In this project, a combination of MAUT concepts and the Analytic Hierarchy Process (AHP; Saaty, 1980) was used.

Being part of a very general decision making methodology, the elicitation procedure is all but problem-specific. Examples of AHP applications range from engineering problems such as assessment of expert judgment credibility and selection of materials (Zio and Apostolakis, 1996; Paulos and Apostolakis) to social sciences issues, such as the assessment of the most livable US cities (Saaty, 1986a). Before starting a
description of AHP we review some prioritization techniques. Their discussion will substantiate why AHP was eventually chosen.

3.2.1 Direct assessment

The most straightforward way to assess the weights seems to be a direct assessment. This procedure can be viewed as a direct allocation of “100 chits”: given the list of elements to be prioritized, the decision maker allocates the chits in proportion to the importance of each element.

This apparently simple method, however, is not satisfactory because it does not delve deeper into the stakeholder preferences. First, the method is known to generate too “flat” weights (i.e. all the elements generally receive slightly different weights). Second, we recognize that such a method is bound to be fraught with uncertainties. Finally, it is impossible to check if there is self-consistency in the assessment. This is important, since it is well-known that people have difficulties expressing their preferences in a consistent manner.

To explain what is meant by consistency, let us note that, in principle, three elements can be prioritized eliciting the importance ratio between two of them. In fact, if, say, A is more important than B and B, in its turn, is more important than C, one might seemingly infer that A is more important than C. One can even assess the importance ratio of A to C from the ratios of A to B and B to C, with no need of a third elicitation. Unfortunately, this is not true in real life cases. Thus, a method that allows us to elicit information at a more elementary level is desirable.

3.2.2 The monetary equivalent

An approach often used aims at estimating a monetary equivalent for each performance measure. In our example, having an equivalent for the loss of cultural resources would make it trivial to determine whether the additional cost is worth the
expense. This homogenization procedure, extended to all the performance measures, allows a straightforward comparison of the overall value of the alternatives. Therefore, not only does it solve the problem of finding the weights, but it even bypasses the problem of eliciting the single-attribute utility functions.

An example of how monetary equivalents are determined can be found in Keeney and von Winterfeldt (1994), where national strategies to manage nuclear waste from power plants are compared. The issue of finding a monetary equivalent is trivial for immediate costs and is more involved for future costs, even if it does not require particular conversion tools beyond present worth calculations. Other quantities are not as easy to convert to costs. In this work, health impacts were quantified on the basis of site specific DoE reports: the unit value tradeoffs per worker and public preclosure statistical fatality were set to $1 million and $4 million, respectively. Finally, other performance measures needed the definition of a constructed scale, i.e. the definition of some impact levels, ranging from no impact to catastrophic consequences. At each level was associated a cost. For instance, in this study it was assumed that it would be worth $400 million to avoid a level five environmental consequence, defined to be a “major aesthetic impact and disruption of an endangered species habitat, plus impact to historical and or archaeological site of major significance”.

A drawback of the method is the emotional response of the stakeholders when they are asked to express all PMs in terms of dollars, especially when delicate issues, such as the value of life, are involved. This response can be blamed to be irrational in nature, but should not be dismissed, because past experience shows how it may lead to the failure of a project.

3.2.3 Standard MAUT tools

A third approach is suggested by the multi-attribute nature of the problem: to use the tools developed in the context of MAUT. Although theoretically sound, these
methods are often found to be quite difficult to apply, particularly when a vast and differentiated public is involved.

A difference of opinion among users of utility theory is found in the literature on single-attribute utilities. Curved functions are probably more precise representations of people's preferences; on the other hand, one of the major criticisms of multi-attribute utility analysis points to its deficiencies in practical assessments even when experienced practitioners of decision analysis and technically competent stakeholders are involved (McCord and deNeufville, 1983). This is in sharp contrast with the needs of our methodology, where ease of stakeholder participation is an issue of major concern. The main obstacles are the concepts of hypothetical trade-offs and lotteries, which do not always seem intuitive or anyway simple to the stakeholders. Moreover, it is not uncommon to find that in real situations the assumption of consistent behavior is not applicable. In later Chapters the discussion of inconsistencies and the workshop experience will provide evidence to back this statement.

In conclusion, the need to deal with simplicity and judgment inconsistencies suggested the use of different techniques. Two seemed to fit our case particularly well: AHP and SMART. Before discussing these in detail it is necessary to show how the development of the value tree prompts to their use.

3.3 Using the value tree in prioritization

As anticipated, the task of prioritizing the objectives is carried out in close connection with the value tree structure. As it is evident from eq. (3), all we need is the weights of the performance measures. Strictly speaking, then, one could skip the assessment of higher level weights, and compare directly the performance measures with each other. Nonetheless, the elicitation was carried out in several steps. This way of proceeding has the double advantage of breaking the task in easier steps and of reducing the total number of comparisons involved.
With reference to the value tree (see, for instance, Figure 2.3), the procedure starts from the objective category tier: all the categories are given a weight, ranging from 0 to 1, so that the ratios will reflect the relative importance of the categories and that the sum of all the weights is 1. We will see shortly how.

As an example, using input from stakeholder #2, Figure 2.3 shows that the objective "Promote Environmental Justice" is assigned a weight of 0.833, while the objective "Promote Community Quality of Life" is assigned a weight of 0.167 (both are under the category "Socioeconomic Impacts" and are shown in parenthesis). At the same time, under "Human Health and Safety", the objective "Minimize Long Term Risk to Public Health and Safety" was judged as important as the objective "Minimize Short Term Risk to Public & Worker Health & Safety". Each objective received a relative weight of 0.50. This does not mean, however, that the objective "Promote Environmental Justice" is this stakeholder's main concern. In fact, the category "Human Health and Safety" is more important than the category "Socioeconomic Impacts" (0.376 vs. 0.021). The absolute weights of the objectives are obtained by multiplying the relative weights (i.e., within each category) by the weight of the category itself. Thus, the objectives "Minimize Long Term Risk to Public Health & Safety" and "Minimize Short-Term Risk to Public & Worker Health & Safety" receive a weight of 0.188 (0.50 x 0.376) each, while the objective "Promote Environmental Justice" receives 0.017 (0.833 x 0.021) only, reflecting this stakeholder's concern for Health and Safety.

The last step is the comparison of performance measures relative to the same objective. When the weights are multiplied by the absolute weights of the objectives, absolute weights are obtained for the performance measures. This procedure implies, that, as it should be, \( \sum_{j} w_j = 1 \), where the \( w_j \) are the \( w \)s of eq. (4), i.e. the absolute weights of the performance measures.

The problem is now reduced to the elementary prioritization procedure, that is the weight assessment within a given level. Two different techniques are presented: SMART and AHP.
3.4 Estimating priority weights via SMART

SMART is based on the direct elicitation of ratio judgments of importance. To show how the weights are elicited, let us suppose that four elements, A, B, C and D are to be prioritized. First, the elements are ordered by decreasing importance and the lowest ranked element is given the score of 10. This allows the elicitation of ratios lower than 2 without recourse to fractional numbers. Second, all other scores are obtained comparing the elements to the lowest ranked, one at a time. The score is given on a 10 to 100 scale on the basis of a ratio judgment. In the example of Table 3.1, C is twice as important as D, and it is thus given a score of 20 (column I).

When all the scores are elicited, one could divide them by their sum to get the normalized weights. This, though, would rely on the hidden assumption that one can assess the importance ratio of two elements comparing them indirectly with a third one. In our example, the relative preferences over A, B, and C are obtained indirectly through comparisons with D (the lowest ranked element). Strictly speaking, this information is sufficient to supply all the weights but in real cases this seemingly obvious “transitivity property” - and with it the “rational” behavior of the stakeholder postulated in the axioms of Utility Theory - is not warranted. This means that if the apparently unnecessary direct comparisons among A, B and C, are made, it is very likely that the result will be different from the expected one (i.e. the one obtained through indirect comparison with D).

<table>
<thead>
<tr>
<th>Values</th>
<th>Column I</th>
<th>Column II</th>
<th>Column III</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>80</td>
<td>40</td>
<td>20</td>
<td>0.53</td>
</tr>
<tr>
<td>B</td>
<td>40</td>
<td>20</td>
<td>10</td>
<td>0.27</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
<td>10</td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td></td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.1: Example of weight derivation via SMART
Concrete examples taken from workshop experience are presented in section 5.3.

This inconsistency is not reducible to a matter of approximation inevitably present in human judgment. Some problems are somehow intrinsically inconsistent. Even when dealing with a perfect measurement, inconsistency might rise. In section 4.2 it is discussed at length not only that inconsistency in comparisons should be expected, but even that it can lead to useful insights into stakeholder’s preferences. For this reason, more than the “strictly necessary” information is considered of interest. In general, when n elements are compared all the possible n*(n-1)/2 pairwise comparisons are needed in place of n-1. In this example n equals 4 and there are 6 possible pairwise comparisons.

SMART deals with inconsistencies by eliciting more information from the stakeholder: the second column of Table 3.1 illustrates how. D is now ignored and C, the second lowest element, is set to the arbitrary value of 10. Then, the weights of A and B are elicited by comparison with C. If the ratios of these scores are reasonably close to those of the first column, the test can be repeated setting B to 10 and assessing the relative importance of A (column III). If the ratio is consistent with the previous ones, the test is successfully completed. On the contrary, if this is not the case, the inconsistency is brought to the assessor’s attention and the numbers are iteratively revised until consistency is obtained. Often the respondents learn the arithmetic involved and accordingly adjust their answers to obtain consistency (Edwards, von Winterfeldt, Moody, 1988, pg. 449). This is not always desired, since it brings us back to the problems outlined for direct assessments: the assessor can start from the first column and then derive the successive ones, thus making useless the consistency check. The discussion of AHP will make clear that inconsistencies can be, rather than a nuisance due to human judgment imperfections that has to eliminated, an interesting part of stakeholder assessments.

The final step of SMART prescribes that, when a consistent pattern is achieved, the weights be calculated dividing the first column by its sum. This result is shown to the assessor to confirm that it indeed captures his/her preferences.
3.5 Estimating priority weights via AHP

SMART explicitly recognizes the natural occurrence of inconsistent judgments and provides the analyst with a first tool to deal with them.

The Analytic Hierarchy Process (AHP), not only accepts the existence of inconsistencies, but even builds a mathematical framework to investigate them. A second difference from SMART is the elicitation scale, which can be semantic rather than numerical. As in SMART judgments are ratio judgments, but are expressed in linguistic terms as "weakly preferred" or "strongly preferred". They are converted to a 1 to 9 scale by means of the semantic equivalents of Table 3.2. A detailed description of AHP and of the cognitive science rationales at the basis of the scale can be found in Saaty (1980). Unlike SMART, the technique was born for hierarchies even more complex than value trees (see Appendix B), but it can be naturally adapted to this simpler case.

To use the same example discussed for SMART, let us consider again the prioritization of the four elements A, B, C and D. Pairwise comparisons are performed for all the possible \( \frac{4 \times (4-1)}{2} = 6 \) pairs. For each pair the assessor is asked which element is more important, and how strongly, in terms of the semantic scale. Using Table 3.2, this assessment is converted to a number and the score \( s \) is designated as the entry \( a_{ij} \) of A, the matrix of pairwise comparisons. A is defined so that each column \( j \) and each row \( i \) is associated with one of the elements being compared. The entry \( a_{ij} \) indicates by how much

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>if the two elements are equally important.</td>
</tr>
<tr>
<td>3</td>
<td>if one element is weakly more important than the other element.</td>
</tr>
<tr>
<td>5</td>
<td>if one element is strongly more important than the other element.</td>
</tr>
<tr>
<td>7</td>
<td>if one element is demonstrably or very strongly more important than the other element.</td>
</tr>
<tr>
<td>9</td>
<td>if one element is absolutely more important than the other element.</td>
</tr>
</tbody>
</table>

Numbers 2, 4, 6 and 8 can be used to express compromise between slightly differing judgments.

Table 3.2: Semantic scale used in an AHP prioritization
the $i$-th element prevails over the $j$-th element. If the opposite is true, i.e. the $j$-th element prevails over the $i$-th element, the inverse of $s$ is used for $a_{ij}$, i.e. $a_{ij} = 1/s$ and $a_{ji} = s$. Note that this is not assuming any “transitivity” property. Here the assumption is that comparing, say, $A$ to $B$ will generate the opposite result of comparing $B$ to $A$. No third element is involved. From this definition, it follows that the diagonal elements of $A$ are equal to unity, since any element is as important as itself.

Returning to our example, if the stakeholder estimates that $C$ is somewhere between equally and weakly more important than $D$, from Table 3.2 we see that $s$ is 2. $C$ is more important, so $s$ is inserted in the matrix in the 3rd row, 4th column (i.e. $a_{34} = s$). The value of $1/2$ is inserted in the symmetric position, i.e. $a_{43} = 1/s$. Similarly, all the matrix can be determined. Table 3.3 shows the matrix that is obtained if the preferences are the same expressed in Table 3.1.

As a further example, the entry $a_{11}$ is equal to 1/8. This means that the assessor judged element $A$ to be between demonstrably and absolutely more important than $D$. Since $D$ has a rating of 1/8 with respect to $A$ and $C$ has a rating of 2 with respect to $D$, consistency requires that $A$ have a rating of 4 with respect to $C$, i.e., $A$ must be somewhere between weakly and strongly more important than $C$. This would preserve the transitivity property among the elements $A$, $C$, and $D$. As the table shows, this is indeed the case here.

The priority weights are derived from the matrix $A$. Among the many feasible approaches Saaty recommends to compute the weights as the components of the eigenvector of $A$ associated to the largest eigenvalue $\lambda_{\text{max}}$ (principal eigenvector),

<table>
<thead>
<tr>
<th>Matrix</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>0.53</td>
</tr>
<tr>
<td>B</td>
<td>1/2</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0.27</td>
</tr>
<tr>
<td>C</td>
<td>1/4</td>
<td>1/2</td>
<td>1</td>
<td>2</td>
<td>0.13</td>
</tr>
<tr>
<td>D</td>
<td>1/8</td>
<td>1/4</td>
<td>1/2</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3: Example of weight derivation via AHP
normalized so that they add up to one. The last column of Table 3.3 contains the priority weights for the elements of A, B, C and D of the example. This is the principal eigenvector of A. Note that, in this case of consistent entries, all the columns are proportional to the principal eigenvector, while the rows are proportional to the inverse of its elements. In the inconsistent case these properties do not hold. Deviation from consistency is reflected by the largest eigenvalue of the matrix, and an inconsistency index can actually be defined. A deeper discussion of the eigenvector method and the inconsistency index require some mathematical background and are therefore delayed to Chapter 4.

As stated earlier, after the calculations are carried out, it is imperative that the results be shown to the stakeholders, to give them the opportunity to confirm that they indeed reflect their perception of the relative importance of the elements. This is of extreme importance. It would be naive to assume that any procedure of this nature would be capable of producing the correct weights directly. Rather, the application of AHP should be interpreted as a means of obtaining an approximate set of weights which is to be refined through successive iterations with the stakeholder. In other words, we never rely exclusively on the chosen method to produce the weights.

Note that while in SMART the stakeholder can, and is asked to, change the entries to reach consistency, in AHP, due to the more involved calculation, this is not immediately possible. This is certainly not a shortcoming of AHP. In fact, in SMART the stakeholder is somehow asked to force the results in a consistent form and this results in canceling part of the input richness. From a certain point of view, one can say that in SMART inconsistencies are treated as a background noise that should be eliminated. In AHP inconsistencies are accommodated into and permeate the result.

A simple example can be reported to clarify this point. Let us suppose that three pairwise comparisons were elicited in the prioritization of three elements, A, B and C. A was judged weakly more important than B, B weakly more important than C and C weakly more important than A (Table 3.4).

If the principal eigenvector is computed, the obvious result that all the alternatives are equally desirable is found. Note how in this case SMART could be difficult to apply
from its first step, since a clear ranking of the alternatives is difficult to reach. On the other hand, AHP proves not only to be easily applicable, but even to reach a reasonable result in a case where inconsistencies are an inherent part of the ranking computed\(^5\).

Finally, it is important to emphasize that the pairwise comparisons are carried out in the context of the specific problem under investigation. When the stakeholders are asked to perform pairwise comparisons at the impact category level they are not comparing “Human Health and Safety” to “Socioeconomic Impacts” (for example) in general, but, rather, for the specific problem at hand. This requires at least a qualitative discussion with the stakeholders of the range of impacts that are expected for this problem. In the case of the CWL, the contamination was not high and the stakeholders were fully aware of the fact that no severe impacts were anticipated. In fact, one stakeholder demanded to see the actual results for the performance measures before she made any comparisons. As the risk assessments had not been done yet, a qualitative discussion supplemented with some quantitative range estimates was all the information that could be provided. When the actual results became available a few months later, she was asked whether she would like to modify the weights that she had assigned. She declined stating that the original set was representative of her preferences.

\(^5\) Cases such as this are not of a purely academic interest. In Chapter 4 it is discussed how they are found in practice. In Chapter 5 are presented some examples taken from this case study.
3.6 Conclusions

The natural output of risk assessments is a set of purely quantitative results. Generally speaking, these results are rather technical and often do not lead to an immediate decision *per se*. Very frequently the decision depends, as it should be, on considerations that go beyond a purely technical fact. In this case, the set of technical data has to be integrated with a set of values reflecting political and personal views. A quantitative prioritization of the preferences is the tool needed to translate qualitative preferences in a form suitable for rigorous processing and thus bridge these two worlds.

In this Chapter several candidate prioritization procedures were discussed. The need for a rigorous method, easy to teach and implement and capable of handling inconsistencies, lead to the choice of AHP. This method is far from being perfect and more than one caution should be taken in its use. In the following Chapter, the foundations of the method are presented along with its limitations, criticisms and the countermeasures that have to be taken to overcome them.
4 DISCUSSION OF AHP

In Chapter 3 an overview of AHP as a viable prioritization tool was presented. This method has been the object of several studies which spotted some weak points of the methodology. In particular the main criticisms concern the possibility of rank reversal, the limitations imposed by the semantic scale and an alleged incompatibility with MAUT. From a more mathematical standpoint, the eigenvector method for extracting the priority vector from the matrix of preferences was investigated. Finally, in the course of the project, the use of the inconsistency ratio was questioned, since it has the potential to be used to discredit the stakeholder input.

A survey of the literature revealed that none of the mentioned problems has been left completely unsolved. Yet, for some questions, namely the semantic scale, the classic conclusion that no solution is universal but every application is peculiar and, as such, needs peculiar adjustments was reached. Before introducing these points, this Chapter presents some more advanced features of AHP in a rigorous mathematical context. This will provide the basis for the discussion of the aforementioned issues and for the explanation of the solutions adopted in this project.

4.1 The mathematical basis of AHP

In Chapter 3 it was discussed through a particular case that when it is necessary to rank n alternatives, various approaches are at our disposition. While in Utility Theory absolute measurement is the tool most widely adopted, AHP relies on the use of relative measurement. The stakeholder is not asked to assign a score to the alternatives on an absolute scale, but to rank them considering them two at a time and answering a set of questions in the comparative form: “How strongly is alternative $A_i$ preferable to alternative $A_j$?".
As anticipated in section 3.4, N-1 comparisons - at least - are needed to work out the relative standings of N elements. It was discussed how this conjecture relies on the implicit assumption that asking the "unnecessary" comparison would give a trivial answer and, thus, no further information. In common practice perfect consistency is not guaranteed at all: indeed, in a real case, it is very likely that the result of the "unnecessary" comparison will be different from the ideal, consistent, one.

Until very recently it was believed that transitivity of preferences was essential for constructing value functions that truly represented the preferences of decision makers. However this is not so. The literature is full of examples that illustrate the intransitivity of decision makers. (Vargas, 1986)

and again

This happens frequently and it is not a disaster (Saaty, 1980).

For this reason, all the possible pairwise comparisons, N*(N-1)/2, are needed and elicited. Finally, it was described how AHP organizes the results of the elicitation in the matrix of the pairwise comparisons, in which entry $a_{ij}$ represents by how much alternative $A_i$ is preferred to alternative $A_j$. These definitions imply two important properties:

1. $a_{ij}^{-1} = a_{ji}$, so the transpose matrix entries are the reciprocal of the matrix entries. This is imposed to take into account that if $A_i$ is $a_{ij}$ times as important as $A_j$, then $A_j$ must be $1/a_{ij}$ times as important as $A_i$. 6

2. setting $i=j$ in the equation above, one obtains $a_{ii}^{-1} = 1/a_{ii}$ and $a_{ii} = 1$, because $a_{ij}$ must always be positive. It is obvious that an element is as important as itself.

Matrices such as these are known as reciprocal matrices. Two numerical examples were presented in section 3.5.

---

6 Note that this has nothing to do with the issue of consistency. Here we are comparing an element with another and we assume that the symmetric comparison is trivial. Consistency is dealt with when three or more elements are involved and has to do with the transitivity of the comparisons, not with their symmetry.
The priority weights are derived from the matrix $A$. From a mathematical point of view, the problem is that of deriving a vector (the priority vector $w$) from a matrix. In this sense, eliciting redundant information results in having more than one way for working out the weights. Many methods have been proposed to calculate $w$. Among these the most commonly found in literature are (Saaty and Vargas, 1984):

1. The method of least squares (LSM), which finds $w$ minimizing the Euclidean metric

$$\sum_{i,j=1}^{n} (a_{ij} - w_i/w_j)^2$$

2. The method of logarithmic least squares (LLSM), which minimizes

$$\sum_{i,j=1}^{n} (\log a_{ij} - \log(w_i/w_j))^2$$

This method is also known as the geometric row mean method, as it finds the vector which has minimum geometric distance from the rows of the matrix.

3. The method of the principal eigenvector (EM), which solves the problem:

$$A w = \lambda_{\text{max}} w$$

in which $\lambda_{\text{max}}$ is the principal (the largest in magnitude) eigenvalue of $A$.

These approaches have been widely debated in literature. Fichtner (1986), showed how the problem of the choice of the appropriate algorithm can be solved starting from four axioms\(^7\). The first two (Correctness in the Consistent Case and Comparison Order Invariance) screen out some rough methods such as taking a column of $A$ and normalizing it. The third (Smoothness) eliminates other “unreasonable” methods as well.

\(^7\) An axiomatic foundation of AHP was provided by T.L. Saaty (1986b) as well.
as an often proposed one: LSM. A fourth axiom is needed if a definitive solution is being sought. Postulating Power Invariance leads to LLSM, while postulating Rank Preservation leads to EM. This is the method suggested by AHP. The weights are computed as the components of the right eigenvector of A associated to the largest eigenvalue $\lambda_{\text{max}}$ (principal eigenvector), normalized so that they add up to one.

The choice of the axioms is arbitrary:

There is little hope that somebody will find a proof to show that [an axiom rather than another one] is better. Here we are beyond mathematical truth. [...] for real situations [...] in most cases we get very similar situations for both methods. So the discussion on best methods for deriving priority vectors from matrices of pairwise comparisons should not prevent successful applications (Fichtner, 1986).

No proof can then be given to show that the EM method is the only right way of deriving the weights $w$. Accordingly, all one can present is an intuitive justification of what at a first glance might seem an esoteric method (Saaty, 1980). The conclusion that the EM method is probably the one with the best mathematical properties was reached by Saaty and Vargas (1984).

As a first step it is assumed that the judgments are the result of an absolutely exact physical measurement (e.g. the height of some mountain). In this case we have immediately, by definition of weight, the vector $w$. From these, the entries of the matrix can be calculated as:

$$a_{ij} = \frac{w_i}{w_j}$$

i, j = 1, 2, ..., N.

The i-th row looks like:

$$[w_i/w_1, w_i/w_2, ..., w_i/w_i, ..., w_i/w_N]$$

On this basis, we can infer that if we were given a matrix and we multiplied the first element of a row by $w_1$, the second by $w_2$ and so on we should get always $w_i$. Indeed, this is what happens in the consistent case.
Let us consider now the more realistic case of measurements affected by deviations. So, starting from the results coming from an instrument that measures the ratio of the weights, following the previous procedure yields a set of estimates for $w_i$. It is reasonable to assume as $w_i$ the mean of the $w_j$s so obtained. In mathematical symbols:

$$w_i = \frac{1}{N} \sum_{j=1}^{N} a_{ij} w_j$$

which in matrix notation becomes:

$$A \mathbf{w} = N \mathbf{w}$$

Not always this problem has a non-trivial solution for $\mathbf{w}$. Indeed, it is possible to find a vector which satisfies this relation in the ideal case only. Furthermore, if the entries of $A$ are allowed to deviate from the consistent case value, a solution is possible only if $N$ is allowed to deviate from the number of alternatives. We indicate this new value with $\lambda_{\text{max}}$. The formulation of the well-known eigenvalue problem is finally found:

$$A \mathbf{w} = \lambda_{\text{max}} \mathbf{w}$$

It is possible to prove that for reciprocal matrices small deviations of the entries of $A$ result in small deviations of $\lambda_{\text{max}}$ from $N$. This result is extremely important, since it allows to quantify deviation from consistency by means of $\lambda_{\text{max}}$. Note that in the consistent case $\lambda_{\text{max}}$ is equal to $N^8$ and increases with increasing inconsistencies. On this basis it is possible to define the inconsistency index:

---

This is easy to prove. In the consistent case all the rows (columns) are proportional. The rank of the matrix is then one and all the eigenvalues of the matrix must be 0, but one. Since the trace of the matrix, which is equal to both the sum of the diagonal elements and the sum of the eigenvalues, is $N$, the only non-zero eigenvalue is the maximum and it must be $N$. 

---
\[
\frac{\lambda_{\text{max}} - N}{N - 1}
\]

and the inconsistency ratio\(^9\), which is the ratio of the inconsistency index over the expected value of the inconsistency index of an \(N \times N\) matrix derived from random assessments (random index). This definition is based on the observation that, for reciprocal matrices, \(\lambda_{\text{max}}\) is always real and greater than \(N\). If, for any reason, the columns of the matrix are not proportional (e.g. because the measurement is affected by random errors), the ratio will move from zero, increasing slightly and revealing inconsistency. Dividing by the random index has the effect of normalizing the inconsistency ratio. Thus, in terms of the inconsistency ratio, a pairwise comparison questionnaire filled out randomly results, on the average, in a value of 1. On the other hand, if the judgments are all consistent, the inconsistency ratio is always 0. As a rule of thumb, an inconsistency ratio higher than 0.1 is the standard value assumed as consistency limit. This limit should not be interpreted too strictly. Saaty (1980) recommends to set this boundary as suggested by the case at hand. As it will be clear from our results, in this project a value of 0.15 is more appropriate.

We note that this ratio is for the analyst's use and it is not communicated to the stakeholder. When inconsistency is relatively high, the analyst looks for the inconsistencies in the stakeholder's pairwise comparisons and points them out to the stakeholder who is, then, free to change them so that greater consistency can be achieved. Telling the stakeholder that a mathematical quantity called the inconsistency ratio is high does not serve any purpose and, in fact, can be detrimental.

As a final remark we wish to emphasize the deeper mathematical meaning of the inconsistency ratio. If the axiom of "Correctness in the Consistent Case" is postulated, the following is straightforwardly demonstrated (Fichtner, 1986):

**THEOREM:** If and only if a way of calculating the vector of the weights fulfills this axiom, it can be derived from a metric.

---

\(^9\) Note that Saaty calls it the consistency ratio.
This means that the priority vector can always be calculated using a metric, that is, minimizing the distance function defining the metric. Once that the metric is chosen, the priority vector is a normalized column of the consistent matrix having the minimum distance from the matrix of comparisons A. This distance is the inconsistency of A: an example of distance function is the inconsistency ratio presented above.

4.2 Inconsistency sources

In light of the previous section, one can see how AHP approaches the problem in a completely different way from SMART. Inconsistencies are incorporated rather than considered aberrations. Furthermore, it was shown how inconsistency rises from the non-transitivity of preferences. Such cases are more than pathological exaggerations. The best way to convince the reader is perhaps to present two real cases, one taken from everyday life, the second documented in literature.

The most obvious case is found in sports, where it is possible that Team 1 defeat Team 2, Team 2 defeat Team 3, which can, in its turn, defeat Team 1.

May (1954) has studied the idea that intransitivity among preferences may be a natural phenomenon. The conclusion is that it is impossible to avoid considering intransitivity as such. In an experiment 62 college students were asked to rank three possible spouses A, B and C. Three “performance measures” were chosen: intelligence, beauty and wealth. A was described as very intelligent, plain looking and well off; B as intelligent, very good looking and poor; C as fairly intelligent, good looking and rich. It was made the point that no candidate was so poor, plain or stupid to be clearly out of the competition. The survey indicated that A was preferred to B by 39 to 23, B to C by 57 to 5 and C to A by 33 to 29. The reasons for this ranking were investigated and found to be not unreasonable. Even if this is no definitive proof that preferences are not transitive, it makes still sense to explore the conditions under which transitivity fails.
Luce and Raiffa (1985) explicitly note that intransitivities exist, go against the transitivity axiom of Utility Theory which, then, can not deal with such cases, unless it is argued that the consistent case is a close approximation of real cases.

To this point, a high consistency ratio might be explained in, at least, two ways:

1. According to the mathematical definition, it is the result of a questionnaire filled out randomly. In this case the consistency ratio is likely to be close to 1.

2. It is a warning signal: the problem might be intransitive in its nature. This interpretation would be supported if the vast majority of the questionnaires reported a high inconsistency ratio. This did not happen in our case.

Along with these theoretical reasons, apparently trivial causes of inconsistency deserved the attention of the literature (Forman, 1986):

3. Lack of stakeholder concentration during the judgment process.

4. Clerical errors, such as wrong computer input, like inserting “the inverse of what was intended”.

5. Lack of information could lead to lack of understanding. Judgments will appear to be random and we are reduced to case 1.

6. Inadequate model structure. A well constructed decision problem involves making sure that at all levels comparisons are being made between factors of similar importance (e.g. of the same order of magnitude). This is related to:

7. High inconsistencies can result if extreme judgments are needed. If factors of very different importance are being compared, the 1 to 9 scale can be saturated.
It is not surprising that most of these causes were found in our case. The presentation of the results will show that non-technical stakeholders and stakeholders who, for reasons independent of the Team’s responsibility, were under time pressure showed relatively high inconsistency ratios. The input in the computer of “the inverse of what was intended” really happened, despite careful double checking by two different people. Two stakeholders recognized a similar clerical error in their input. These errors are somehow inevitable and, despite their detrimental potential, often go undetected in most numerical analyses. From this perspective, the inconsistency index is an effective debugging tool.

Since one of the project’s main goals was risk communication to a wide public and the development of a decision making structure, the issues of lack of information and inadequate model structure were addressed carefully at a different level, well beyond the mere prioritization procedure. The discussion of the problems involved at the site in the course of the workshop was extensive and, from what we can tell from the questionnaires and stakeholder feedback, exhaustive. Accordingly, there is no evidence that inconsistency was caused by lack of knowledge.

However, realizing that high inconsistency does not imply that the calculated ranking is senseless must not lead to the opposite extreme, that of accepting any inconsistent result, without any further consideration. Some mathematical procedures were developed to refine the matrix to reach consistency; but the conclusion is that using brute force to improve consistency distorts the answer rather than improving the results. As Forman notes:

Low inconsistency [must] not become the goal of the decision making process. A low inconsistency is necessary but not sufficient for a good decision. It is more important to be accurate than consistent. It is possible to be perfectly consistent but consistently wrong.

Even though some iterative methods based on matrix manipulation were developed to reduce inconsistency, Saaty (1980) warns against their use, which tends to distort the answer. A better consistency should be reached by a higher familiarity with the elicitation procedure. Accordingly, great importance is given to making the assessor more and more
acquainted with the elicitation procedure. As it was shown by workshop experience, a higher familiarity can be reached in a couple of evaluations and does result in better consistency.

At this point of the discussion, it is clear how accepting inconsistent judgments does not undermine their validity, as far as the inconsistency is being monitored. The inconsistency ratio is used to verify its extent and to trace back its origin, focusing the discussion on the points that have the highest potential of enhancing the accuracy of stakeholder input. This opportunity is even more valuable in a context of limited resources and time pressure.

4.3 The semantic scale

In section 3.5 it was shown that a semantic scale (Table 3.2) is used to assess priorities. The scale maps degrees of preference expressed verbally to a limited, numerical, 1 to 9 scale. This aspect has been criticized (Dyer, 1990a). As the author recognizes, this is not a major criticism; nonetheless it raises an important issue.

A linear 1 to 9 scale was chosen for AHP among a variety of scales, ranging from 1 to 5 to 1 to 90, on the basis of fundamental studies on human perceptions. The choice was the result of a compromise between the opposite need of being able to perceive the difference between slightly different levels of the scale and leaving an adequate number of degrees of freedom to the assessor. To use an example, on a 1 to 90 scale it is difficult to argue that the assessor can actually tell the difference between a score of 77 and a score of 78. At the other end, on a 1 to 3 scale the difference between the levels is certainly clear, but it is impossible to rank 4 elements, due to the lack of degrees of freedom. A simple solution is the use of fractional numbers, which are certainly not forbidden, but do not come naturally to the mind.

Holder (1990) pointed at a “violence to the normal usage of English language” inherent in this semantic scale. The author argues that if a stakeholder wants to be consistent and assesses that A is “weakly” more important than B (3 on the semantic
scale) and B is "weakly" more important than C (3 on the scale), then it must follow that A is $3 \times 3 = 9$ times more important than C, that is, accordingly to the scale, A is "absolutely" more important than C. In fact, in the consistent case, we can write\(^\text{10}\)

$$a_{ij} = \frac{w_i}{w_j}, \quad \forall i,j$$

this immediately implies

$$a_{ij} a_{jk} = \frac{w_i}{w_j} \frac{w_j}{w_k} = a_{ik}$$

Alternatively, one can verify that a 9 is needed in the matrix of comparisons to preserve the proportionality of the rows (Table 3.5).

In reply to criticisms concerning the abrupt interruption at 9 of the possible values and the limited freedom given to the assessor, it should be stressed that the scale itself and the verbal form of the assessment are not fundamental part of AHP. The scale can be tailored to suit the assessor needs without undermining the validity of the method. As for the criticism on the use of words and their conversion to numbers, it must be noted that this is not the only possible elicitation format. If preferred by the stakeholder, a numerical format is possible (the assessor gives directly a 1 to 9 score) as well as a graphic format, where the assessor expresses his/her degree of preference by indicating among several

<table>
<thead>
<tr>
<th>Matrix A</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>1/9</td>
<td>1/3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.5: Holder's argument of semantic scale inconsistency

\(^{10}\) Recall from section 3.1 that in the consistent case the matrix elements can be written with a "backward" process from the weights.
sample pie or bar graphs which looks closer to the strength of the preference.

All these alternatives were considered in the course of the project. It was noted, however, that even a more "objective" approach, such as a numerical format, could present significant drawbacks. In several applications of AHP the analyst has to measure quantities for which no agreed measurement procedure exists. For instance, in this context concepts such as "relative importance", "relative impact", and "desirability" have to be determined. In such cases it would be very difficult to give a universal definition of numerical values: a statement such as "Roses are four times as sweet-scented as daffodils" needs a definition of the scale used to assign the value four.

These assessments are not trivial at all even in cases where priorities are connected to physical quantities. It has been noted (Fichtner, 1986) that the age of a car is certainly a decision driver, but the ratio of ages does not necessarily provide the priority ratio. Similarly, a $20,000 car can be much more expensive than a $15,000 car for a limited income family, while this difference is not considered in a national-scale budget. Thus, using numbers rather than words can make too tempting the shortcut of using the ratios of the numerical quantities as priority ratios. In this sense numbers may be even more subjective than words and the semantic scale is a means of providing a glossary common to the evaluator and the analyst.

In the project the issue was addressed using the semantic scale and waiting for stakeholder feedback. The fact that nobody felt uncomfortable expressing his/her judgment nor asked for further clarification was considered a positive indication. The best verification was the buy-in obtained upon showing ranking and weights: unsatisfactory results were due to different reasons that were identified and eventually eliminated in close cooperation with the stakeholders.

4.4 Rank reversal and the AHP absolute mode

The output of a prioritization methodology is the ranking of the alternatives under consideration. A major problem of AHP was first identified by Belton and Gear (1983)
and later discussed by Dyer (1990a, 1990b) and Holder (1990). The question is whether the ranking is preserved when the set of alternatives is changed either by adding or dropping one or more alternatives\textsuperscript{11}. A reply was given by Saaty (1990) and Harker and Vargas (1990), with reference to the previous publication by Saaty (1987).

In this paper it was shown that rank is unaffected, in the consistent case, when the judgment values of an alternative dominate those of the other alternative in all of the pairwise comparison matrices. Yet rank reversal may occur when an alternative dominates another one under one criterion but the opposite is true when other criteria are considered. Since rank reversal is the most important issue debated in AHP literature, it is important to understand what it means and its origin.

The problem appears when AHP is used in the relative mode, which is the original version of the method, but can be completely eliminated using AHP in the so called absolute mode. In order to do this, the classic AHP decision hierarchy (see Appendix B), which is a relative mode hierarchy, must be slightly modified. The decision tree discussed in Chapter 2 already incorporates the adjustments necessary to transform a relative mode hierarchy in an absolute mode hierarchy. Thus, in the example below, a relative mode AHP hierarchy was built to illustrate the issue.

A decision tree was developed for the choice of a new car (Figure 4.1). For simplicity’s sake, two criteria only are being considered: cost and performance. The relative importance of these criteria can be assessed through a pairwise comparison: if

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{decision_hierarchy.png}
\caption{Decision hierarchy for the car selection problem}
\end{figure}

\textsuperscript{11} Note then the difference with the Rank Preservation axiom, which deals with a given number of alternatives.
they are judged to be of equal importance, they will both have a weight of 0.5. Furthermore, let us assume that two alternatives are possible: A, an expensive but highly performing car and B, much cheaper but less brilliant. The only difference with an absolute mode hierarchy is at the interface between the alternatives and the performance measure level. In the absolute mode the alternatives are considered one at a time. They are characterized by a set of numbers: the numerical values of their performance measures, in this case the cost and, say, the maximum speed. These numbers are substituted in the single-utility functions, which provide an absolute standard against which the alternatives are compared. In the relative mode the alternatives are considered all at the same time and are compared pairwise under each performance measure. In this simple example are being considered, for the time being, two alternatives only. To compare them under each performance measure the AHP methods of the previous sections are used. Therefore, two 2 x 2 matrices are obtained. We will consider later three alternatives and two 3 x 3 matrices under each performance measure will be built.

Let us assume that the pairwise comparisons of Table 4.6 were obtained. Two priority vectors are calculated. The first expresses the prevalence of B under cost, the second the prevalence of A under performance. In this case it can be seen that A is faster than B is cheaper, and since the two performance measures were given the same weight, we expect that A will be the preferred choice. Some easy arithmetic confirms this. In fact, A scores: 0.25 x 0.5 + 0.8 x 0.5 = 0.525; while car B scores: 0.75 x 0.5 + 0.2 x 0.5 = 0.475.

To see how rank reversal occurs, we introduce a third car C, an exact replica of A. It follows immediately that A and C are equally desirable and that the relative comparisons of C to B are the same as the comparisons between A and B. The matrices of Table 4.7

<table>
<thead>
<tr>
<th>Cost</th>
<th>A</th>
<th>B</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1/3</td>
<td>0.25</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>1</td>
<td>0.75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance</th>
<th>A</th>
<th>B</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>4</td>
<td>0.8</td>
</tr>
<tr>
<td>B</td>
<td>1/4</td>
<td>1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table 4.6: Pairwise comparisons at the alternative level (2 alternatives)
are obtained. From the eigenvectors of Table 4.7 we can see that now B is cheaper than A and C are faster. Therefore, the ranking becomes:

\[ B: 0.6 \times 0.5 + 0.11 \times 0.5 = 0.3505 \]
\[ A \text{ and } C \text{ (tie): } 0.2 \times 0.5 + 0.44 \times 0.5 = 0.32 \]

The introduction of the third alternative has caused the rank reversal between A and B, which was ranked below A in the previous example.

The result is certainly surprising but it is easy to gain further insights considering in more detail what happened. As the criteria were assigned the same weights (0.5), alternative A, prevailing by 4 under performance and losing by 3 only under cost, is preferred to alternative B. The introduction of C causes A to lose 0.36 points under the criterion Performance, while B loses 0.09 points only. In fact, being a copy of A, C is a very effective competitor in the allocation of the points under each criterion: for instance, if under a certain criterion A is getting nearly the whole available score, then C will nearly take half of its score, and B will be left nearly unmodified. Meanwhile, under the criterion where B prevails, the addition of C causes B to lose more points (0.15) than A (0.05), but this is not enough for A to recover the previous loss (0.27). The overall effect is, then, a net loss and this causes rank reversal.

Table 4.7: Pairwise comparisons at the alternative level (3 alternatives)

<table>
<thead>
<tr>
<th>Cost</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1/3</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0.6</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1/3</td>
<td>1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Performance</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0.44</td>
</tr>
<tr>
<td>B</td>
<td>1/4</td>
<td>1</td>
<td>1/4</td>
<td>0.11</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0.44</td>
</tr>
</tbody>
</table>

12 If alternative A has a score \( \alpha \), the introduction of the alternative C will multiply it by a factor \( \frac{1}{1+\alpha} \). Thus, if \( \alpha = 1 \), the score of alternative A will be nearly halved.

13 This is proved noting that the variation of the score of B under a certain criterion is given by \( \frac{\alpha^2 - \alpha}{\alpha + 1} \) and that in the case considered \( \alpha = 1 \).
An interesting observation is that the use of relative measurement leads us to a normalization which can be interpreted as a structural modification of the criteria of choice:

[...] relative measurement may be regarded as an operation which always introduces a new criterion that operates on the existing criteria by modifying their priorities.

Such a criterion is called structural criterion (Saaty, 1987, and 1986a). To show this in more detail, we note that the calculations above can be rewritten to show how the introduction of a new alternative modifies the scores of the alternatives under each criterion through a modification the weights of the criteria. For the first two alternatives only we can write:

\[
\begin{align*}
A: 0.5 \times \frac{1}{4} + 0.5 \times \frac{4}{5} &= (0.5 \times \frac{1}{4}) \times 1 + (0.5 \times \frac{1}{5}) \times 4 = 0.525 \\
B: 0.5 \times \frac{3}{4} + 0.5 \times \frac{1}{5} &= (0.5 \times \frac{1}{4}) \times 3 + (0.5 \times \frac{1}{5}) \times 1 = 0.475.
\end{align*}
\]

These equations are easily obtained from the first calculation for A and B priorities. They can be interpreted by saying that under the criterion Cost A scores 1 while B scores 3. We have to divide these values by 4 to normalize them and obtain the eigenvector. The normalization factor 4 is common to all the scores under the criterion and can be included in the weight of the criterion: this is what is meant by the parentheses. When the third alternative is added things change.

For the set of three alternatives:

\[
\begin{align*}
B: 0.5 \times \frac{3}{5} + 0.5 \times \frac{1}{9} &= (0.5 \times \frac{1}{5}) \times 3 + (0.5 \times \frac{1}{9}) \times 1 = 0.3505 \\
A \text{ and } C \text{ (tie)}: 0.5 \times \frac{1}{5} + 0.5 \times \frac{4}{9} &= (0.5 \times \frac{1}{5}) \times 1 + (0.5 \times \frac{1}{9}) \times 4 = 0.32
\end{align*}
\]

Now, again under the criterion Cost, A and C score 1 while B scores 3. Now the normalization factor is 5 and it can be included in a revised new weight whose value is 0.1. While under this criterion the change is not dramatic, things are completely different
under the criterion Performance. In fact, while under Cost the weight was divided by 5, under Performance it is now divided by 9. In this sense we can say that A’s score is always a 4; but the introduction of the new alternative has modified the weights structure, causing this score to fall from $4/5$ to $4/9$. Following this interpretation, the effect of dilution is evident.

In conclusion, when relative measurement is used, the preference for an alternative is determined by all other alternatives. In this sense the alternatives are not independent from each other. This dependence can be interpreted as a rescaling of the criteria priorities. This feature is not encountered in absolute measurement nor in Utility Theory.

### 4.4.1 Is rank reversal a flaw of the methodology?

In traditional decision theory it is known that rank reversal may occur when new criteria are introduced. Indeed, the effect is not always unwelcome, and in some cases it is even desirable. A good example is found in Forman (1986).

Consider the evaluation and ranking of employees in a small office with half a dozen or so employees. John, who is reasonably good at analysis and good at relating to customers is the only employee who is proficient in the use of personal computers. As in many offices today, the ability to use PCs well is fairly important and because of this John is evaluated as the most valuable of the employees, with Susan ranking as second most valuable. Subsequently two more employees are hired, each of whom can use PCs fairly well, but not as well as John. A new evaluation is performed and Susan is evaluated as the most valuable with John now second. This is reasonable and occurs because the allocation of value for ability with PCs has been diluted, with John getting the largest share but not as much as before.

Generalizing, we can conclude that normalization is associated with the idea of the lack or presence of a criterion. If a criterion is very common, then it is not important in making the difference among the alternatives. In the example reported above, the capability of allowing such dilution is more a strength than a weakness of the methodology.
Huber and Puto (1983) have found that the choice process can be affected in two different ways by the presence of close alternatives. The first can be described as a proportionality effect: the new item, i.e. the copy, performs as well as the original and does not substitute for it. In this sense there is no similarity effect and rank reversal is impossible. The second possible consequence (substitutability) is that the copy has a negative similarity effect on the original, making it less attractive and thus causing rank reversal. A very witty example (Saaty, 1987) is the one of the lady who slightly preferred a hat to another in the first shop she visited. When she found that the first hat was very common in other shops in town, she went back to the first shop to buy the second hat.

Of course this is not always the case; if it is felt that allowing rank reversals is inappropriate in a particular context, then AHP has to be used in its absolute mode. Forman (1986) noted that this approach makes AHP almost identical to Utility Theory. In this regard, it is stressed how the only relevant difference between the two approaches is the formulation of questions, which is carried out in the form of pairwise comparisons rather than by the use of lotteries. It is not a surprise to find that recently AHP is being used more and more often in the context of Utility Theory applications.

In conclusion, while AHP adversaries claim that rank reversal is a flaw of the algorithm, AHP supporters point out that rank reversal is not always undesirable. In the cases in which it effectively is, one should be aware that the absolute mode is the correct tool to be used. In other words, their position is that rank reversal is the result of the application of a good tool to the wrong problem. Avoiding rank reversal was a major consideration in the application of AHP in our decision making methodology.

4.5 Final remarks

The use of AHP has been criticized mainly under two aspects. The first concerns the use of the semantic 1 to 9 linear scale. The assessors were given a complete presentation of the scale and how it was used to derive priorities. No discomfort due to the use of the scale was reported nor transpired from stakeholder input.
The second criticism concerns the possibility of rank reversal when AHP is being used in its relative mode. The difficulty was overcome using AHP in its absolute mode. The technique is now very similar to standard MAUT approaches. This decision resulted in further noticeable advantages.

First, it is worth noting that the decision tree developed for this decision is a very particular case of both AHP and MAUT. In a classic relative mode AHP approach one must compare all the alternatives pairwise under all performance measures. In our case this would result in the comparison of 6 alternatives under 14 performance measures and \(6 \times 5 / 2 \times 15 = 225\) pairwise comparisons would be needed. Moreover, if new alternatives are developed, at least \(5 \times 15 = 75\) new comparisons for each new alternative introduced will have to be elicited. This is particularly important when it is noted that the entire process of selecting an alternative is iterative in nature, as explained in section 1.2.

The disadvantage of this choice is the need to build the absolute scales against which the performance measures are compared. In our case these scales are the single-attribute utility functions. The development of these is outside the focus of this thesis and yet it is extremely important for completeness' sake, since additional proportionality factors can be introduced through the single-attribute utility functions. This has the potential of altering the performance measure weights. Appendix A is dedicated to the assessment of these functions.

4.6 Elicitation forms and input needed from the stakeholders

The forms used for the elicitation of preferences are gathered in Appendix C. The input needed from the stakeholders is limited to the pairwise comparisons. It is important to explain to the assessors what is meant by “pairwise comparison” and that the elements are compared, at each stage, one against the other, one pair at a time, with respect to the top goal, an objective category, or an objective. There is no need to explain the mathematical foundations of the method unless this is explicitly requested by the same
stakeholders. One of our stakeholders did, in fact, request an explanation of the mathematics behind the method.
5 RESULTS AND DISCUSSION

This Chapter contains the prioritization results. While for simplicity’s sake the discussion is limited to the impact category level, the results are shown at every level.

5.1 Results for all the stakeholders

As explained in Chapter 2, the development of the value trees and the elicitation of preferences were carried out in the course of several workshops attended by both the stakeholders and the analysts. Following the value tree structure, the elicitation started from the objective categories to proceed through the objectives to the performance measure level. The assessment procedure was discussed at length in Chapters 3 and 4.

5.1.1 The first workshop

During the first day of the workshop, AHP was presented to the stakeholders and the evaluation sheet for the impact categories with respect to the top goal was handed out (Appendix C). Going to prioritization directly after the discussion of the decision tree was considered advantageous, because the stakeholders would have had fresh in their mind the meaning of the objective categories. However, the stakeholders were left free to return the forms either immediately or two days later, at the second session.

Overall, the stakeholders expressed their being at ease with the assessment procedure and the task did not pose any particular difficulties: two stakeholders even turned in their sheets before the end of the meeting and only a non-technical stakeholder, a late pensioner, asked for further assistance. His/her definitive rankings were obtained two months later, at the second workshop.
After the calculations, the weights were shown to the stakeholders. Most of them agreed that the procedure had reached a correct ranking and sufficiently accurate weights. Two of them (stakeholder #6 and stakeholder #11) felt the need for some fine tuning. These cases are discussed in section 5.3.

At this level, there was no need to have different questionnaires, as the objective categories were left unchanged by the value tree modification suggested by the stakeholders and discussed in section 2.4. Of course, their definition is different and this makes direct comparisons of the weights across stakeholders using different definitions a difficult matter. This is why the results are reported separately in Table 5.1, which shows for all the stakeholders the weights produced by our analysis. The top part contains the results for the stakeholders who maintained “Public Health and Safety” -under the category “Human Health and Safety” (value tree V1), while the bottom part reports the

<table>
<thead>
<tr>
<th>Category</th>
<th>Programmatic</th>
<th>Cost</th>
<th>Socioeconomic</th>
<th>Cultural</th>
<th>Environment</th>
<th>Human Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH2</td>
<td>0.022</td>
<td>0.147</td>
<td>0.021</td>
<td>0.059</td>
<td>0.376</td>
<td>0.376</td>
</tr>
<tr>
<td>SH5</td>
<td>0.032</td>
<td>0.100</td>
<td>0.042</td>
<td>0.106</td>
<td>0.195</td>
<td>0.524</td>
</tr>
<tr>
<td>Mean</td>
<td>0.027</td>
<td>0.124</td>
<td>0.032</td>
<td>0.083</td>
<td>0.285</td>
<td>0.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Programmatic</th>
<th>Cost</th>
<th>Socioeconomic</th>
<th>Cultural</th>
<th>Environment</th>
<th>Human Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHI</td>
<td>0.078</td>
<td>0.115</td>
<td>0.039</td>
<td>0.042</td>
<td>0.340</td>
<td>0.386</td>
</tr>
<tr>
<td>SH3</td>
<td>0.026</td>
<td>0.070</td>
<td>0.042</td>
<td>0.078</td>
<td>0.387</td>
<td>0.398</td>
</tr>
<tr>
<td>SH4</td>
<td>0.050</td>
<td>0.077</td>
<td>0.249</td>
<td>0.041</td>
<td>0.168</td>
<td>0.415</td>
</tr>
<tr>
<td>SH6</td>
<td>0.120</td>
<td>0.045</td>
<td>0.127</td>
<td>0.104</td>
<td>0.272</td>
<td>0.333</td>
</tr>
<tr>
<td>SH7</td>
<td>0.020</td>
<td>0.019</td>
<td>0.052</td>
<td>0.266</td>
<td>0.507</td>
<td>0.137</td>
</tr>
<tr>
<td>SH8</td>
<td>0.035</td>
<td>0.029</td>
<td>0.080</td>
<td>0.155</td>
<td>0.393</td>
<td>0.308</td>
</tr>
<tr>
<td>SH9</td>
<td>0.033</td>
<td>0.053</td>
<td>0.100</td>
<td>0.026</td>
<td>0.317</td>
<td>0.471</td>
</tr>
<tr>
<td>SH10</td>
<td>0.196</td>
<td>0.148</td>
<td>0.211</td>
<td>0.060</td>
<td>0.107</td>
<td>0.277</td>
</tr>
<tr>
<td>SH11</td>
<td>0.046</td>
<td>0.045</td>
<td>0.055</td>
<td>0.078</td>
<td>0.276</td>
<td>0.500</td>
</tr>
<tr>
<td>Mean</td>
<td>0.067</td>
<td>0.067</td>
<td>0.106</td>
<td>0.094</td>
<td>0.307</td>
<td>0.358</td>
</tr>
</tbody>
</table>

Table 5.1: Impact category rankings
weights of those stakeholders who opted to place “Public Health and Safety” under the impact category “Environment” (value tree V2).

Stakeholders affiliated with the site owner tend to give higher weight to the category “Life Cycle Cost” than other stakeholders. Moreover, the categories “Human Health and Safety” and “Environment” score the highest in both groups, a result that is not unexpected.

Since the definition of the category “Human Health” in V2 does not include the objective “Long Term Public Health”, the corresponding weights are expected to be lower than the ones derived for V1. Indeed the highest value of the weight on “Health & Safety” (52.4) is found for a stakeholder using V1. Moreover, the average value of this category is indeed higher for the first group than for the second (0.45 vs. 0.358). The same observation applies to the category “Environment”, but in this case the averages for the two groups are too close to clearly confirm that the second group shows a higher priority. In fact, the two lowest weights assigned to the objective category “Environment” are found for stakeholders adhering to the V2 structure, and four out of seven stakeholders using V2 have a “Health & Safety” weight higher than the lowest value in V1. Finally, if we consider the ratio of the two weights (“Environment” / ”Health & Safety”) for each stakeholder, we would expect higher values for the stakeholders using V2. Indeed, the highest value is found for a stakeholder using V2 and the smallest for a stakeholder using V1; despite this, the next to highest ratio is found for the only other stakeholder using V1.

In this light, it is difficult to identify clear trends, especially if it is noted that two stakeholders only are using V1. Also, it may be worth noting that most stakeholders filled out the prioritization questionnaires using V1 to define the categories and later, when at the second workshop they were asked if V2 captured better their perception of the problem, chose to switch. When asked if they needed to revise the weights, they did not feel to repeat the assessments. Evidently, the stakeholders who switched felt that they had always been using the definitions of the objective categories adopted for V2.
5.1.2 The second workshop

The whole workshop was designed to assess the weights at the objective and performance measure level. Two forms were prepared (Appendix C), one for each of the two different value trees. The only difference is under the "Environment" and "Human / Worker Health and Safety" objective categories. At the beginning of this workshop, the team briefly reviewed the two value trees, going over their differences again. After a review of each objective category, the stakeholders were asked to carry out the pairwise comparisons.

The results are shown:

- at the objective level, with respect to the objective categories, in Table 5.2;
- at the performance measure level, with respect to the objective level, in Table 5.3 for V1 and in Table 5.4 for V2.
- at the performance measure level, with respect to the top goal, in Table 5.5 for V1 and in Table 5.6 for V2.
- for each stakeholder trees such as those in Figure 2.3 and Figure 2.4 were drawn (Appendix D). In these are shown both the absolute and the relative weights (as obtained at each level, in parentheses).

Not all the stakeholders present at the first workshop attended the second. Moreover, while all the stakeholders present at the first workshop turned in their sheets, not all of the participants in the second did: this explains the smaller number of stakeholders whose results are reported. By the end of the second workshop the elicitation of the weights \( w_i \) of eq. (4) was completed.
<table>
<thead>
<tr>
<th>Objective Category</th>
<th>Objective</th>
<th>SH1</th>
<th>SH3</th>
<th>SH4</th>
<th>SH6</th>
<th>SH8</th>
<th>SH9</th>
<th>Objective</th>
<th>SH2</th>
<th>SH5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programmatic Assumptions</td>
<td>Minimize waste</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Minimize waste</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Life Cycle Cost</td>
<td>Minimize direct costs</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Minimize direct costs</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Socioeconomic Issues</td>
<td>Promote community quality of life</td>
<td>0.125</td>
<td>0.167</td>
<td>0.2</td>
<td>0.75</td>
<td>0.167</td>
<td>0.8</td>
<td>Promote community quality of life</td>
<td>0.167</td>
<td>0.875</td>
</tr>
<tr>
<td></td>
<td>Promote environmental justice</td>
<td>0.875</td>
<td>0.833</td>
<td>0.8</td>
<td>0.25</td>
<td>0.833</td>
<td>0.2</td>
<td>Promote environmental justice</td>
<td>0.833</td>
<td>0.125</td>
</tr>
<tr>
<td>CAH resources</td>
<td>Protect CAH resources</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Protect CAH resources</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Environment</td>
<td>Protect environmental resources</td>
<td>0.5</td>
<td>0.9</td>
<td>0.25</td>
<td>0.25</td>
<td>0.2</td>
<td>0.2</td>
<td>Protect environmental resources</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Minimize risk to public health &amp; safety</td>
<td>0.5</td>
<td>0.1</td>
<td>0.75</td>
<td>0.75</td>
<td>0.8</td>
<td>0.8</td>
<td>Minimize long term risk to public health &amp; safety</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>Human / Worker Health &amp; Safety</td>
<td>Minimize risk to worker health &amp; safety</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Minimize short term risk to public &amp; worker health &amp; safety</td>
<td>0.5</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Table 5.2: Priority weights of the objectives with respect to their corresponding objective categories
<table>
<thead>
<tr>
<th>Category</th>
<th>Objective</th>
<th>Performance Measure</th>
<th>SH 2</th>
<th>SH 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programmatic Assumptions</td>
<td>Minimize waste</td>
<td>Quantity of waste transported</td>
<td>0.685</td>
<td>0.618</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantity of process waste generated</td>
<td>0.234</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantity of ER waste generated</td>
<td>0.081</td>
<td>0.297</td>
</tr>
<tr>
<td>Life Cycle Cost</td>
<td>Minimize direct costs</td>
<td>Implementation costs</td>
<td>0.25</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Completion costs</td>
<td>0.75</td>
<td>0.833</td>
</tr>
<tr>
<td>Socioeconomic Issues</td>
<td>Promote community quality of life</td>
<td>Impact on local economy</td>
<td>0.167</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Promote environmental justice</td>
<td>Changes in ambient condition</td>
<td>0.833</td>
<td>0.75</td>
</tr>
<tr>
<td>Cultural,Archaeological &amp; Historic (CAH) Resources</td>
<td>Protect CAH resources</td>
<td>Compare total population health effects</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Environment</td>
<td>Protect environmental resources</td>
<td>Contaminant concentration</td>
<td>0.167</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Changes in resources</td>
<td>0.833</td>
<td>0.833</td>
</tr>
<tr>
<td>Human Health &amp; Safety</td>
<td>Minimize long term risk to public health &amp; safety</td>
<td>Individual health risk</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Minimize short term risk to public &amp; worker health &amp; safety</td>
<td>Individual worker health risk</td>
<td>0.5</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Individual public health risk</td>
<td>0.5</td>
<td>0.875</td>
</tr>
</tbody>
</table>

Table 5.3: *Priority weights of the performance measures with respect to their corresponding objectives for VI*
<table>
<thead>
<tr>
<th>Category</th>
<th>Objective</th>
<th>Performance Measure</th>
<th>SH1</th>
<th>SH3</th>
<th>SH4</th>
<th>SH6</th>
<th>SH8</th>
<th>SH9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programmatic Assumptions</td>
<td>Minimize waste</td>
<td>Quantity of waste transported</td>
<td>0.779</td>
<td>0.200</td>
<td>0.665</td>
<td>0.333</td>
<td>0.167</td>
<td>0.413</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantity of process waste generated</td>
<td>0.161</td>
<td>0.200</td>
<td>0.245</td>
<td>0.333</td>
<td>0.667</td>
<td>0.327</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantity of ER waste generated</td>
<td>0.060</td>
<td>0.600</td>
<td>0.090</td>
<td>0.333</td>
<td>0.167</td>
<td>0.009</td>
</tr>
<tr>
<td>Life Cycle Cost</td>
<td>Minimize direct costs</td>
<td>Implementation costs</td>
<td>0.167</td>
<td>0.833</td>
<td>0.25</td>
<td>0.333</td>
<td>0.5</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Completion costs</td>
<td>0.833</td>
<td>0.167</td>
<td>0.75</td>
<td>0.666</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>Socioeconomic Issues</td>
<td>Promote community quality of life</td>
<td>Impact on local economy</td>
<td>0.1</td>
<td>0.167</td>
<td>0.8</td>
<td>0.25</td>
<td>0.667</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Changes in ambient condition</td>
<td>0.9</td>
<td>0.833</td>
<td>0.2</td>
<td>0.75</td>
<td>0.333</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Promote environmental justice</td>
<td>Compare total population health effects</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cultural, Archaeological &amp; Historic (CAH) Resources</td>
<td>Protect CAH resources</td>
<td>Number impacted / severity impacted</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Environment</td>
<td>Protect environmental resources</td>
<td>Contaminant concentration</td>
<td>0.9</td>
<td>0.875</td>
<td>0.167</td>
<td>0.75</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Changes in resources</td>
<td>0.1</td>
<td>0.125</td>
<td>0.833</td>
<td>0.25</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Minimize risk to public health &amp; safety</td>
<td>Individual health risk</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Worker Health &amp; Safety</td>
<td>Minimize risk to worker health &amp; safety</td>
<td>Individual worker health risk</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.4: Priority weights of the performance measures with respect to their corresponding objectives for V2
<table>
<thead>
<tr>
<th>Category</th>
<th>Objective</th>
<th>Performance Measure</th>
<th>SH2</th>
<th>SH5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programmatic Assumptions</td>
<td>Minimize Waste</td>
<td>Quantity of Waste Transported</td>
<td>0.015</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantity of Process Waste Generated</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quantity of ER Waste Generated</td>
<td>0.002</td>
<td>0.009</td>
</tr>
<tr>
<td>Life Cycle Cost</td>
<td>Minimize Direct Costs</td>
<td>Implementation Costs</td>
<td>0.037</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Completion Costs</td>
<td>0.110</td>
<td>0.083</td>
</tr>
<tr>
<td>Socioeconomic Issues</td>
<td>Promote Community Quality of Life</td>
<td>Impact on Local Economy</td>
<td>0.0007</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Changes in Ambient Condition</td>
<td>0.0033</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Promote Environmental Justice</td>
<td>Compare Total Population Health Effects</td>
<td>0.017</td>
<td>0.005</td>
</tr>
<tr>
<td>Cultural, Archaeological &amp; Historic (CAH) Resources</td>
<td>Protect CAH Resources</td>
<td>Number Impacted / Severity Impacted</td>
<td>0.059</td>
<td>0.106</td>
</tr>
<tr>
<td>Environment</td>
<td>Protect Environmental Resources</td>
<td>Contaminant Concentration</td>
<td>0.063</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Changes in Resources</td>
<td>0.313</td>
<td>0.162</td>
</tr>
<tr>
<td>Human Health &amp; Safety</td>
<td>Minimize Long Term Risk to Public Health &amp; Safety</td>
<td>Individual Health Risk</td>
<td>0.188</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>Minimize Short Term Risk to Public &amp; Worker Health &amp; Safety</td>
<td>Individual Worker Health Risk</td>
<td>0.094</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Individual Public Health Risk</td>
<td>0.094</td>
<td>0.344</td>
</tr>
</tbody>
</table>

Table 5.5: Performance measure weights for stakeholders using V1. In parentheses, the ranking of the PMs.
Table 5.6: Performance measure weights for stakeholder using V2. In parentheses the PM ranking.

5.2 The inconsistency index as a debugging tool.

The workshop experience showed that the stakeholders were both pleased and impressed by having their questionnaires processed directly at the meeting. All that is needed is software capable of handling matrices. The computation itself is certainly not a challenge, but errors in data input are possible, especially when several questionnaires
have to be processed under time pressure. To avoid this as much as possible it is advisable to use software providing user-friendly data input interfaces and to have more people double check what is being typed.

Despite these precautions, an error occurred in the calculation for stakeholder #2. Her comparison between the categories “Cultural” and “Socioeconomic” was “strongly” (i.e. 7) in favor of the former, but the inverse was hit. This is shown in the positions $a_{65}$ and $a_{56}$ of Table 5.7. Of course, the error was not immediately realized. The analyst’s attention was drawn by the inconsistency ratio, which was 0.27: this suggested further investigation.

The first test performed was the plot of the normalized rows. Figure 5.1 was produced. In this figure each line represents a row of the matrix of the pairwise

![Figure 5.1: Row plot for the wrong input. The dashed line is the principal eigenvector. We five, rather than six lines are seen because two are perfectly overlapped.](image)

<table>
<thead>
<tr>
<th>SH2</th>
<th>Program</th>
<th>LCC</th>
<th>Environ</th>
<th>HH&amp;S</th>
<th>Socioec</th>
<th>Cultural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>1</td>
<td>1/9</td>
<td>1/9</td>
<td>1/9</td>
<td>1</td>
<td>1/5</td>
</tr>
<tr>
<td>LCC</td>
<td>9</td>
<td>1</td>
<td>1/7</td>
<td>1/7</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Environ</td>
<td>9</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>HH&amp;S</td>
<td>9</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Socioec</td>
<td>1</td>
<td>1/9</td>
<td>1/9</td>
<td>1/9</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Cultural</td>
<td>5</td>
<td>1/7</td>
<td>1/9</td>
<td>1/9</td>
<td>1/7</td>
<td>1</td>
</tr>
</tbody>
</table>
comparisons. The x axis is associated to the columns and the y axis shows the matrix element $a$ divided by the row sum. In the consistent case we would see a perfect superimposition of six lines. When small inconsistencies are foreseen, a perfect overlap is not expected, yet the six lines should be close to each other and show a common trend. In this case, however, Figure 5.1 shows that two lines behave very differently from the other four in the last two positions (columns 5 and 6). While the general trend is to have for column 6 a value lower than for column 5, the two suspected lines show the opposite. This is what is expected when an inversion occurs.

To understand this point better, let us have a look at Table 5.7. This table reports the matrix of the pairwise comparisons as was typed, that is with the wrong entries (shaded cells). If there is an inversion, we should swap the two shaded entries. It is easy to see that this would result in inverting the slope of the two lines under investigation, which would then show a final descent, consistently with other lines. The computer data were thus checked with the questionnaire and the error was found and corrected. The inconsistency index fell to 0.18. The result in terms of a row plot is shown in Figure 5.2. Here the more consistent behavior of the lines$^{14}$ is evident.

---

$^{14}$ A second problem is seen at position 2. This was a partial “saturation” effect. For further details see section 0.
5.3 Discussion of results and stakeholder interaction

In this section, the elicitation results shown above are further discussed. The second step of the prioritization procedure is not only a test of result robustness. Through interaction with the stakeholders fine tuning is being pursued.

The representative from the County Environmental Health Department, stakeholder #11, provided the assessments summarized in Table 5.8. The calculated weights for the impact categories are reported in the same table. When he was shown these results, the stakeholder suggested that the category “Life Cycle Cost” should be ranked fourth, i.e., above the “Socioeconomic Impacts” and “Programmatic Assumptions” categories. The stakeholder added, however, that this was not a major concern to him, because the weights of these three categories should be, as the results showed, very close to each other.

We pointed out to the stakeholder that his suggestion was inconsistent with his earlier input. Indeed, as the matrix in Table 5.8 shows, in the direct comparison between “Life Cycle Cost” and “Programmatic Assumptions”, the stakeholder had given priority to the former (the corresponding matrix entry, $a_{ij}$, equals 3, indicating that this stakeholder considered LCC as weakly more important than “Programmatic Assumptions”). This stakeholder’s fairly high inconsistency ratio, 0.2, drew the analyst’s

<table>
<thead>
<tr>
<th>SH11</th>
<th>Environ</th>
<th>Cultural</th>
<th>HH&amp;S</th>
<th>Socioec</th>
<th>Program</th>
<th>LCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environ</td>
<td>1</td>
<td>7</td>
<td>1/7</td>
<td>7</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Cultural</td>
<td>1/7</td>
<td>1</td>
<td>1/5</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>HH&amp;S</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Socioec</td>
<td>1/7</td>
<td>1/3</td>
<td>1/5</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Program</td>
<td>1/7</td>
<td>1</td>
<td>1/5</td>
<td>1</td>
<td>1</td>
<td>1/3</td>
</tr>
<tr>
<td>LCC</td>
<td>1/9</td>
<td>1/3</td>
<td>1/7</td>
<td>1/3</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.8: Matrix of pairwise comparisons and weights for stakeholder #11.
attention and a more detailed scrutiny of the stakeholder’s stated preferences was undertaken.

Considering the first four entries of the last two rows of the matrix in Table 5.8 indicates that the category “Programmatic Assumptions” always dominates “Life Cycle Cost”. In fact, in the second entry, “Life Cycle Cost” is judged to be three times less important than the category “Cultural”, which is judged as important as “Programmatic Assumptions.” If the stakeholder were consistent, this would imply that the category “Programmatic Assumptions” should be three times as important as “Life Cycle Cost”. As this information was repeated in every column (except, of course, for the direct comparison), it is not surprising that the AHP ranked “Life Cycle Cost” last. Thus, the stakeholder’s desire to rank “Life Cycle Cost” higher is not to be ascribed to a flaw in the algorithm, but to this stakeholder’s inconsistency. This is an example of how the elicitation of redundant information can be exploited to analyze consistency.

Having identified the problem with the last two rows, it seems reasonable to investigate what happens when these inconsistencies are corrected. Thus, we produce

---

15 As a matter of fact, it is possible to verify that dropping one of the other categories (that is, dropping from the matrix the corresponding column and row) is sufficient to restore the proper order.
Table 5.10: Preferences of Stakeholder #6

<table>
<thead>
<tr>
<th>SH6 (I)</th>
<th>HH&amp;S</th>
<th>Environ</th>
<th>Socioec</th>
<th>Cultural</th>
<th>Program</th>
<th>LCC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>0.333</td>
<td>0.272</td>
<td>0.137</td>
<td>0.114</td>
<td>0.100</td>
<td>0.045</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SH6 (II)</th>
<th>HH&amp;S</th>
<th>Environ</th>
<th>Socioec</th>
<th>Program</th>
<th>Cultural</th>
<th>LCC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>0.333</td>
<td>0.272</td>
<td>0.127</td>
<td>0.120</td>
<td>0.104</td>
<td>0.045</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.9, which differs from Table 5.8 in that the last two rows (and two columns) have been interchanged. In this case, the inconsistency ratio is improved to 0.15. The ranking is now exactly the ranking that the stakeholder suggested.

Another interesting insight came from stakeholder #6, whose ranking was the one shown in Table 5.10(I). After reviewing this result, the stakeholder did not agree and directly expressed the preferences of Table 5.10(II). The major difference is the reversal in the ranking of the categories “Programmatic Assumptions” and “Cultural”. Although in the questionnaire itself the direct comparison was in favor of the “Cultural” category, as shown in the first weight vector of Table 5.10(I), the stakeholder was surprised that she had done so. She, then, proceeded to change the weights to the new set shown in Table 5.10(II). This is a good example of direct fine tuning of the results by the stakeholder. The AHP analysis gives a first set of preferences, which the stakeholder is free to adjust after reconsideration of the original pairwise comparisons.

In the next example, unlike preceding two examples, the interaction with the stakeholder was initiated by the analysts (i.e. us), after we noticed a high inconsistency ratio. Table 5.11 shows the results for stakeholder #1. The inconsistency ratio is 0.19. If the category “Life Cycle Cost” were not considered, the inconsistency ratio would drop dramatically to 0.0074. This is a very valuable hint which points to a major source of inconsistency. The matrix shows that the category “Environment” is “very strongly” more important than the category “Cultural” (corresponding entry $a_{12} = 7$) and equally important as “Life Cycle Cost” ($a_{16} = 1$). Entry $a_{26} = 5$, however, indicates that the
Table 5.11: Matrix of pairwise comparisons for SH1

<table>
<thead>
<tr>
<th></th>
<th>Environ</th>
<th>Cultural</th>
<th>HH&amp;S</th>
<th>Socioec</th>
<th>Program</th>
<th>LCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environ</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Cultural</td>
<td>1/7</td>
<td>1</td>
<td>1/9</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>HH&amp;S</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>7</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Socioec</td>
<td>1/5</td>
<td>1</td>
<td>1/7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Program</td>
<td>1/9</td>
<td>1</td>
<td>1/9</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>LCC</td>
<td>1</td>
<td>1/5</td>
<td>1/9</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

category “Cultural” is “strongly” more important than “Life Cycle Cost,” a preference that contradicts the preceding preferences. When the stakeholder was told of this, she attributed it to a trivial error in filling out the questionnaire and asked to change the last entry. As the stakeholder wished to make other changes as well, she was given a new form at the following workshop. This is a good example of what was observed about the revision of judgment in Chapter 4. Getting more familiar with the questionnaire, the stakeholder improved her consistency to 0.14.

5.4 Extreme inconsistency

Table 5.12 shows the most inconsistent matrix that was produced (stakeholder #7). The stakeholder agreed with the results, but the inconsistency ratio, 0.26, was exceptionally high. This is not surprising, if we note that, while in its second column the objective categories “Human Health & Safety” and “Socioeconomic Impacts” seem to be equally important, each being seven times less important than the category “Cultural”, their direct comparison records a 9 ($a_{34} = 9$) in favor of “Human Health & Safety,” whereas a consistent judgment would require $a_{34} = 1$.

A second reason for the high inconsistency ratio of the complete matrix is found in the last two columns: it is evident that the categories “Environment”, “Cultural,” and “Human Health & Safety” should be of equal importance. Despite this, we find in the
Table 5.12: *Matrix of pairwise comparisons for the most inconsistent questionnaire (SH7)*

third column that the category “Cultural” is “very strongly” more important than “Human Health and Safety”, while “Environment” is “absolutely” more important than “Human Health and Safety”.

This is a case of “scale saturation”. The categories “Environment”, “Cultural,” and “Human Health and Safety” are preferred absolutely when compared to categories perceived as much less important (“Programmatic Assumptions” and “Life Cycle Cost”). This is expressed with the highest possible score \((a_{13} = a_{14} = a_{15} = 9)\) resulting in a “saturation” of the scale, which makes all these categories appear as equally important. However, this does not appear to be true when these “important categories” are compared among themselves (for instance, \(a_{12} = 5\) in lieu of 1), thus leading to inconsistency.

Another surprising result is that this stakeholder ranks the category “Life Cycle Cost” last, even though he is a representative of the site owner.

A solution could be the use of a wider scale, i.e. not limited to 9. Anyway, as a second, tailored, questionnaire confirmed, it was expected that the stakeholder would have behaved as before, using again the highest possible scores. Therefore, backed by the fact that the stakeholder agreed, despite the high consistency ratio, that the weights were a good representation of his perceptions, we chose a second, more obvious solution:
A higher than usual inconsistency ratio may result because of the extreme judgments necessary. If one recognizes this as the cause, [...] one can accept the inconsistency ratio even though it is greater than 10%. (Forman, 1986).

The discussion of these results among the analysts brought up the general attitude of this stakeholder. There was unanimous agreement that he had not taken the whole exercise seriously and that he seemed to be more interested in challenging the analysts than in participating in the process. This was confirmed later, when the stakeholder stopped coming to the meetings, depriving us of the opportunity to work with him to revise his assessments.

5.5 Whom do the stakeholders represent?

During the elicitation, a question was raised as to whether the stakeholders should represent themselves or their organizations, since their personal views on some matters did not necessarily coincide with those of their organizations. A stakeholder even claimed that she could represent at least three different perspectives: the private citizen’s, the employee’s, and the employer’s. It was decided that a stakeholder should represent his/her institution. Accordingly, it was expected that, if an institution were represented by different representatives at the workshops, the weights obtained from their questionnaires would look very similar. This was confirmed in our case study.

The County Environmental Health Department was represented by a new person at

<table>
<thead>
<tr>
<th></th>
<th>HH&amp;S</th>
<th>Environ</th>
<th>Cultural</th>
<th>LCC</th>
<th>Socioec</th>
<th>Program</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>0.513</td>
<td>0.274</td>
<td>0.074</td>
<td>0.059</td>
<td>0.052</td>
<td>0.028</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>HH&amp;S</th>
<th>Environ</th>
<th>Cultural</th>
<th>LCC</th>
<th>Socioec</th>
<th>Program</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>0.398</td>
<td>0.387</td>
<td>0.078</td>
<td>0.070</td>
<td>0.042</td>
<td>0.025</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.13: *Comparison of rankings for the two representatives.*
the second workshop. To ensure consistency, this person was asked to fill out the same AHP questionnaires that his colleague had completed at the first workshop two months earlier. The first representative was stakeholder #11 whose final set of weights is shown in Table 5.9. Table 5.13 shows the weights of both stakeholders. Even though the actual weights are different, the overall ranking of the categories by the two stakeholders is the same. No other stakeholders have the same ranking (see Table 7.2).

5.6 Setting a limit for the inconsistency value

In the discussion of the inconsistency ratio it has been pointed out that no a priori boundary is set to identify inconsistent cases. It was concluded that each application has a particular value that should be fixed by the analyst.

To solve this problem the inconsistency ratios were collected in Table 5.14. From the analysis of the stakeholder ratios it seemed reasonable to assume as "limit warranting further investigation" 0.15. Four of the ten stakeholders show an inconsistency ratio higher than this limit. The first concern was to investigate the effects of the time dedicated to fill out the pairwise comparison questionnaires. A connection between time and consistency was identified. In fact, of these four questionnaires, two were returned at the end of the first session, despite the fact that the stakeholders were given two days to perform the assessments; one of these is the most inconsistent, and it has already been discussed in section 5.4. The third one was returned by a stakeholder who had to leave early during the second session; the last one is the one of the stakeholder who repeated the assessment at the following workshop.

Of some interest can be the comparison with the inconsistency ratios obtained by

<table>
<thead>
<tr>
<th>Stakeholders</th>
<th>0.26</th>
<th>0.20</th>
<th>0.19</th>
<th>0.18</th>
<th>0.11</th>
<th>0.11</th>
<th>0.10</th>
<th>0.10</th>
<th>0.05</th>
<th>0.05</th>
<th>0.14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysts</td>
<td>0.29</td>
<td>0.24</td>
<td>0.18</td>
<td>0.16</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td></td>
<td></td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 5.14: Summary of inconsistency ratios
the analysts, who filled out the questionnaires in a dry-run. It is a constant pattern that the stakeholders are more consistent than the analysts, who quickly answered the questionnaires.

From the examples above, it is clear that a careful use of the consistency ratio can be of great help in understanding stakeholder input and making sure that it is being correctly interpreted. A few feedback interactions are obviously necessary to achieve successful stakeholder involvement. In this perspective, a means of focusing the efforts on the weak points of the evaluation is of fundamental importance. The experience of the workshop supports what is suggested in the literature: the use of the inconsistency ratio can be very effective.
6 CONCLUSIONS

In this work we presented, through the application to an Environmental Management problem, a simple and yet very general method for assessing stakeholder priorities. The basic approach to the problem was that no mathematical method could be relied upon to produce accurate weights for the impact categories. Most of the proposed methods in the literature have advantages and disadvantages. A combination of the Analytic Hierarchy Process (AHP) with value trees, influence diagrams (ID) and multi-attribute utility theory (MAUT) principles was used to rank the decision alternatives. More precisely, AHP was used to elicit stakeholder priority weights for the various objective categories, objectives and performance measures identified by the preliminary problem structuring. The priorities obtained from the calculations had to be checked and confirmed by each stakeholder. The methodology is designed to be a practical and reliable working tool.

Practicality was addressed proposing AHP as basic weight elicitation technique. Its systematic treatment of inconsistencies and the simplicity of the questions needed for the assessment were key to its selection among several candidates. Regarding the second issue, reliability, AHP is nowadays an established technique, particularly suited to hierarchically structured problems. Our general philosophy was to use the mathematical techniques only to generate, from stakeholder input, successively more refined approximations in an iterative process that always gives the last word to the stakeholder. This seemed to be the best warranty of reliability.

As any technique, AHP must be used with full awareness of all its aspects. In particular, we were concerned with some limitations that may occur when rank reversal must be avoided and when the 1 to 9 scale is used with the semantic equivalents of Table 3.2 (Holder, 1990). The former issue was resolved by using AHP in its absolute mode;
the latter by focusing interactions with the stakeholders on potential problem areas as
identified by the inconsistency ratio.

It is our experience that nothing as time-effectiveness is more appreciated by the
stakeholders. In this perspective, it is fundamental to implement practical tools to focus
the efforts on weak points almost in real time. Any matrix manipulation and imaging
software coupled with the use of the inconsistency ratio is a practical alternative. Our
direct experience indicates that the use of the inconsistency ratio can be very effective not
only in the identification of clerical errors, but even in disclosing more hidden aspects,
such as the suitability of the 1 to 9 scale, and in giving hints about the accuracy and
accountability of stakeholders.

In some cases the public initial attitude towards what are felt as purely
mathematical abstractions is very sceptic. The insights gained during the process were
useful in defending the computation results with very comprehensible, pertinent and
down-to-earth arguments, greatly enhancing the participant’s trust in the analysts. The
more technical stakeholders asked questions about the mathematical foundations of the
AHP, while the non-technical stakeholders did not particularly care about the
mathematics. However, they all seemed to be impressed by the fact that we were able to
point inconsistencies in their inputs which they also recognised as such and they were
willing to modify.

We found that the results were generally satisfactory to the stakeholders. We also
found that the inconsistency ratio was very valuable to the analysts in discussing the
results with the stakeholders and pointing out to them inconsistencies between their
pairwise comparisons and their overall rankings of the impact categories.
7 EPILOGUE

"QVO USQVE TANDEM ABUTERE, CATILINA, PATIENTIA NOSTRA?"16

The focus of this thesis is on the quantitative prioritization of performance measures. In the preceding Chapters it was discussed how the problem is born, its theoretical foundations, the context of our particular case and the tools used to tackle its specific challenges. The tools adapted were carefully scrutinized and their potential limitations and weaknesses were investigated. Finally, the results obtained were presented along with an extensive comment of their significance. Of course, these data were of the foremost importance in the integration process; yet it is beyond the scope of this thesis to explain in detail how they were used during the culmination of the process, that is the deliberative part.

However, it would be unfortunate if a reader patient enough to reach this point were left in a state of frustrated apprehension about what happened to the stakeholders and the CWL. In this Epilogue the results of integration are shown and a very brief review of how these lead to the final decision is given. This will provide further details that go past the bare prioritization, but are fundamental in understanding the process as a whole. The reader interested in the integration and rank analysis process will find very interesting the work of Zio (1997), while Apostolakis and Pickett is the reference point for all the aspects related to the deliberative part of the process.

---

16 "In heaven’s name, Catiline, how long will you abuse our patience?", Cicero, M. T., In Catilinam I, translated by Louis E. Lord, Cicero - the speeches with English translation, Harvard University Press, Cambridge, Massachusetts, MCMLXIV.
7.1 Obtaining the PI

As it was explained in section 3.1, to aggregate the evaluations of the performance measure impacts of a given alternative into the Performance Index PI, the priority weights generated by AHP are combined with properly constructed single-attribute utility functions as from eq. (4), which is repeated here for the reader's convenience:

\[ PI^j = \sum_{i=1}^{N} w_i \, u_i(x_i^j) \]  

In this equation:

- the performance measure weights \( w_i \)'s, were obtained from our analysis and are reported in Table 5.5 and Table 5.6
- the functions \( u_i \)'s were obtained during a third workshop. The procedure is summarized in Appendix A
- the functions \( u_i \)'s are evaluated at the performance measure values \( x_i \)'s that is the consequences of the performance measures resulting from the implementation of the various RAAs. The \( x_i \)'s are provided by the risk assessments.

Eq. (4) is then used to evaluate the expected value of the PI for each alternative and each stakeholder. At the end of the process six PIs are calculated for each stakeholder. The associated uncertainties are assessed by straightforward application of the variance formulas applicable to a linear transformation of the many random variables \( u_i \). The PI uncertainty bands are calculated exploiting the 20 realizations of \( x_i \) calculated in the risk assessments for each performance measure.
7.2 Ranking analysis

The PIs are first analyzed stakeholder by stakeholder. Ranking the PIs is equivalent to ranking the alternatives: the higher the PI the more desirable the alternative. These rankings can be reported in error bar charts or in the form of tables. As an example, the results for stakeholder #3 are presented in Figure 7.1 and Table 7.1.

All these rankings are thoroughly analyzed. Three major tools are used to this end (Zio, 1997):

1. **Major contributor analysis**: for each alternative the PI is broken in the single terms of the sum that defines it. In other words the products $w_i \cdot u_i$ are compared against each other to identify strength and weaknesses of the alternatives. Moreover, the product can be analyzed in terms of its factors, $w_i$ and $u_i$, to understand the influence of personal weights and actual performance on the result.

2. **Sensitivity analysis**: it investigates the stability of the ranking and its dependence on the weights at the objective category, objective and performance measure level.

3. **Uncertainty analysis**: it investigates the effect of uncertainties as estimated by risk assessments on the rankings. This helps the intuition of “distance" between the alternatives. In fact, if two alternatives are separated by a utility difference small with respect to the uncertainty attached to the mean values, it could be meaningless to prefer one to the

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>D</th>
<th>C</th>
<th>B</th>
<th>E</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranking</td>
<td>1</td>
<td>2</td>
<td>.3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Lower</td>
<td>.1659</td>
<td>.1697</td>
<td>.1633</td>
<td>.1387</td>
<td>.1239</td>
<td>.052</td>
</tr>
<tr>
<td>Higher</td>
<td>.1957</td>
<td>.1875</td>
<td>.1909</td>
<td>.1699</td>
<td>.1409</td>
<td>.0902</td>
</tr>
<tr>
<td>Total Utility (- constant)</td>
<td>.1808</td>
<td>.1786</td>
<td>.1771</td>
<td>.1543</td>
<td>.1324</td>
<td>.0711</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>.0149</td>
<td>.0089</td>
<td>.0138</td>
<td>.0156</td>
<td>.0085</td>
<td>.0191</td>
</tr>
</tbody>
</table>

Table 7.1: Total utility mean values and uncertainty bands
other. For instance, in Figure 7.1 it is seen that the first three alternatives are very close to each other. Moreover, if this stakeholder is risk averse, it could very well be that he will prefer alternative D, which has a slightly lower expected value but smaller uncertainties, to alternative F, for which is expected a better result but with a significantly higher uncertainty. This analysis is of great help in visualizing that the rankings are far from being well defined.

The ranking analysis identifies the reasons for and against each alternative. This provides the analysts with valuable insights which are especially useful in the perspective of deliberation (Apostolakis and Pickett, 1997).

Once the rankings of the stakeholders are well understood, it is interesting to analyze the rankings across the stakeholders. Table 7.2 was prepared to anticipate agreements and disagreements. As a general comment, it was seen that for most stakeholders, the expected PIs are quite close and do not determine outstanding dominances.

An immediate result is that four of the stakeholders consider the no action alternative (F) as the most preferred one, while the other two show a slight preference for alternative E. Observation such as these are formulated by the analysts and are used in the course of deliberation to stimulate the discussion and lead the debate towards consensus,
starting from the least controversial points. In our case six tentative conclusions were prepared:

1. A is the least preferred alternative
2. C is neither strongly disliked nor liked by the stakeholders
3. D is less preferred than C by all except possibly one stakeholder
4. F is a candidate to be the preferred alternative
5. E is a candidate to be recommended
6. F and E are the two preferred options

Initially, RAA F, no action, was included for comparison purposes, but it turned out to be, if not the best alternative, a really competitive one. The reasons were carefully
examined: they can be roughly summarized saying that for the type of contamination involved and the site hydrogeological conditions, the only way to hurt somebody was to have some workers exposed to the hazard.

The conclusion that no action was the best option, however did not obtain consensus. Indeed, it was recommended that no action was not an option. This was due to an attitude difficult to capture by mathematical means. Most people, especially the representatives of citizen groups, felt that the project had to send a message beyond the particular problem of the CWL. The recommendation had to include somehow that the Department of Energy had “to pay for its past sins” by considering remediation at all contaminated sites, as they were doing for the CWL. The stakeholders agreed that this message could have been less manifest if no action was to be recommended. This consideration was not included in the initial decision tree despite the early involvement of stakeholders. From a methodological point of view this is not a disaster, since it could be included in a second iteration of the analysis. However, due to time and resource limits, this step was not carried out for this project.

Nevertheless, the stakeholders learnt that, against intuition, no action was a very good alternative from an overall point of view. This was particularly true for metal contamination. In fact, the risk assessments made clear that no action would not have inhibited the natural process which reduces Cr$^{6+}$ to Cr$^{3+}$, ion of minimal risk. If something had to be done, it made more sense to do it on the organic contamination side. All the alternatives were divided in their parts targeting metal and organic contamination. The attention now was on the latter only. On the basis of cost and future land use, consensus was reached that an hybrid alternative, named F+, was to be the unanimous recommendation: this alternative included no action for chromium and soil vapor extraction to address TCE contamination.

It is worth stressing that these conclusions were helped by the decomposition process of major contributor analysis, especially when it came to provide a quick evaluation of which alternative gave the best results under a limited number of aspects.
APPENDICES
APPENDIX A

SINGLE-ATTRIBUTE UTILITY ELICITATION

This thesis has discussed how the weights $w_i$ of eq. (4) are obtained by use of AHP and Hierarchical Composition. Since the $x_i$'s are obtained from the risk assessments, we are left with the task of constructing the single-attribute utility functions $u_i$. This is beyond the aim of this thesis but must be included for completeness' sake. In this Appendix we summarize how this is done. The details and a discussion of the method can be found in Zio (1997).

The main idea is to perform the assessment using the intuitive concepts of "worst", "moderate" and "best" outcomes, rather than a comparison between the numerical values of these same outcomes. To this end, the stakeholders are given, for each performance measure, a range comprising all the possible outcomes of the alternatives, as found by risk assessments. The stakeholders are asked, first, to divide the performance measure ranges in three subranges of performance; and then to assess, consistently with these definitions, the relative utilities of a "worst", a "moderate" and a "best" outcome. For example, a stakeholder may consider the implementation cost of a remediation technology "worst" when it costs more than five million dollars, "moderate" if it is within one and five millions, and "best" if it is less than three millions. Note how these sets need not necessarily be exclusive.

More information than what is necessary to build linear utilities is obtained; yet, this is not sufficient to allow a perfect evaluation of the utility curve. The following procedure was developed to achieve a reasonable compromise between the conflicting needs of simplicity and accuracy.

For a given performance measure the utility function should be such that an alternative leading to an outcome which falls in the "worst" range will receive a low
utility value, to be determined in agreement with the preference of the stakeholder. Symmetrically, if the outcome for that alternative falls within the "best" range, this must be reflected by a high utility value.

Subjective concepts such as "worst", "moderate" and "best" consequences are far from being universal and clearly defined; on the contrary, it is expected that every stakeholder will have a different perception of these concepts. These considerations find a rigorous mathematical accommodation in fuzzy logic, which has defined membership functions (Klir and Yuan, 1995).

Membership functions attempt to convert to mathematical form the human perception of fuzzy concepts such as "best", "moderate", and "worst". While many values may be clearly categorized, e.g., are felt as being thoroughly defined saying that they definitely are a "best" outcome, in most cases they are not completely characterized by any of these terms. Instead, they are better described by a mix of different properties combined to different extents: for example, a cost of five million dollars can be perceived as being "quite a worst" outcome, describable as "pretty moderate", and "definitely not a best one". So, for each given term, say "moderate", we can plot (Figure A.1) on the performance measure axis the extent to which each value is perceived as having that property, i.e. to which it is considered "moderate". This is an example of membership function.

A membership function defines a fuzzy set. The set is fuzzy in that one can not say for each value whether it belongs or not to the set of moderate outcomes, as would

Figure A.1: Sample membership function
happen with a classic (crisp) set, where these two eventualities are the only possible.

From the discussion above, it is expected that values close to the ends of the subranges of definition will belong to more than one subrange. Therefore the subranges "best", "moderate" and "worse" can, and should, overlap, as it is seen in Figure A.2, were the three fuzzy sets defined for a performance measure are shown on the same axis. This set of membership functions describing different terms is an example of a linguistic variable. In our case we can define "performance measure consequence" as a linguistic variable: it consists of the fuzzy sets "worst", "moderate" and "best".

Consider now a given performance measure and a corresponding consequence value $x'$ (Figure A.3): for the hypothetical stakeholder who assessed the membership functions for "worst", "moderate" and "best" for this performance measure, an alternative resulting in a value $x'$ would belong, according to this criterion only, to the fuzzy set "best" with a membership of 0.2, to the set "moderate" with a membership of 0.6 and to the set "worst" with zero membership. To construct a utility function from this information we need to assess, consistently with these same perceptions of "best", "moderate" and "worst", by how much a best outcome is preferred to a "moderate" and a "worst" outcome. This can be done via AHP, a technique with the considerable advantage of being familiar to the stakeholders.
Figure A.3: *Membership values corresponding to a consequence* $x'$

Under each performance measure, each stakeholder will be asked to carry out three comparisons. For example, let us assume that the following were elicited from this hypothetical stakeholder:

1. Best vs. Worst 6
2. Worst vs. Moderate 3
3. Moderate vs. Best 2

The resulting matrix of pairwise comparisons is in Table A.1. The priorities in the fourth column are given by the principal eigenvector. Similarly, the ratio of two components has the interpretation of relative importance of the two corresponding terms. These priority values can be interpreted (Hughes, 1986) as single-attribute utility evaluations for values of the performance measure falling, if it were possible, in the "best", "moderate" and "worst" set only.

<table>
<thead>
<tr>
<th></th>
<th>best</th>
<th>moderate</th>
<th>worst</th>
</tr>
</thead>
<tbody>
<tr>
<td>best</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>moderate</td>
<td>1/2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>worst</td>
<td>1/6</td>
<td>1/3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table A.1: *Matrix of comparisons for the assessment of the relative importance of the terms "best", "moderate" and "worst".*
Thus far, we have a tool to translate into linguistic terms of satisfaction the performance measure values; moreover, the utility values corresponding to these linguistic expressions can be assessed. We now need to build a bridge between the utility and the performance values: this can be done very naturally at the linguistic level with fuzzy inference rules (von Altrock, 1995): their simplicity makes them self-evident.

If performance value = worst, then Utility = Utility (worst)
If performance value = moderate, then Utility = Utility (moderate)
If performance value = best, then Utility = Utility (best)

In a fuzzy logic context, though, such statements are not to be regarded "crisply" as either true or false: they can be true or false to a certain degree. In our example we have that the consequence under consideration is "worst" ($u=0.1$) with zero membership (i.e., it is not "worst" at all), "moderate" ($u=0.3$) with membership 0.6 (i.e., it is considered as "fairly moderate") and "best" ($u=0.2$) with membership 0.2 (i.e., it is "somehow best").

<table>
<thead>
<tr>
<th>Worst</th>
<th>0</th>
<th>Utility = 0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate</td>
<td>0.6</td>
<td>Utility = 0.3</td>
</tr>
<tr>
<td>Best</td>
<td>0.2</td>
<td>Utility = 0.6</td>
</tr>
</tbody>
</table>

In order to be able to proceed with a utility analysis, we now need to "defuzzify" these utilities; i.e. synthesize them in a single, crisp number. To this end, the center of maximum defuzzification rule was chosen (von Altrock, 1995). A simple justification of the method is given pointing out that the method is really a weighted average:

$$u(x') = \frac{(0.2 \times 0.6 + 0.6 \times 0.3 + 0 \times 0.1)}{0 + 0.6 + 0.2} = 0.375$$

The result of the defuzzification is a crisp number representing the single-attribute utility value corresponding to a consequence value $x'$ for the performance measure considered. If the procedure is repeated for all the values of the axis, the utility function
can be plotted (Figure A.4). The procedure can be repeated for all consequence values of all performance measures thus providing the $u_i$'s.

The extrema of the range upon which the subranges were elicited were the minimum value obtained from the risk assessments minus the 20% of the range, and the maximum value increased by the 20% of the range. This "recipe" was chosen as a compromise between two conflicting needs: covering all the possible outcomes and maintaining sufficient resolution to classify them under the three different terms. In fact, if the range is too large, all the results might fall under one class only, say "moderate", and this would make them indistinguishable. In utility terms, the utility function would be flat. The natural remedy would be the use of more than three terms in the linguistic variable, but this would result in a significant increase in the amount of information required: the use of five terms would require the elicitation of two more ranges and seven more pairwise comparisons.

The information on the subranges is used to build the membership functions. In this prototype phase of the methodology development, the simplest method was adopted. The minimum of the range given for "best" was set to a membership of one whenever the low consequence value implies a desirable outcome (e.g. Contaminant Concentration or Completion Cost). The highest value which can to some extent be considered as best, i.e.

![Figure A.4: Schematic diagram of utility function and membership function](image)

102
the maximum of the range given by the stakeholder, is set to zero and between the two
points a straight line is drawn. The opposite is true whenever a low consequence value
implies a bad outcome (e.g. Impact on Local Economy). The minimum is set to zero and
the maximum is set to one. The term “worst” was symmetrically treated. The “moderate”
range is assumed to be the base of an isosceles triangle, where the base midpoint is
assumed to have membership one. These features can be recognized in Figure A.4.

The theoretical foundations need not be presented to the stakeholders to allow
them to perform single-attribute utility assessments. Yet, it is key to explain some
implications that the definitions of these concepts have. The facilitator team should devote
a particular effort to explaining the following delicate issues, which lend themselves to
easy misunderstandings:

- The terms “best”, “moderate” and “worst” are supposed to indicate
what, within the range, is felt as “best”, “moderate” and “worst”. Therefore, the terms are defined differently for different performance
measures.

- If within the two extrema of the range no difference is felt, this will be
expressed assessing “best” and “worst” as equally desirable.

- The ranges of definition of the terms can, and should, overlap. The
reason is clear from their definition. It is important to remember that the
possibility of overlapping may appear strange to the stakeholders, who
face such forms for the first time; the facilitator team should devote
particular care in explaining this issue to the stakeholders.

- As the terms are performance-measure specific, the comparison
between “best”, “moderate” and “worst” can not be performed once and
for all. It is tempting for the same stakeholders to think so; again, the
facilitator team must explicitly bring the issue to the stakeholders’
attention.
It is important to refer to the site: the ranges are in fact determined by the actual outcomes. It may be important to calibrate the stakeholders by providing them some information regarding similar environmental restoration problems in order to help them in the assessment of how “best” compares to “worst”. This has the drawback of potentially biasing their judgments. Again it will be key for the facilitator team to reach, with the stakeholder help, the best compromise.

The innovative use of fuzzy logic brings a noticeable improvement to the elicitation/communication aspects of the methodology. In fact, while in a conventional combination of AHP and utility approaches (Hughes, 1986) we should elicit comparisons between two numerical characterizations of the alternatives, the fuzzy approach allows the use of comparisons in a linguistic, and therefore more “friendly”, fashion. In this procedure the stakeholders will be given two linguistic questionnaires rather than the purely numerical one involved in a non-fuzzy approach.

The conventional approach (lower part of Figure A.5) is more direct, but has the drawback of presenting the stakeholder with a questionnaire which deals with the “technical level” only. In such a questionnaire, in fact, we would ask the stakeholder to

---

**Figure A.5: Schematic diagram of the utility assessment process (after von Altrock, 1995)**

- Linguistic level
- Technical level
- Alternatives: linguistic assessment (PM)
- Fuzzification: range questionnaire
- Alternatives: numerical evaluation (PM value)
- Inference (fuzzy rules)
- Utility: linguistic assessment
- Defuzzification: AHP questionnaire
- Utility numerical evaluation(score)
- Conventional approach: AHP yields utilities by comparisons of numbers
compare the two extrema of the performance measure range, thus forcing the stakeholder to compare two numerical values using numbers. From this comparison, we can work out the relative priority of the two points and a utility function is obtained fitting a straight line between these points. This approach is contained in the fuzzy approach: in fact, it seems reasonable to assume that the comparison of “best” vs. “worst” will give a good approximation of what we would have obtained from the comparison between the numerical values: the information elicited with the form already provides us with what is needed in the classical approach.

In this second approach, though, a second linearity assumption (after the one underlying eq. (4)) must be introduced: it should be again stressed that curved utility functions are probably more precise representations of people’s preferences; however such curvature almost never makes any difference to the decision and strong arguments support a linear utility whenever the underlying performance measure is conditionally monotone (that is, either more is better than less or vice-versa throughout the plausible range of the performance measure, regardless of the values of the other performance measures). The availability of a second tool to construct utilities is an effective means of testing the effectiveness of the fuzzy procedure introduced.
APPENDIX B

AHP Hierarchies

In the discussion of rank reversal it was stated that relative measurement AHP hierarchies are slightly different from absolute mode hierarchies. In this Appendix the distinction is explained through an example.

In a relative measurement AHP approach, a hierarchy typically looks like the one sketched in Figure B.1. In this example, four alternatives, say A, B, C and D, are being compared under three different criteria I, II and III. We can see how all the alternatives are compared pairwise under each criterion. In light of what was exposed in section 4.4, this leads to three matrices (one for each criterion). These matrices compare 4 elements and are, therefore, 4 x 4 matrices. For each matrix the (column) eigenvector $w_i$ ($i = I, II, III$) can be computed and yields the priority vector. These eigenvectors, thus, express the relative priorities under each criterion. To obtain the overall preferences, one has to weigh these priorities with the relative weight of the relative criterion. The criteria I, II, and III are, then, prioritized by means of three pairwise comparisons, that are organized in their turn, in a matrix. Let its principal (column) eigenvector be $v$. Its first component is the priority weight of criterion I, the second the weight of criterion II and so on.

It is evident that the eigenvector $w_i$ has to be weighted by the first component on $v$, the eigenvector $w_II$ by the second component of $v$, and so on. In matrix notation the overall priorities can be expressed as $W \times v$, being $W$ the matrix whose columns are the vectors $w_i$, $w_{II}$ and $w_{III}$. This method is easily extended to an arbitrary number of tiers and is the most classic use of an AHP structure.

Figure B.1
The value tree used in our methodology is much simpler in that its branches feed in one box of the upper level only. In this sense our structure is a particular case of the more complex hierarchy shown above. But there is more.

Our methodology relies upon a modification known as absolute mode of AHP. In this approach the alternatives are not directly compared one against the other. Rather, the different values of the performance measures are compared against an absolute scale, the utility function, which associates to every outcome a score. Once the utility function is built under each performance measure, the overall utility is evaluated as the weighted sum of these single-attribute utilities. The weights of the top level are provided exactly as it was done in the previous case for v. This process can be represented as in Figure B.2.

These two sketches offer a pictorial representation of two slightly different hierarchies. This fundamental difference is found at the lower tier, where the alternatives enter the hierarchy. While in the second case the alternatives are compared to an absolute standard, in the first case they are compared against each other. It is now recognized that, while in the relative mode the ranking is dependent on the set of alternatives chosen, in the absolute mode it is not; adding or removing alternatives can not affect the final ranking, and rank reversal is therefore impossible.
APPENDIX C

ELICITATION QUESTIONNAIRES

This Appendix collects the three questionnaires used for the elicitation of the performance measure weights.

The first questionnaire was used for the prioritization of the objective categories. Note that at this level there is no need to have different forms for the two value trees. Of course, each stakeholder is giving a different meaning to the categories “Environment” and “Human Health & Safety”, but this does not affect the form. Some stakeholders wished to rename the latter category “Worker Health & Safety”. They did it correcting the questionnaire by hand.

The second and third questionnaires (Form 1 and Form 2) were used at the objective and performance measure level. They are different in that different comparisons are needed under the categories “Environment” and “Human/Worker Health & Safety”.

EVALUATION SHEET

Date:

NAME ____________________________

ORGANIZATION _____________________

• SCALE

For two elements A and B, the comparison with respect to a given element is expressed as follows:

I. Circle the one that you believe to be more dominant with respect to the objective in question
II. Write in the space indicated how strong this dominance is according to the following scale:

1: if the circled element and the other element are equally important.
3: if the circled element is weakly more important than the other element.
5: if the circled element is strongly more important than the other element.
7: if the circled element is demonstrably or very strongly more important than the other element.
9: if the circled element is absolutely more important than the other element.

Feel free to use the numbers 2, 4, 6 and 8 to express compromise between slightly differing judgments.
• RELATIVE IMPORTANCE ASSESSMENT

Compare the following with respect to the OVERALL DESIRABILITY objective

1. Socioeconomic Impacts vs. Life Cycle Cost

2. Cultural Resources vs. Socioeconomic Impacts

3. Programmatic Impacts vs. Environment

4. Life Cycle Cost vs. Human Health & Safety

5. Environment vs. Human Health & Safety

6. Cultural Resources vs. Life Cycle Cost

7. Environment vs. Life Cycle Cost

8. Socioeconomic Impacts vs. Environment

Please go on to the next page

• Key for the evaluation:
  1 equally  3 weakly  5 strongly  7 demonstrably or very strongly  9 absolutely
  Use even numbers to express compromise.
9. Cultural Resources vs. Human Health & Safety

10. Cultural Resources vs. Environment

11. Human Health & Safety vs. Programmatic Impacts

12. Programmatic Impacts vs. Life Cycle Cost

13. Human Health & Safety vs. Socioeconomic Impacts

14. Programmatic Impacts vs. Socioeconomic Impacts

15. Cultural Resources vs. Programmatic Impacts

• Key for the evaluation:
  1 equally  3 weakly  5 strongly  7 demonstrably or very strongly  9 absolutely
  Use even numbers to express compromise.
EVALUATION SHEET

Date:

NAME ____________________________

ORGANIZATION _________________

- SCALE

For two elements A and B, the comparison with respect to a given element is expressed as follows:

I. Circle the one that you believe to be more dominant with respect to the objective in question

II. Write in the space indicated how strong this dominance is according to the following scale:

1: if the circled element and the other element are equally important.

3: if the circled element is weakly more important than the other element.

5: if the circled element is strongly more important than the other element.

7: if the circled element is demonstrably or very strongly more important than the other element.

9: if the circled element is absolutely more important than the other element.

Feel free to use the numbers 2, 4, 6 and 8 to express compromise between slightly differing judgments.
- RELATIVE IMPORTANCE ASSESSMENT

Objective category:

PROGRAMMATIC ASSUMPTIONS

Compare the following performance measures with respect to the Minimize Waste objective

1. Quantity of Transported Waste vs. Quantity of Process Waste Generated

2. Quantity of ER Waste Generated vs. Quantity of Transported Waste

3. Quantity of Process Waste Generated vs. Quantity of ER Waste Generated

Key for the evaluation:

1 equally
3 weakly
5 strongly
7 demonstrably or very strongly
9 absolutely

Use even numbers to express compromise.
Objective category:
LIFE CYCLE COST

Compare the following performance measures with respect to the Minimize Direct Costs objective

1. Implementation Costs vs. Completion Costs

Key for the evaluation:

1 equally      3 weakly      5 strongly      7 demonstrably or very strongly      9 absolutely

Use even numbers to express compromise.
Objective category:
SOCIOECONOMIC ISSUES

Compare the following objectives with respect to the Socioeconomic Issues objective category

1. Promote Community Quality of Life vs. Promote Environmental Justice

Compare the following performance measures with respect to the Promote Community Quality of Life objective

1. Impact on Local Economy vs. Esthetics
2. Esthetics vs. Changes in Ambient Conditions
3. Changes in Ambient Conditions vs. Impact on Local Economy

Compare the following performance measures with respect to the Promote Environmental Justice objective

1. Compare Total Population Health Effects vs. Compare Socioeconomic Impacts

• Key for the evaluation:
  1 equally  3 weakly  5 strongly  7 demonstrably or very strongly  9 absolutely
  Use even numbers to express compromise.
Objective category:

ENVIRONMENT

Compare the following performance measures with respect to the Protect Environmental Resources objective

1. Contaminant Concentration vs. Changes in Resources

Key for the evaluation:
1 equally 3 weakly 5 strongly 7 demonstrably or very strongly 9 absolutely
Use even numbers to express compromise.
Objective category:

HUMAN HEALTH & SAFETY

Compare the following objectives with respect to the Human Health & Safety objective category

1. Minimize Long-Term Risk to Public Health & Safety vs. Minimize Short-Term Risk to Public and Worker H & S

Compare the following performance measures with respect to the Minimize Short-Term Risk to Public and Worker H & S objective

1. Individual Public Health Risk vs. Individual Worker Health Risk

Key for the evaluation:

1 equally 3 weakly 5 strongly 7 demonstrably or very strongly 9 absolutely
Use even numbers to express compromise.
EVALUATION SHEET

Date: 

NAME ________________________________

ORGANIZATION _________________________

• SCALE

For two elements A and B, the comparison with respect to a given element is expressed as follows:

I. Circle the one that you believe to be more dominant with respect to the objective in question

II. Write in the space indicated how strong this dominance is according to the following scale:

1: if the circled element and the other element are equally important.
3: if the circled element is weakly more important than the other element.
5: if the circled element is strongly more important than the other element.
7: if the circled element is demonstrably or very strongly more important than
   the other element.
9: if the circled element is absolutely more important than the other element.

Feel free to use the numbers 2, 4, 6 and 8 to express compromise between slightly differing judgments.
• RELATIVE IMPORTANCE ASSESSMENT

Objective category:

PROGRAMMATIC ASSUMPTIONS

Compare the following performance measures with respect to the Minimize Waste objective

1. Quantity of Transported Waste vs. Quantity of Process Waste Generated

2. Quantity of ER Waste Generated vs. Quantity of Transported Waste

3. Quantity of Process Waste Generated vs. Quantity of ER Waste Generated

Key for the evaluation:
1 equally 3 weakly 5 strongly 7 demonstrably or very strongly 9 absolutely
Use even numbers to express compromise.
Objective category:
LIFE CYCLE COST

Compare the following performance measures with respect to the Minimize Direct Costs objective

1. Implementation Costs vs. Completion Costs

Key for the evaluation:
1 equally 3 weakly 5 strongly 7 demonstrably or very strongly 9 absolutely
Use even numbers to express compromise.
Objective category:
SOCIOECONOMIC ISSUES

Compare the following objectives with respect to the Socioeconomic Issues objective category

1. Promote Community Quality of Life vs. Promote Environmental Justice

Compare the following performance measures with respect to the Promote Community Quality of Life objective

1. Impact on Local Economy vs. Esthetics

2. Esthetics vs. Changes in Ambient Conditions

3. Changes in Ambient Conditions vs. Impact on Local Economy

Compare the following performance measures with respect to the Promote Environmental Justice objective

1. Compare Total Population Health Effects vs. Compare Socioeconomic Impacts

Key for the evaluation:
1 equally  3 weakly  5 strongly  7 demonstrably or very strongly  9 absolutely
Use even numbers to express compromise.
Objective category:

ENVIRONMENT

Compare the following objectives with respect to the Environment objective category

1. Protect Environmental Resources vs. Minimize Risk to Public Health & Safety

Compare the following performance measures with respect to the Protect Environmental Resources objective

1. Contaminant Concentration vs. Changes in Resources

Key for the evaluation:

1 equally 3 weakly 5 strongly 7 demonstrably or very strongly 9 absolutely

Use even numbers to express compromise.
APPENDIX D

PRIORITIZED VALUE TREES

In this Appendix are reported the prioritization results in graphic format.

The diagrams for stakeholder #2 and #6 are reported in Figure 2.3 and Figure 2.4. Stakeholder #7 never completed the prioritization process beyond the objective category tier. His case is discussed in section 5.4. Stakeholder #8 and #9 left the process after prioritization, before the elicitation of the single attribute utility functions.

All stakeholders but stakeholder #2 and #5 used V2.
Minimize Overall Risk

- Programmatic Assumptions 0.078
  - Minimize Waste 0.078 (1)
    - Quantity of Waste Transported 0.061 (0.779)
    - Implementation Costs 0.019 (0.167)
    - Quantity of Process Waste Generated 0.012 (0.161)
  - Life Cycle Cost 0.115 (1)
    - Quantity of ER Waste Generated 0.005 (0.960)
    - Completion Cost 0.096 (0.833)
    - Changes in Ambient Condition 0.0045 (0.9)
  - Socioeconomic Issues 0.039
    - Impact on Local Economy 0.0085 (0.1)
    - Changes in Economic Activity 0.0125 (0.1)
  - Cultural, Archaeological & Historic (CAH) Resources 0.042
    - Protect CAH Resources 0.042 (1)
      - Number Impacted/Severity Impacted 0.042 (1)
    - Compare Total Population Health Effects 0.034 (1)
  - Environment 0.340
    - Contaminant Concentration 0.153 (0.9)
    - Changes in Resources 0.017 (0.1)
  - Worker Health & Safety 0.386
    - Minimize Risk to Worker Health & Safety 0.386 (1)
    - Individual Health Risk 0.170 (1)
    - Changes in Resources 0.017 (0.1)

STAKEHOLDER 1
Minimize Overall Risk

Programmatic Assumptions
0.050

Life Cycle Cost
0.077

Socioeconomic Issues
0.249

Cultural, Archaeological & Historic (CAH) Resources
0.041

Environment
0.168

Worker Health & Safety
0.415

Minimize Waste
0.050
(1)

Minimize Direct Costs
0.077
(1)

Promote Community Quality of Life
0.05
(0.2)

Protect Environmental Justice
0.199
(0.8)

Protect CAH Resources
0.041
(1)

Protect Environmental Resources
0.142
(0.25)

Minimize Risk to Public Health & Safety
0.126
(0.75)

Minimize Risk to Worker Health & Safety
0.415
(1)

Quantity of Waste Transported
0.033
(0.665)

Implementation Costs
0.019
(0.25)

Impact on Local Economy
0.04
(0.8)

Number Impacted/Severity Impacted
0.041
(1)

Contaminant Concentration
0.007
(0.167)

Changes in Resources
0.035
(0.833)

Quantity of Process Waste Generated
0.012
(0.245)

Completion Cost
0.058
(0.75)

Changes in Ambient Condition
0.01
(0.2)

Individual Health Risk
0.126
(1)

Individual Worker Health Risk
0.415
(1)

Stakeholder 4


Apostolakis, George E. and Pickett, Susan E., “Deliberation: Integrating Analytical Results into Environmental Decisions Involving Multiple Stakeholders”, accepted for publication in Risk Analysis.


Paulos, Todd, and Apostolakis, George E., "A Methodology to Select a Wire Insulation for Use in Habitable Spacecraft", accepted for publication in *Risk Analysis*.


