VTT TIEDOTTEITA – MEDDELANDEN – RESEARCH NOTES 1972

Evaluation of a Delphi technique based expert judgement method for LCA valuation

DELPHI II

Yrjö Virtanen & Sirpa Torkkeli
VTT Chemical Technology

Bob Wilson
Landbank Environmental Research and Consulting, London

TECHNICAL RESEARCH CENTRE OF FINLAND
ESPOO 1999

ISBN 951–38–5461–2 (soft back ed.)
ISSN 1235–0605 (soft back ed.)

ISBN 951–38–5462–0 (URL: http://www.inf.vtt.fi/pdf/)
ISSN 1455–0865 (URL: http://www.inf.vtt.fi/pdf/)

JULKAISIJA – UTGIVARE – PUBLISHER

Valtion teknillinen tutkimuskeskus (VTT), Vuorimiehentie 5, PL 2000, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 4374

Statens tekniska forskningscentral (VTT), Bergsmansvägen 5, PB 2000, 02044 VTT
tel. växel (09) 4561, fax (09) 456 4374

Technical Research Centre of Finland (VTT), Vuorimiehentie 5, P.O.Box 2000, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 4374

VTT Kemiantekniikka, Ympäristötekniikka, Tekniikantie 4 B, PL 14031, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 7043

VTT Kemiteknik, Miljöteknik, Teknikvägen 4 B, PB 14031, 02044 VTT
tel. växel (09) 4561, fax (09) 456 7043

VTT Chemical Technology, Environmental Technology,
Tekniikantie 4 B, P.O.Box 14031, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 7043

Technical editing Leena Ukskoski

Libella Painopalvelu Oy, Espoo 1999

Virtanen, Yrjö, Torkkeli, Sirpa & Wilson, Bob. Evaluation of a Delphi technique based expert judgement method for LCA valuation – DELPHI II. Espoo 1999. Technical Research Centre of Finland, VTT Tiedotteita – Meddelanden – Research Notes 1972. 125 p. + app. 17 p.

Keywords LCA, Life Cycle Assessment, Delphi method, DELPHI II, valuation, environmental impacts, statistical analysis

Abstract

Because of the complexity and trade-offs between different points of the life cycles of the analysed systems, a method which measures the environmental damage caused by each intervention is needed in order to make a choice between the products. However, there is no commonly agreed methodology for this particular purpose. In most of the methods the valuation is implicitly or explicitly based on economic criteria. For various reasons, however, economically obtained criteria do not necessarily reflect ecological arguments correctly. Thus, there is a need for new, ecologically based valuation methods. One such approach is the expert judgement method, based on the Delphi technique, which rejects the economic basis in favour of the judgements of a group of environmental experts. However, it is not self evident that the expert judgement based environmental rating of interventions will be essentially more correct and certain than other methods. In this study the method was evaluated at different points of the procedure in order to obtain a picture of the quality of the indexes produced. The evaluation was based on an actual Delphi study made in 1995–1996 in Finland, Sweden and Norway.

The main questions addressed were the significance of the results and the operational quality of the Delphi procedure. The results obtained by applying the expert method indexes were also compared with the results obtained with other valuation methods for the background life cycle inventory of the case study. Additional material included feedback data from panellists of the case study, collected with a questionnaire. The questionnaire data was analysed to identify major dimensions in the criteria for evaluating interventions and correlation of the final indexes of the Delphi I study with these dimensions. The rest of the questionnaire material was used to document panellists' opinions and experiences of the Delphi process, familiarity with the environmental impacts of various interventions, and classification in typologies of cultural theory. The quality of results and methodological aspects, such as effects of task instructions, selection of the index basis, and effects of the final standardisation were analysed statistically. Accordingly, the effects of various postulates made on the conformity of the environmental harm conceptions of the experts, and the influence of the moderators' decisions were assessed on the basis of standard statistical indicators.

The state of consensus and its development in the Delphi process were studied with the aid of *K*-entropy analysis.

The study showed that transparency and certainty, which are essential qualities for an acceptable and trusted valuation method, are only partially accomplished by the expert judgement method in the format in which it was developed in the analysed case. As for the technical procedure, the method is well documented and transparency is good. Argumentation of the judgements, however, should be increased. The quality of the valuation indexes is explicitly available, but their certainty is very low for most interventions. The opinions of the experts vary greatly. How much this depends on different values and how much on differences in knowledge etc. is impossible to assess. Also, how much the technique used and the statistical processing of the experts' answers may have influenced the eventual scores of different interventions is difficult to assess.

The application of expert judgement to LCA valuation is a new idea, and the method is still very much under development and far from maturity. Nevertheless, utilisation of expert knowledge can be a significant addition to model approaches to ecological impact assessment, which, because of the chaotic behaviour of ecosystems, are limited and uncertain in predicting the ecological consequences of interventions to the environment. This should be taken into account when considering the results of the evaluation of the case study, which was the third of its kind in Europe.

Preface

The aim of the project was to evaluate a Delphi technique based expert judgement method in producing harmfulness indexes for different environmental interventions. The evaluation was based on an actual Delphi study made in 1995–1996 in Finland, Sweden and Norway (Wilson and Jones, 1996). The main questions addressed were the significance of the indexes and the operational quality the Delphi-procedure. Valuation results obtained using the expert judgement indexes were also compared with the results obtained with other valuation methods, and possibilities for wider use of the expert judgement with Delphi-technique were considered.

The study was carried out as a part of the research programme SIHTI. The research parties were VTT Chemical Technology, Neste Corporation and Landbank Environmental Research and Consulting. The Industrial Environmental Economics Research Group of VTT Chemical Technology was responsible for the project with Landbank and Neste as sub-contractors.

Thanks are due to Osmo Kuusi (VATT Economical Research Centre of Finland), Eva Heiskanen (Helsinki School of Economics and Business Administration), Juha Koponen (Ministry of Environment) and Pontus Mattson and Jouko Nikkonen (Neste Corporation) who actively participated in the project's follow up group and gave valuable comments to the issue. Thanks are also due to all those representatives of companies, the public sector and citizen movements who participated in the workshop and brought up their views on the LCA valuation.

Espoo, January 1999

Authors

Contents

APPENDICES

1. Introduction

Because of the complexity and trade-offs between different points in the life cycles of the analysed systems, it is practically impossible that a product could have consistently lower levels than alternative products for all interventions. When compared to an alternative, a product usually has higher levels for some interventions and lower levels for others. Therefore, a method which measures the environmental damage caused by each intervention is needed in order to make a choice between the products. However, there is no commonly agreed methodology for this particular purpose. One reason for the prevailing situation is that the proposed assessment methods include value judgements. There is no scientifically based method for reducing LCA results to a single overall score representing the environmental harm caused by a product. Another issue is that most of the methods presently proposed produce relative results which are very loosely, if at all, connected to the real ecology. Therefore they are less feasible for purposes such as assessing the sustainability of products, which requires actual ecological arguments.

Values and subjectivity are typical of any ranking, weighting and aggregation across categories of life cycle impacts because no natural scientific methods are available. In many cases it is also obvious that relevant natural scientific methods simply cannot be constructed because of the complexity of the interactions and internal structure of the technical and ecological systems. In most of the methods the valuation is implicitly or explicitly based on economic criteria. For various reasons, however, economically obtained criteria do not necessarily reflect ecological arguments correctly. The very large number of interventions which product systems cause makes it impossible to obtain relevant economic prices for them all. Eventually, the fact that ecological services are considered free in today's economic thinking prevents meaningful price setting for any damages which reduce the capability of nature to provide those services.

Because of the increasing concern about the sustainability of ecosystem services and the disability of economic methods to provide reliable information for decisions, there is a need for new, ecologically based valuation methods. One such approach is valuation based on expert judgement, which rejects the economic basis in favour of the judgements of a group of environmental experts. It is founded upon four main assumptions (Wilson and Jones, 1996):

1. A group of experienced scientists with a good understanding of environmental problems are the most knowledgeable and capable members of society for judging the relative significance of a number of interventions.
2. The judgements of the experts involved in the process are based on purely environmental rather than economic grounds.

3. The expertise of the experts enables them to judge which interventions are more damaging to the environment and to human health, and which are less damaging.
4. On the basis of their expert knowledge, the experts can rate each intervention in a way which reflects its true environmental significance in comparison with other interventions.

These are considerable assumptions, which, in case they fail, substantially reduce the relevance of the obtained ratings. Another important factor, which obviously will affect the results, is the method for consulting a panel of scientific experts. In order to produce valid results, it is necessary that the consulting method enables the experts to understand the task correctly and to express their opinions honestly and free from all external pressures. Impacts of environmental interventions are so complex that judgements about their relative significance, in terms of the damage they do to the ecosystems, including human species, may be controversial. It is necessary, therefore, that the method of consulting the scientific experts does not expose them to public controversy or censure when making the judgements in valuation. It is also quite important that the experts understand what is the method and the procedure to assess the relative importance of the interventions and how they should record their judgements. And, finally, adequate background information on the scene of the interventions must be available to the experts.

One possible communications procedure for eliciting information about the experts' judgements is the Delphi technique. Other possible communications procedures would be face-to-face interviews, round table meetings or one-off questionnaires.

It is not self evident that the environmental indexes produced with the expert judgement and Delphi technique for different interventions will be essentially more correct and certain than any other indexes produced with other kinds of methods. To get a picture of the quality of those indexes it is important, therefore, to evaluate the procedure and its results at different points in the procedure. This is done in the report at hand. The evaluation is based on an actual Delphi study made in 1995–1996 in Finland, Sweden and Norway (Wilson and Jones, 1996). This study will be referred to as the Delphi I study, or as the case study. The main questions addressed are the significance of the results and the operational quality the Delphi procedure. The results obtained by applying the Delphi indexes are also compared with the results obtained with other valuation methods for the background life cycle inventory of the Delphi case study. The possibilities for wider use of expert judgement with the Delphi technique were studied by surveying the opinions of panellists of the case study about generalisation of the method and the results of the case study. The possibilities of updating and expanding the data acquired from the expert panels by utilising Mackay models, developed for environmental impact estimation panels, were examined.

The study was carried out as a part of the SIHTI research programme. The research parties were VTT Chemical Technology, Neste Corporation and Landbank Environmental Research and Consulting. The Industrial Environmental Economics Research Group of VTT Chemical Technology was responsible for the project, with Landbank and Neste acting as sub-contractors.

2. Objectives and tasks

The following three main goals have been set for the assessment of the methodological quality of the Delphi procedure and the significance of its results in LCA valuation:

(i) to evaluate the actual expert-valuation procedure from the methodological point of view along with the assumptions made, instructions given, and actions taken at each stage.
(ii) to examine and to compare the valuation results indexes produced from a DELPHI exercise with results from other valuation methods
(iii) to evaluate the possibilities of generalising the method and preconditions for different uses of life cycle assessment.

According to the main goals the study is divided into three main tasks:

(i) Analysis of the method and data produced by the DELPHI procedure

In the first phase of the study, expert judgement based on the Delphi technique and its application to the evaluation of environmental impacts was analysed with the aid of a study made by Landbank Environmental Research & Consulting for Neste Corporation. The analysis included a study of the technique applied and possible alternative procedures for its different phases.

The course of the valuation process, the aggregation of the votes of individual judges into index statistics, and the computing of the final valuation weights were also studied. Special attention was paid to the logic of the rating of interventions, to the selection of the index basis, to functioning of the task instructions, and to the influence of the conductors of the Delphi exercise. The analysis was to a large extent statistical. The quality of the results was assessed by statistical methods under various postulates made on the conformity of the environmental harm conceptions of the experts. The state of consensus and its development in the Delphi process were studied with the aid of entropy analysis. Based on the analyses, an estimate was made for the distribution and confidence intervals of the Delphi indexes and LCA valuation results produced based on them, utilising the background LCI of the Delphi case study. Statistical analyses were complemented by a feedback questionnaire and interviews with the experts in order to clarify the factors which had affected the formation and expression of their opinions in the Delphi study, to assess the feasibility of expert judgement based on the Delphi technique for LCA valuation, and to identify needs and possibilities of developing the method for future applications.

(ii) The Delphi method in relation to other valuation methods

Valuation results computed using the indexes produced in the Delphi case study were compared to results obtained by other available valuation methods to clarify the position of expert judgement among various LCA valuation approaches in practice. Among the methods compared were the Swedish EPS method and methods produced in some recent Finnish studies.

(iii) The possibilities of generalising the method

The possibilities for (and possible restrictions on) wider use of expert judgement based on the Delphi technique were studied by surveying the opinions of panellists of the case study about generalisation of the method and the results of the case study, about the acceptance of the end result, and about broadening the scope of interventions. The possibilities of updating and expanding the data acquired from the expert panels by utilising Mackay models, developed for environmental impact estimation panels, were examined.

The acceptance and possibilities to generalise the results based on expert judgement based on the Delphi technique were analysed within groups dealing with environmental issues in decision making or other activities of interest. The views of possible users of the results in companies, in the public sector and in citizen movements (Ministry of Environment, Finnish Association of Nature Conservation, etc.) were collected in the form of a workshop utilising the method description produced in the study.

3. Methods and material

The analysis largely utilised material from a real Delphi study made by Landbank Environmental Research & Consulting for Neste Corporation.

Additional material included feedback data from panellists of the case study, collected by means of a questionnaire (eight of the sixteen panellists responded), and interviews with the questionnaire respondents. The questionnaire data was analysed to identify major dimensions in the criteria for evaluating interventions and correlation of the final indexes of the Delphi I study with these dimensions. The method used to detect the dimensions was reciprocal averaging. The rest of the questionnaire material was used to document the panellists' opinions and experiences of the Delphi process, familiarity with the environmental impacts of various interventions, and classification in typologies of cultural theory.

The quality of the results and methodological aspects such as effects of the task instructions, selection of the index basis, and the effects and meaning of standardisation were analysed statistically. Accordingly, the effects of assumptions and influence of the conductors' decisions were assessed on the basis of standard statistical indicators under various postulates made on the conformity of the environmental harm conceptions of the experts. The state of consensus and its development in the Delphi process were studied with the aid of *K*-entropy analysis. *K*-entropy, which is a special form of information entropy, is introduced as a measure of disagreement of opinions. Since it is a reasonably new term in LCA, there is also a brief theoretical introduction to it.

4. Life cycle impact assessment (LCIA)

4.1 Introduction

The objective of the impact assessment is to assist in the interpretation of inventory results. This is accomplished in two steps: (i) by aggregating the inventory results into indicators of environmental problems, such as global warming, stratospheric ozone depletion, acid deposition, photochemical smog, eutrophication, resource use, human health, etc. and (ii) by producing an overall estimate of the environmental harm caused by a product. This estimate may be based directly on the inventoried interventions or on the impact indicators. There are two principal approaches to LCIA: the assessment of potential impacts with the aid of equivalency factors, and the prediction of actual effects. The equivalency factor approach is, at least so far, applied in most European LCA studies.

Impact assessment methods can be divided into two broad groups according to how they apply the inventory data (Guinée, 1994). The first and, one might say, the traditional group, the so-called single-step methods, does not classify and characterise the impacts of the interventions but uses direct rating. The objective of these methods is to provide a single overall score representing the environmental harm caused by a product. Examples of single-step methods are the Ecoscarcity method (Ahbe et al., 1990) and the EPS system (Steen and Ryding, 1992). Also expert judgement based on the Delphi technique, as implemented in Delphi I, belongs to this group. The second group, the so-called multi-step methods, tends to separate the environmental science based assessment parts from the ethically and ideologically based ones. The emphasis is placed on the classification and characterisation of the interventions, and the possible valuation is built on the estimated impacts (potentials). However, attempts were not made to develop valuation methods for the multi-step approach until rather recently (Kortman et al., 1994 and Kalisvaart and Remmerswaal, 1994), and the feasibility of such valuation methods is still uncertain.

The idea of the procedure in Delphi I does not necessarily fit into the present standard structure of LCA. This is so because the expert judgements were elicited in a single-step procedure, applied directly to inventory. This means that the systematic proposed in the prevailing standard LCIA approach is not exactly followed. Nevertheless, an expert process has an inherent systematic, on the basis of which the interventions are ranked and rated. Even though this systematic is not made explicit in the process, as is the case in the standard LCIA approach, it can, without doubt, be assumed to fully correspond to the explicit systematic of the standard LCIA, which is discussed in the following.

4.2 Multi-step LCIA – the standard approach

According to ISO standard, ISO 14040 (ISO, 1997), which defines the principles of a framework for life-cycle assessment, the purpose of life cycle impact assessment (LCIA) is to evaluate the significance of potential environmental impacts of the studied system, using the results of life cycle inventory analysis. In the impact assessment phase, the inventory data is associated with specific environmental impacts. The level of detail, the choice of impacts evaluated, and the methodologies used to assess the impacts all depend on the goal and scope of the study. The position of impact assessment in the LCA process is shown in Figure 1.

The multi-step methods try to follow the standard structure of the life cycle impact assessment, which according to present thinking includes three basic elements, two optional elements and possibly a complementary element to enhance the interpretation (see Figure 2). This structure departs considerably from the earlier concepts, which usually had one or two basic components and treated valuation separately from impact assessment.

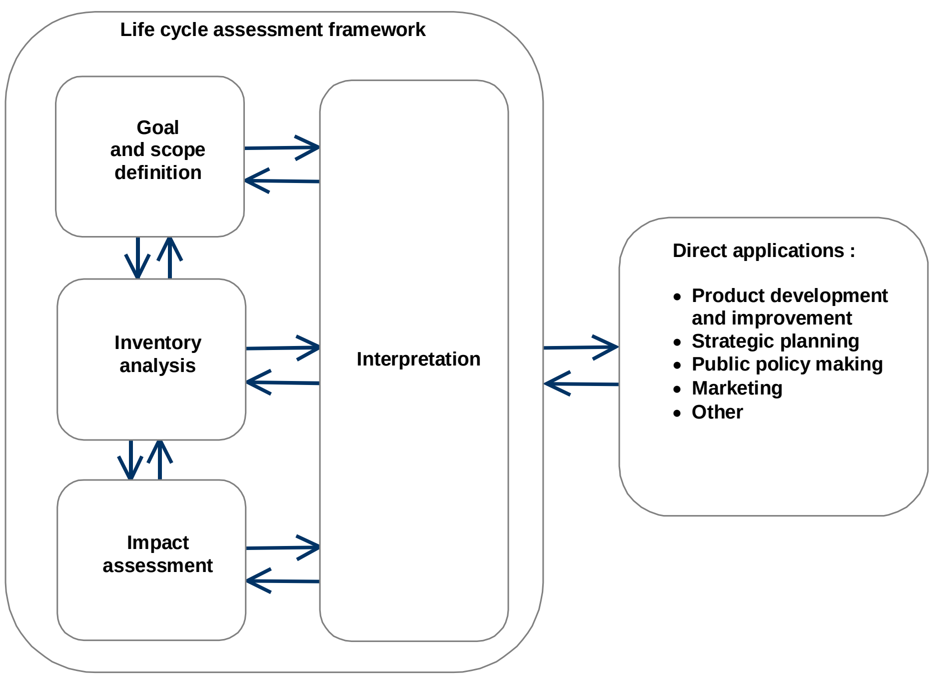

Figure 1. Standard life cycle assessment framework (ISO, 1997).

The basic elements which shall be included in all life cycle impact assessments are:

I. *Selection and definition of impact categories* constitutes the effects and the corresponding indicators that will be addressed. This is a new element in the impact

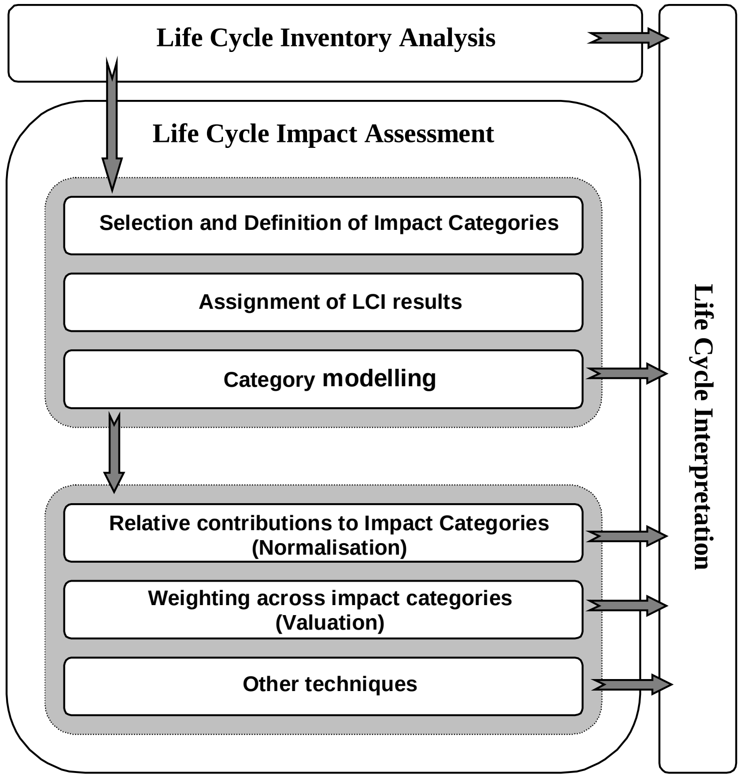

Figure 2. Typical information flows in the life cycle impact assessment (Modified from a committee draft of ISO 14042, 1997).

assessment. The requirements for this task include transparent documentation, scientific soundness and international acceptance. No categories are normative and in most studies impact categories are selected from previously proposed categories. These usually represent three main impact themes, i.e. resource depletion, human health impacts and ecological impacts. A summary of the proposed categories is given in Box 1. (Lindfors et al., 1995).

II. *Assignment of life cycle inventory results* (classification) to the selected impact categories is a qualitative step. The assignment is based on known environmental processes. In order to avoid multiple counting it may be necessary in some cases to divide an intervention into different categories. For instance, SO_2 can contribute to acidification and to toxic effects. However, a single SO_2 molecule cannot contribute to both effects concurrently. Therefore it is necessary to consider the shares of SO_2 for each category.

III *Category modelling* (characterisation), encompasses modelling of the inventory data within impact categories. This is a quantitative procedure in which the contributions of each input and output to its assigned impact categories are calculated in quantities of the indicators representing the categories, and the contributions are summed up within each category. An important aspect is that category models and, accordingly, the values they give to the characterisation factors (specific impacts) of interventions should depend, for example, on regional differences in background levels, the persistence of substances in the environment and dose-response characteristics, which they presently do not. Therefore, present characterisation models may be valid for global impacts, but they are not as feasible for regional and for local impacts because the ecological responses to the effect potentials are different in different environmental conditions. In this respect, the Delphi method, as applied in the Delphi I study, has an advantage because it aims to take the local aspects into account.

The optional elements of life cycle impact assessment are:

I. *Relative contribution to impact categories* (normalisation) is obtained by dividing the category results by the total indicator values of the categories, which are calculated using the current total emissions contributing to the categories. For instance, the normalised global warming potential is calculated as

$$GWP* = \sum_i m_i q_{CO_2,i} \Big/ \sum_i m_{tot,i} q_{CO_2,i} \quad . \qquad (1)$$

with $q_{CO_2,i}$ as the global warming potential (CO_2 -equivalent) and $m_{tot,i}$ as the total rate of an emission for the time period and area being considered. Normalised results may be helpful, for instance, in detecting data errors, and in ranking the impacts of the product system from a relative importance point of view. On the other hand, normalisation means a further manipulation of data and the introduction of new uncertainties. For this reason it should be avoided. But when normalisation supports valuation, which is the case for some proposed valuation methods, it is necessary. In any case, normalised category results are not a measure of environmental impacts.

IV. *Weighting across impact categories* (valuation), which means aggregating the results in very specific cases and only when meaningful. Valuation can be either a qualitative or a quantitative procedure in which the relative importance of the different potential environmental impacts are weighted against each other. It is done in order to make judgements on the relative environmental benefits and disadvantages of different systems compared to each other. Valuation employs other than natural scientific criteria, e.g. economical, political and other subjective arguments, and because of the fact that people have different values, it has been and will be a disputed issue in life cycle assessment. An overview of valuation approaches is given in chapter 5. Valuation in LCA.

Box 1. Summary of impact categories proposed for the life cycle impact assessment.

Impact categories proposed for the impact assessment can be grouped under three main environmental impact themes:

(i) Resource depletion including
- *a) Energy and materials*
- *b) Water*
- *c) Land*

(ii) Human health impacts including
- *a) Toxicological impacts*
- *b) Physical impacts*
- *c) Psychological impacts*
- *d) Diseases caused by biological organisms*

(iii) Ecological impacts including
- *a) Global warming*
- *b) Depletion of stratospheric ozone*
- *c) Acidification*
- *d) Eutrophication of waters*
- *e) Eutrophication of lands*
- *f) Photo-oxidant formation*
- *g) Ecotoxicological impacts*
- *h) Habitat alterations and impacts on biological diversity*

In addition to the actual impact categories given above, it has been suggested that two indicative categories should be added in the impact assessment, i.e. inflows and outflows which are not followed to the system environment.

Life cycle impact assessment may also include other techniques and information to better understand the relevance or the accuracy of life cycle impact assessment results, to remove negligible results or to guide further study. Analytical techniques applied may include dominance analysis to identify the items with the greatest influence on the life cycle impact results; uncertainty analysis to assess the level of accuracy of the life cycle impact results; and sensitivity analysis to assess how variations in data and methods will change life cycle impact results. Environmental data may be utilised to study where threshold values are exceeded and thus to assess the relevance of the results.

Despite a reasonably well developed formal structure, the methodology and science of life cycle impact assessment are still rather underdeveloped. Models for various impact categories are in different stages of development. There are no generally accepted methodologies for consistently and accurately associating inventory data with specific potential environmental impacts. Relevant valuation methods, which would rely on the results of characterisation, are, anyhow, impossible to develop before valid characterisation methods for all impact categories are available. To better relate the impact categories to the valuation step, intervention-damage relationships should be developed further by modelling the cause-and-effect chains from the interventions up to the end-points, the final damages, which would then be valued (Udo de Haes, 1996). It would be best to base the valuation on final damages, caused by the environmental interventions, such as health effects or loss of biological production. But assessment of the damages due to complicated impact chains is very laborious. In many cases it is also obvious that relevant natural scientific methods simply cannot be constructed because of the complexity of the interactions and internal structure of the technical and ecological systems.

4.3 Single-step LCIA – the usual approach

The problem with the multi-step impact assessment is that it presently restricts the valuation, which is perhaps the most important element in the LCA process. The vast majority of valuation methods are single-step methods, which employ direct rating of interventions to provide an overall score for the environmental harm. Accordingly, valuation and impact assessment are treated as separate processes, as indicated schematically in Figure 3. This approach, which has been and still is widely applied in the LCA society, allows valuation to be performed both directly upon the inventory and following impact assessment, including the optional normalisation. For instance, the Nordic guidelines on life cycle assessment (Lindfors et al., 1995) recommend this kind of a structure of LCA, in which valuation is performed directly from the inventory or from impact assessment.

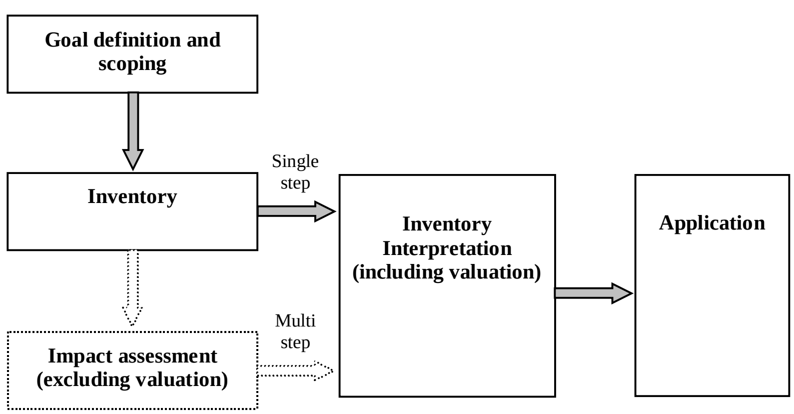

Figure 3. Present structure of Life Cycle Assessment (modified from Lindfors et al., 1995).

Expert judgement based on the Delphi technique employs basically the single-step concept. All single-step methods have some kind of inherent systematic on which the different interventions are rated. In most cases, however, this systematic is not based on actual causalities or ecological consequences, but on proxy measures, which are normally derived from economic relations. Expert judgement in Delphi I differs in this sense from most of the other single-step methods, since it tries to be clearly ecological in the argumentation of judgements. Valuation in life cycle assessment is discussed further in the following.

5. Valuation in LCA

5.1 Introduction

Decision-making situations where environmental information is used vary, so different motives and criteria will be included. The purpose could be just to comply with legislation and with agreements, or with competitive requirements. Pressure from customers and other stakeholders may sometimes be decisive. Decision-making may be influenced by various societal factors, such as views on the society, market economy, and democracy etc., and considerations such as equality between the present and future generations, views on nature and ecological systems, equality of species, and so on. Further arguments may relate to environmental risks.

Valuation includes political, ideological and ethical values. It raises several fundamental questions, such as:

- Why should the weighting be performed ?
- Should absolute priority be given to some aspects, such as irreversibility of non-renewable resources or effects on working conditions, or effects on human health, or violation of human, animal or other natural rights.
- If weighting is performed, which methodological approach should be chosen?
- Which weighting factors should be used?

and further:

- Whose values should be respected, citizens', experts' or politicians' ?
- How should information on preferences be obtained?
- Should information be based on primary or secondary sources, like empirical studies?

The concerns and preferences of stakeholders relate to their views on nature and ecological systems and their attitudes to decision-making, risk and justice. In the decision-making situations, political attitudes or views may be assumed to have an indirect but important role. According to the kind of view on nature, cultural theory distinguishes between four broad stakeholder types (e.g. Douglas and Wildavsky, 1982, and Schwarz and Thompson, 1990), namely individualist, hierarchist, fatalist and egalitarian (Box 2). Each type represents a typical political culture with characteristic values and rationales concerning nature and its management. These four main attitude types are briefly described in the following. Figure 4. illustrates the different views on nature.

According to the individualist's view, nature is benign, predictable, bountiful, robust, and stable in global equilibrium. Thus it is forgiving of any insults that humankind might inflict upon it. This corresponds to a Laissez-faire attitude towards the management of the environment. Individualists emphasise economic values and believe, accordingly, that markets will solve the environmental problems. Individualistic behaviour is typical of individualised corporations, liberal politicians, normal consumers, and business-oriented eco-establishments.

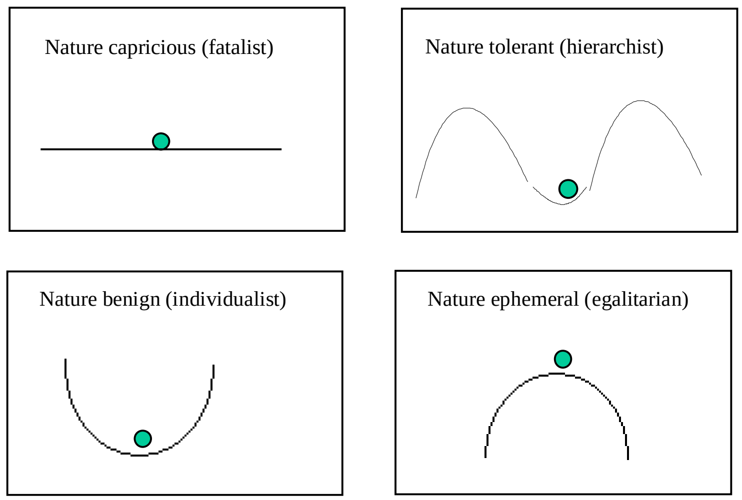

Figure 4. Illustration of the views of the four main attitude types on nature.

A hierarchist sees nature as perverse and tolerant within limits, when not pushed too far, and thinks that care must be taken not to exceed the safety limits. Environmental risks are understood as technical and physical phenomena. Examples of this attitude type are hierarchist corporations, conservative and old socialist politicians, regulators, and political and expert establishments. According to the hierarchist's view, problems can be solved with technology and know how, and authoritative regulation is preferred.

Nature may also be regarded as capricious, unpredictable and beyond the control of an individual. Changes may turn out well or badly, but it is not possible to predict the direction beforehand. Effects in the environment are recognised only when concrete and sensible consequences are perceived. The corresponding attitude type is called fatalist, and representatives are groups with marginal influence, such as poor consumers, and businesses subject to extremely severe competitive conditions. The management principle of a fatalist is no management or learning, just coping as best one can with erratic and unpredictable events.

From the egalitarian viewpoint, nature is as ephemeral, precarious, fragile and unforgiving, and in danger of being driven into a catastrophic collapse by human carelessness. Thus the precautionary principle should be followed. According to the egalitarian view, societies must practice control. Values other than economic ones are important. Examples of the practitioners are egalitarian corporations, environmentalists, green politicians, and green consumers. A campaigner is a typical egalitarian.

Different views imply different preferences in assessing environmental impacts (Finnveden and Lindfors, 1997). For example, if nature is seen as benign, then more freedom is allowed, or if surprising, then it may be found preferable to act according to the precautionary principle. Thus, if the panel involved in valuation could be grouped according to attitude categories, the results of the valuation could be better interpreted and even forecasted.

The issues to be assessed may be so difficult that it is impossible to reach convergent results on the factual bases. Attitudes and values also have a role. A hypothesis may be put forward that, because of the individual views of each expert, each judge must have individual, underlying patterns, “damage curves” for the interventions' environmental impacts (for an explanation of damage curves, see chapter 5.2), on the basis of which the valuation is made (Wilson, 1997–1998). The hypothesis also presupposes that whatever pattern an expert is using, s/he will use it consistently. Because of the different damage curves, it is not possible to reach a consensus among the experts. There will always be disagreement about the absolute values of environmental harm. Anyhow, there might be an agreement in relative indifference curves. A current study is testing this hypothesis on 200 experts.

Box 2. Cultural theory typologies with respect to the environment (Schwarz and Thompson, 1990).

Type	**View about Nature**	**Management approach**
Individualist	Benign, predictable, bountiful, robust, stable and forgiving of insults from humans	Non-interventionist management approach, market forces will restore equilibrium
Hierarchist	Perverse and tolerant within limits, provided it is not pushed too far	Interventionist management approach, based on research to establish the limits and regulation to ensure limits are not exceeded.
Fatalist	Capricious, unpredictable and beyond human control	Resigned to fate and sees no point in trying to change it.
Egalitarian	Ephemeral, precarious, fragile and unforgiving. Believes that people must tread lightly on the Earth	Guiding management rule is the precautionary principle, and follows the ’small-is-beautiful’ ethic.

Valuation is the most subjective phase in the LCA, although subjectivity cannot be avoided in the preceding phases either, because each step comprises many alternatives, between which choices have to be made. Values change with the passage of time and vary with social conditions. The appropriate method should be able to take account of different values and preferences and to indicate how the final weighting will be affected by various emphases of different values. The inescapable subjectivity should be managed correctly. Whatever the ideological standpoints are, they should always be handled transparently and explicitly.

According to the criteria applied, the main departure points for valuation in LCA are targets and preferences or judgements [1]. Monetarisation, which is often considered as its own type of valuation, eventually reflects the preferences of people or authoritative decisions, as, in fact, do the targets as well.

Those approaches, which are formally based on physical measures such as total energy consumption or material turnover, are relative and there is no indisputable scientific evidence on how these measures correlate with the carrying capacity of the environment.

Environmental policy targets suggest what kind of environmental conditions are to be sought and which actions are needed to bring that about. Targets are normally based on politically decided sustainability levels. Environmental targets are becoming more important both in LCA and in other contexts. When targets are used in valuation, legality and the interests of all involved parties should be taken into account. This is not always self-evidently the case, for instance, because of over-emphasised national interests in setting targets, or because of the rights of socially or geographically remote communities are sometimes ignored. In the Delphi case study, experts are asked to make their judgements solely on ecological grounds. However, the geographical region to be considered was limited to Northern Europe. Thus, the judgements were, at least in theory, free of most political and social restriction, which target-based valuation methods may entail.

When approaching the valuation problem from the standpoint of targets, mutual weights of impacts are normally based on the ratio of the total score of intervention or impact either to the target per se or to distance to the target, which is the difference between the current level and the target. Weighting system may also comprise criteria based on available technological options for reducing the environmental impacts. Sustainability should, of course, be the eventual and overlapping target for all environmental and development policy. In many cases, however, it is not at all certain that this is the case. For instance, many ecologists regard the present targets for greenhouse gas emissions more as poor political compromises rather than really effective limits to prevent the progress of global warming. Therefore, target-based methods may well mask some

[1] The preferences are personal, self-centered desires which may or may not have a rational basis, but cannot be a matter of factual dispute. They make no claim about the outside world. Judgements, on the other hand, while still remaining personal, are oriented towards both self and others and the outside world. As such they are capable of being scrutinised. (Holland, 1997)

important environmental problems from the decision-makers. An ideal feature of the expert judgement method is that it relies on a group of experienced scientists with a good understanding of environmental problems and who are the most knowledgeable and capable members of society to judge the relative significance of interventions. This should, in principle, enable the method to reveal the real ecological harmfulness of each intervention. On the other hand, panel judgement is a social process, which makes it a subjective, even with experts. Depending on the quality of the panellists, judgements will somehow be affected and inspired by facts and scientific information. But the panellists retain the freedom to deviate from scientific standards and knowledge, and instead, to take greater account of the uncertainties and logic of risk management, for example by following a precautionary principle.

Weighting based on judgements may include different techniques and different levels of sophistication, such as Delphi or methods and multi-attribute utility theory. Information on preferences may also be derived from the behaviour of individuals or organisations, or from the expressed preferences of individuals. The values that individuals place on non-market goods and services are estimated using several techniques. Impacts may be monetarised, for example, by using market prices of resources, and shadow prices of maintenance of environmental functions, appearing, for instance, as costs of nature conservation actions, property value change, travel costs, willingness-to-pay etc.. In the following chapters we will give an overview of the state of the art of the valuation methods. The Delphi technique is presented separately in chapter 6. Delphi technique.

5.2 Valuation methods

Valuation methods can be broadly classified as qualitative (including semi-quantitative) and quantitative methods. The latter are presently predominant in number and are far more frequently used. Nevertheless, qualitative valuation methods are an interesting group of techniques even though they are not featured prominently in the methodological discourse of the life cycle impact assessment. In fact, the interpretation of LCA results is usually a more qualitative than quantitative process, because of the subjectivity of the quantitative methods.

5.2.1 Qualitative methods

The idea behind the qualitative valuation methods is to structure the information provided by an LCA into such a format that somebody can draw conclusions from it. Organising data in suitable matrices is a frequently used technique. One type of matrix

(Christiansen et al., 1990 and Graedel et al., 1995) maps the values of environmental parameters at the different life-cycle stages. Environmental parameters may be either inventory parameters, impact categories or some other kind of classification parameters. In an absolute analysis, the severity of the impacts represented by those parameters are indicated by a sign or a number. However, it has turned out to be difficult to find criteria for the severity. When the technique is applied to comparisons, signs or numbers are used in each cell to indicate whether the alternative is better, worse than or equal to a reference. The idea of this kind of mapping is illustrated in Figure 5.

	Fossil energy resources	**Acidification**	**Global warming**
Raw material acquisition	+	–	+
Materials manufacture	–	–	–
Product manufacture	–	+	+
Product use	+	–	+
Reuse and recycling	–	–	–
Final disposal	+	+	+

Figure 5. The idea of life cycle phase – the environmental parameter mapping approach to valuation.

Assessing the importance of each cell is difficult though. The structure suggests that the overall comparison can easily be made simply by just summing up the plusses and minuses. This is not, however, the idea of the method. Eventually, this kind of matrix approach may be less feasible for making an overall evaluation, although they may be quite useful for clarifying the decision situation.

In another technique (Schmitz et al., 1994) the importance of the considered environmental impacts are evaluated using five criteria:

- Ecological threat potential
- Reversibility – irreversibility
- Global, regional, local
- Environmental preferences of the population
- Relationship of actual and previous pollution to quality goals.

The ecological importance of each impact is rated on the basis of the above criteria in a five-level scale ranging from little importance to very great importance. For the overall statement the results of the category modelling are normalised and rated using also a five-level scale from little to very great. Based on the rated effect potentials and the level of the ecological importance of the different impacts, a verbal overall statement is made and conclusions are drawn.

The qualitative valuation techniques are obviously strongly dependent on the persons who make the valuation. Therefore the results obtained by these techniques might not be reproducible since other persons performing the valuation could reach another conclusion. Therefore they cannot be successfully used e.g. for the communication of LCA results, as shown by some cases where an attempt has been made to use them for communicative purposes.

5.2.2 Quantitative methods

In most of the quantitative methods the valuation is based on economic criteria, either explicitly or implicitly. Some methods do not consider actual environmental effects, but use proxy indicators such as energy or materials intensity. There are three main types of quantitative valuation methods according to the basis they employ (Finnveden, 1996):

1. Monetarisation methods
2. Distance-to-target methods
3. Panel methods.

Combinations of these basic types of methods are also possible. Still, it is typical for all quantitative methods that the valuation is based on an additive rating of different interventions or effect potentials, i.e. the overall score (H) is obtained from

$$H = \sum_i H_i''' V_i \ , \tag{2}$$

where H_i''' is the rating for an intervention (or an impact potential) with the inventoried volume V_i. In all methods the ratings are considered to be independent of the intervention volumes. How the various interventions are rated in the three basic approaches is discussed in the following.

Other techniques which have been proposed for the valuation of LCA include, for instance, applications of the multi-attribute utility theory (MAUT) and the analytical hierarchy process (AHP), which operate on the level of pairwise comparisons of environmental objects and their values, rather than on the level of the entire environment and its predicted damage. So far, however, these other techniques are not in general use in LCA valuation.

5.2.2.1 Monetarisation methods

Monetarisation methods either take advantage of some kind of willingness-to-pay data or employ techno-economic assessments to obtain the value for the interventions. The rationale of the willingness-to-pay is usually to avoid unwanted effects on something which is considered valuable. One might be concerned about either the effects per se or the possibility of having them. There are user and non-user values associated with the environment. User values may be direct, such as the timber value of a forest, or indirect, like the recreational value of the forest. The existence value of the forest ecosystems is an example of non-user values. The total value reflected in the willingness-to-pay is a composition of all user and non-user values. This composition may be different on different occasions.

There are a number of techniques for estimating the willingness-to-pay. It can be derived, for instance, from individuals' revealed preferences, individuals' expressed preferences and society' willingness-to-pay. Methods based on individuals' revealed preferences seek data from the market. They assume that people reveal their environmental preferences in market prices. But pollution is not a tradable commodity. Therefore these methods need to utilise substitute data such as damage assessments from insurance companies, etc. For the most part it is only direct user values that can be derived from market prices. However, in some cases indirect user values can also be worked out from market values. Examples of such cases are the so-called travel cost method, which is used to evaluate recreation sites, and hedonistic pricing methods, which are applied to house prices and take account of a number of factors, including environmental aspects (Finnveden, 1996).

People's preferences for reducing environmental damage can also be sought by asking them directly to give values for environmental assets (Turner et al., 1994). This technique is the foundation of the contingent valuation methods (CVM), which are often referred to as expressed preferences methods. In principle CVM could also be used to evaluate interventions and hazards, but it is mostly used to evaluate damages. The problems associated with the technique are well known. Willingness and its indication are prone to the person's ability to pay, perception of the significance of a cleaner environment, ethical viewpoints, education etc. Consequently the answers will strongly depend on who are asked, which makes international comparability difficult. Therefore a complementary concept, 'willingness-to-accept', which basically relies on the knowledge of the environmental processes and on the ethics of the judges, has been developed in order to reduce the dependency on the socio-economic conditions of the respondents. Even this concept includes features which vary with the economic and educational standard of living. Thus, it cannot remove all the difficulties associated with a wider application of the CV method.

Political and governmental decisions and expenditure records can be used to study society's investments in avoiding damages, for instance in saving a life or in keeping emissions within certain limits. Then, the marginal reduction costs found for a certain emission can be regarded as the specific price that society is willing to pay for preventing that emission from causing damage. Or, if an emission has negative impacts on health, the estimated loss of life and life-saving investments could be used to evaluate the damages caused by the emission. Society's valuations could also be assessed from environmental taxes. Taxes collected on emissions can be considered as prices that society regards as being justifiable compensation for the damages caused by the release of those pollutants.

The dose-response technique (Pearce, 1993) assesses the physical and ecological linkages between the pollution (dose) and the impact (response), and values the final impact at a shadow price. For example, if air pollution has negative impacts on buildings or crops, the resulting damage can be estimated by valuing the loss of crops or the damage to property. The dose-response technique has been used, for instance, to generate damage costs for conventional air emissions like carbon dioxide, sulphur dioxide and particulates.

Still another technique is to estimate how much it would cost to reduce pollution, e.g. by abatement technology. In this case, the ratings are based on an estimation of the cost of doing something, but it is not an issue of whether somebody is willing to pay this cost. Therefore, the rates are clearly not measures of willingness-to-pay. The rates based on the abatement cost do not necessarily represent the size of the damage from pollution either. However, given the lack of damage data for some types of pollution, this approach can provide rough proxy figures for valuation.

The EPS and Tellus methods are presently perhaps the best known monetary-based valuation methods. The EPS method employs the concept of 'willingness-to-pay' and uses historical expenditures paid by society in order to control environmental problems. The Tellus method uses the estimated abatement costs of reducing interventions to desired levels.

A perceived shortcoming of the monetarisation methods is the difficulty in establishing values for nature. It has been estimated (Costanza et al., 1997) that ecological systems and natural capital stocks contribute between US$ 16 and 54 trillion (10^{12}), with an average of US$ 33 trillion. To put this in perspective the combined GNP of all of the countries in the world is around US$ 18 trillion per year. These services are rarely, if ever, captured in any economics-based valuation study.

This type of approach represents the views of those who believe that nature has an instrumental value to humankind; there exists another group who reject this approach altogether, arguing that nature has an intrinsic value outside any human-based concept such as economy. (‘You can’t put a price on the environment’).

5.2.2.2 Target-based methods

Target-based methods compare the actual annual load of an intervention with a desired target load. Two such methods have been developed: the 'Eco-point' system developed in Switzerland and the 'Distance-to-target' method developed in the Netherlands. Both methods use the underlying concept of damage curves for interventions, where the harm caused by the intervention is plotted against levels of the intervention. Damage curves are generally considered to be S-shaped, as shown in Figure 6.

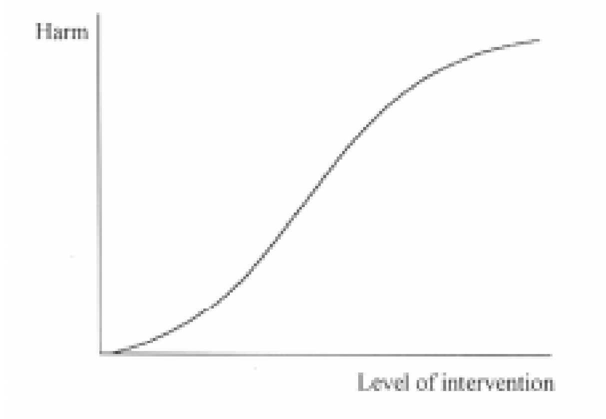

Figure 6. General shape of the damage curve.

The Eco-point method employs national targets which are devised in a political process moderated by central government. It therefore depends upon trading-off an acceptable level of harm with the economic constraints of making reductions. The distance-to-target method uses a target for a 'sustainable' level of each intervention. This again implies a trade-off between economic and environmental benefits, and in both cases there is a danger of double-counting when a manufacturer makes choices involving costs.

In its simplest form the Eco-point method sets up valuation factors calculated by dividing the actual annual load (A) of an intervention by the square of its target value (T). The environmental index (I) for a product is then calculated by multiplying each intervention loading (e) identified in the inventory by its valuation factor and summing over all interventions. This gives the following equation:

$$I = \sum eA / T^2 \quad (3)$$

For the distance-to-target approach the valuation factor is simply the reciprocal of the target, i.e.

$$I = \sum e/T \qquad (4)$$

The latter approach has been criticised because it does not depend on the actual level, A. This means that two interventions having the same target value will have the same valuation factor, even if the actual level of one is much greater than that of the other. Consider again the two hypothetical countries described previously. Since they are identical, both countries should have the same target. Country A, however, will be at or near this target, whereas country B will be greatly above it, and so should have a high valuation. The distance-to-target approach results in the same valuation being assigned to both, however. In the Eco-points system, the intervention with the higher actual value will be more severely punished for having a greater overshoot compared with the target. It can also be argued that this system is more appropriate for countries whose actual levels are in the lowest part of the 'S', where the curve is approximately quadratic.

A problem with both methods is that they assume there is no threshold value for the curve below which there is little or no environmental damage; distance-to-target assumes a straight line through the origin, Eco-points a quadratic. A more appropriate assumption may be to introduce a threshold value below which harm is zero or near zero – a No Significant Adverse Effects Level (NSAEL), say. Then, assuming the target is set at the threshold and a straight line is assumed above the threshold, the environmental index is given by

$$I = \sum e \frac{A-T}{T} \text{ when } A > T \qquad (5a)$$

and

$$I = 0 \text{ otherwise} \qquad (5b)$$

A perceived problem with this method – which might be called 'distance-to-threshold' – is that it gives a valuation of zero for levels below the threshold. This might induce a manufacturer to switch from an area where loads are greater than the threshold to a pristine area. This risk must be considered to be fairly minimal, however, given current levels of concern and regulation. In any case it might be considered better to move production from a highly polluted area to one where the carrying capacity is greater. The merits of this method are that it does not depend on economic assessment – the NSAEL level is determined purely on environmental grounds – and it takes the current load of the intervention into account.

A possible extension of this method may be to determine both the economic target level (T_e) and the threshold level (T_t). Then the index becomes:

$$I = \sum e^{\frac{A - T_t}{T_e - T_e}} \qquad (6)$$

Opinion-based methods are less well developed than either of the two general methods described above, and no widely used example is available as far as the writers are aware, although a number of individual studies have been performed on a case-by-case basis. The are two important considerations in the conduct of such studies: firstly, the method used for soliciting opinions and, secondly, whose opinions are sought (the stakeholders). General decisions to be made when selecting a data collection method are whether the subjects should be interviewed face-to-face or by post, as a panel or individually, and as named or anonymous participants. Possible stakeholders are the general public, politicians, manufacturers, campaigners and experts. The media can also be an important consideration in the process. In an ideal world the general public should be able to express its preferences and make the necessary choices between the environmental effects of different interventions. These preferences should take account of the rights of those in all countries (international), of all species (interspecial) and of all generations (intergenerational). We do not consider this a practical possibility at present, however, and will make the case in the next section for using panels of expert scientists and the Delphi technique (explained more fully in section 6) to collect their opinions.

5.2.2.3 Panel methods

In panel methods a group of people are asked about their opinions, which they should express in a quantitative format. However, the composition and the qualities of the panels may be very different with regard to their members' knowledge and perception of the importance of environmental problems. Also the questions may be formulated and asked in a number of ways. Consequently, there is much room for variation in each relevant factor of the panel method, and, thus, the results obtained with the aid of panels can vary greatly. The following two examples highlight the variability of the panel method. They are both modified from Finnveden (1996).

The first example, explained in the following, is a Dutch study, which used a four-step Delphi-like process and a panel composed of representatives of industry, government, environmental groups, universities and scientific institutes. The aim of the first step was to gain a common understanding of the importance of the impact categories and of the issues which were included in the environmental profiles. One basis for the discussion was a framework in which different aspects of the different categories were defined, such as whether the impact is only on humans or ecosystems or both, the degree of scientific uncertainty, the degree of reversibility of the impact, the scale of the impact, the timing of the impact and other issues. The second step of the process was a first assessment of the weighting factors. This step was confidentially done by each panel member. In the third step, the results of the second step were presented to the members, who continued the discussions. The fourth step was a second assessment. The process was then continued until a final, agreeable set of weighting factors was achieved.

In another panel study, weighting factors were obtained by interviewing environmental experts (Finnveden, 1996). In the interviews, each of the 22 experts was first provided with a description of the status of the environmental problems and then asked to rank them. In the next step they were asked to share 100 points between the ranked environmental problems in a way which would correspond to their understanding of the seriousness of each problem. After this was done, the experts were shown alternative scores produced with a distance-to-target based method and they were given a chance to change their earlier rankings and rates.

In a British study (Wilson and Jones, 1994), the valuation was performed directly by the Delphi technique. The panel consisted of eleven anonymous experts from British Universities. They gave their views on the subject being investigated by completing a questionnaire. The results of the survey were summarised and fed back to each expert by post, showing how his or her view differed from the other participants. The experts were then invited to reconsider their positions. From the judgements obtained in the second iteration, scores reflecting the median values were obtained and then applied to the inventory data.

5.2.2.4 Analytic hierarchy process

The analytic hierarchy process (AHP) is a comprehensive, logical and structural framework, which provides a better understanding of complex decisions by decomposing the problem into a hierarchical structure. The incorporation of all relevant decision criteria, and their pairwise comparison, allows the decision-maker to determine the trade-offs among competing objectives. Such multi-criteria decision problems are typical, for instance, for environmental policy planning. The application of the AHP approach explicitly recognises and incorporates the knowledge and expertise of the participants in the priority-setting process by making use of their subjective judgements – a particularly important feature for decisions to be made on a poor information base. However AHP also integrates objectively measured information, where such information is available.

The AHP approach is based on three principles:

1. Decomposition of the decision problem,
2. Comparative judgement of the elements, and
3. Synthesis of the priorities.

The first step is to structure the decision problem in a hierarchy as depicted in Figure 7. The overall focus of the decision problem, such as evaluating the overall environmental performance of the alternative product systems, is at the top level of the hierarchy. The next level consists of the criteria relevant for this goal and at the bottom level are the alternatives to be evaluated. Such criteria may represent, for instance, various impact categories.

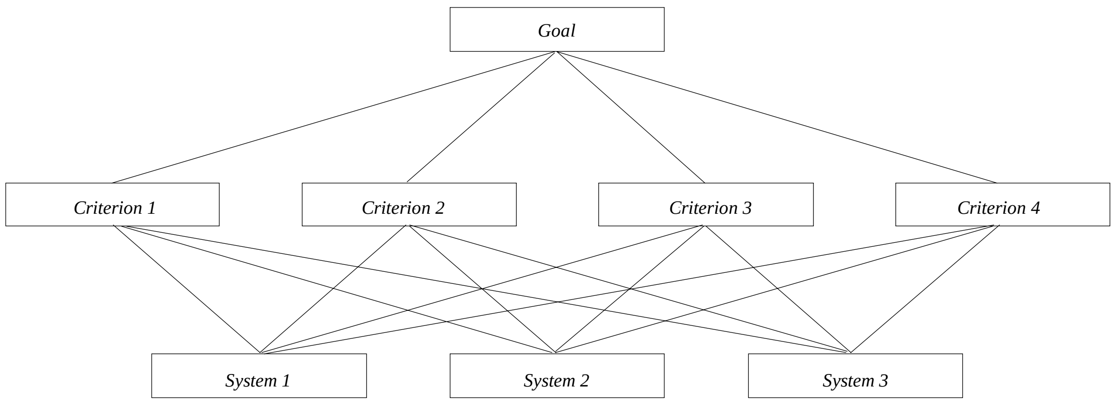

Figure 7. The basic structure of the hierarchy.

The second step is the comparison of the systems and the criteria. They are compared in pairs with respect to each criteria. The scale given in Table 1 can be used for this relative comparison. It allows the comparisons to be expressed in verbal terms, which are then translated into the corresponding numbers. In a similar way the criteria are compared pairwise with respect to the goal to obtain the weight of each criterion.

Table 1. Fundamental scale for pairwise comparisons.

Verbal scale	**Numerical values**
Equally important, likely or preferred	1
Slightly more important, likely or preferred	3
Clearly more important, likely or preferred	5
Very clearly more important, likely or preferred	7
Overwhelmingly more important, likely or preferred	9
Intermediate values to reflect compromise	2, 4, 6, 8

In the last step, the comparisons are synthesised to get the overall priority of each system. This is done by summing up the products of the priority of a system with respect to a criterion and the weight of the respective criterion. This is illustrated in Table 2.

Table 2. Priorities, weights, and the final ranking of the systems.

	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Global priority
Weight of the criterion with respect to the goal	0.483	0.066	0.136	0.315	
	Priorities of the systems with respect to criterion				
System 1	0.571	0.452	0.653	0.085	0.422
System 2	0.286	0.072	0.062	0.644	0.354
System 3	0.14	0.476	0.285	0.271	0.224

AHP is not generally used for LCA valuation at the moment, but the technique has potential because it can assist in structuring and formulating opinions and preferences and thus adds to the transparency of the valuation. AHP may have a major impact on the valuators' understanding of the factors which influence the value of a system. The pairwise comparison may become fairly time consuming if a large number of alternatives and criteria need to be evaluated. It may, however, be replaced by a quicker scaling approach.

5.2.3 Available valuation methods

Some of the valuation methods presently available and used are presented in Box 3.

Box 3. Some presently available and generally used valuation methods.

Eco-scarcity approach (Swiss)	In the Eco-scarcity method the different emissions are weighted against each other directly by using eco-factors. Eco-factors are calculated for each emission by dividing the actual flow of emission by the critical flow. The critical flow is evaluated from the annual load limits for a certain area, which are set, for example, by national environmental protection laws and regulations. The actual flow is the total yearly amount of the emissions in the area.
Effect category (Swedish)	In the characterisation step of the Effect Category method the environmental loads are grouped into effect categories, i.e. selected environmental themes. The result per theme is normalised by dividing by the corresponding total pollution of the same theme within the geographical area relevant to the study. The impact fractions of several themes may be summarised after applying weighting factors for environmental themes. This weighting may be based on expert panels, political goals etc.
EPS (Swedish)	In the EPS (Environmental Priority Strategies in product design) system five safeguard objects are valued. The objects are biodiversity, production, human health, resources and aesthetics. The valuation of the objects is based on society's cost for protecting (biodiversity), OECD market prices (production), society's cost for reducing excess deaths caused by various risks and people's willingness to pay to avoid diseases, suffering and irritation (human health), impact on the other safeguard subjects when restoring the resource (resources), the willingness to pay (aesthetic values) to restore them to their normal status. Emissions, use of resources and other human activities are then valued according to their estimated contribution to the changes in the safeguard objects.
Tellus (USA)	The valuation system is based on the control costs of a number of air pollutants, such as CO, NO_x, particles, SO_x and VOC, etc. The valuation for greenhouse gases is based on the costs of afforestation for a carbon sink. For the purpose of valuation the hazardous substances are ranked on the base of health risk factors (US Council for Environmental Quality). The units of environmental indices are in US$.
Aggregation of critical volumes (Swiss)	The emissions in separate environmental compartments, air, water and ground are weighted and aggregated on the base of (legal) emission limits for the compartments. MIK (maximum emission concentration) – values should guarantee well-being for humans when exposed to emissions according to limits, 24 hours per day, etc. Based on medical research as well as technical-political feasibility.
ECO Indicator (Dutch)	The method is based on scientifically set targets of various environmental impacts. The targets are interpreted to the levels or factors by which the emission has to be decreased to reach a certain acceptable level. (Goedkoop, 1995)
DAIA (Finnish)	The valuation problem is structured with the aid of decision analysis methods. Weighting of environmental impact categories is done by experts. Coefficients for emissions (attributes) are obtained by multiplying the weight of the attribute within the impact category (based on scientific knowledge or preferences of experts) by the category weight and by dividing by the intervention scores of the application area. Concerning the total interventions within the certain area, local conditions are also taken into account to some extent. (Seppälä, 1997).
Panel method	Expert judgements may be performed directly on interventions or impact categories. Interaction among panellists can be enabled by the Delphi technique or with other structured dialogue methods. Experts panels or other stakeholders may be asked about their preferences. The way in which the questions are asked and the use of appropriate units are important.

6. Delphi technique

6.1 Background

The Delphi technique was initially developed in America in the 1950s for the military defence administration to assess and develop cold war strategies. Great complexity and the subjectivity and scarcity of available information are typical of such problems. The technique inquires the views of experts, and seeks a consensus between them. This is done in an iterative manner. To minimise the potential bias of views and hang-ups of the process due to various peer pressures and in situ conflicts, the inquiries are normally carried out separately between each expert and the administrators of the exercise. Thus it is ensured that the participants remain anonymous, and possible distortions caused by rank, by dominant personalities or by experts simply expressing the views of the organisations they represent are minimised.

The Delphi technique is widely used in business, engineering and education, as well as in social and cultural studies. In the 1990s it has been applied, for instance, to technology assessments and technology policy foresights. In the early 1980s its use was extended to environmental management, since which time its range of applications has grown further (Wilson and Jones, 1996). A disadvantage of the technique is that it is basically not possible to reproduce the results once they have been published. Therefore, a pseudo-replication is often performed by running two similar panels independently. This approach was used in the case study investigated in this project.

The Delphi technique, or a technique close to it, has been used in LCA valuation in a few studies in the 1990. In addition to the Delphi I study, a phosphate study in the UK in 1994 (Wilson and Jones, 1994) and a similar study in the Nordic Countries in 1995 (Wilson and Jones, 1995) have applied the actual Delphi technique. A study on chemicals in Holland in the early 1990s (Annema, 1992) employed a Delphi-like technique, in which the composition of the panel was not limited to environmental experts only and consensus was sought in panel sessions. In Germany there is a traditional Delphi study going on at the Fraunhofer institute.

6.2 Procedure

A Delphi study proceeds stepwise. First, one or two expert panels are recruited. It has turned out that it is not necessary to have large numbers of experts on the panels. Eight members is often sufficient to develop a consensus view (Wilson and Jones, 1996). Next, each member is sent an initial questionnaire inquiring his or her opinions. Then, the completed questionnaires are returned to the monitoring team, who summarise the responses and circulate them back to the panel members, accompanied by a new reply form. In this first iteration, the panel members can compare their responses with those of their peer colleagues. There are three basic situations in which a member can stick with his or her opinion about a problem. First, there may be a complete consensus on that problem. In that case it is not necessary for any member to consider the problem further, since new opinions cannot add to the consensus. The second possibility is that there is general consensus, which the member does not share. In this case the member may choose either to move part or all of the way towards the majority opinion or to stick to his/her dissenting opinion. In the latter case the member is invited to state his/her case, and this statement is then circulated to the other panellists in the next round of the procedure. The third situation is that there is no general agreement on a problem. In actual fact, this was perhaps the most frequent situation in the Delphi I study. It was not recognised as a separate case, however, because the goal was to achieve consensus. Instead, this case was treated as the second case, considering the mean opinion as a (pseudo-) consensus. The iteration is repeated until the replied views converge. When there is no further progress towards consensus, the procedure is stopped. How much this process of "opinion work-out" may affect the ratings of the interventions is an interesting question. This is studied in chapter 8. Statistical analysis of the Delphi.

6.3 An evaluation of the Delphi method

Braunschweig et al. (1994) have evaluated expert judgement based on the Delphi technique from the LCA point of view as follows:

(1) **Completeness of the set of interventions covered**

> In principle it is possible to include as many interventions as necessary for a certain LCA. It is also possible to weight unquantified interventions, but how to formulate the corresponding task for the questionnaire might cause a problem.

(2) **Transparency**

> This is provided by reporting appropriate statistical figures and comments of individual peers. However, information about the criteria and preferences that each judge used to derive the scores is not reported. Instead, the rating procedure remains unstructured and untransparent. There should be a set of criteria which all peers should consider when scoring. The structure of the rating procedure, the statistical treatment of the results and some basic criteria to be considered might be standardised.

(3) **Content**

The range in scoring (1–100) does not reflect the relative importance of different environmental problems. To get an idea of the relations between the environmental importance of different interventions, threshold values could be used as the basis for a first estimation. A scoring on the basis of a kg of emission for each intervention seems to underestimate the relations between the environmental effects. There should be a set of criteria binding all judges, such as the degradation behaviour of substances in nature, reaction chains to a certain degree, life time of the intervention, and known environmental problems. This set could be discussed and reformed if necessary for each new weighting procedure in order to achieve a precise formulation of the rating problem which is most important.

(4) **Practicability**

The number of interventions and peers determine the extent of resources required. The method is time intensive and requires active participation of the peers.

6.4 Problems with judgement-based valuation of environmental impacts

Opinion-based valuation of environmental impacts poses particular problems stemming from the complexity of the evaluated issue, differences in the interests of stakeholders, and the technique applied in the evaluation procedure. Important aspects in the conduct of such an exercise are the stakeholder composition of the panel and the method used for soliciting opinions. Possible stakeholders are the general public, politicians, manufacturers, campaigners and experts. The selected panel should reflect the mutual importance of these different stakeholders in a justified manner. The problem is, however, which criteria would fulfil this requirement so that the preferences of each stakeholder group would be taken into account appropriately. Such criteria should include the rights of people in all countries (international), of all species (interspecial) and of all generations (intergenerational). In practice, of course, this is not possible. The idea of using panels of expert scientists relies on the (idealistic) assumption that experts have the best knowledge and morals to consider the effects of environmental interventions for the good of all stakeholders. The reliance on such an assumption is justified for the following reasons (Wilson and Jones, 1996):

a) Expert opinion will inform any environmental decision made by politicians, manufacturers, campaigners or the general public. Often the opinion will be moderated and possibly distorted by other groups, mainly by the media. Given the distortion which must occur in this process, it is more sensible to access the opinion directly from the expert source. There is considerable confusion among the general public about the nature and causes of many of the environmental problems facing us today, perhaps because of the distortions arising from the information provided by various stakeholders. Consequently, a valuation relying on public opinion would probably not reflect the importance of various interventions correctly.

b) Where experts are used to describe to the general public the likely effects of environmental problems and the options for tackling them, the public may simply feed back the experts' value judgements. The Landbank report (Wilson and Jones, 1996) mentions a study conducted recently in the USA, which possibly supports this view. It used two groups, one consisting of 400 members of the general public randomly selected in four US cities, and the other consisting of experts selected from college directories across the whole country. Both groups were invited to complete the same questionnaire about whether they agreed or disagreed with a number of policy options regarding two well-known problems – global warming and solid waste disposal. The lay group were then invited to attend a briefing where they were shown videos about the two issues, which were prepared by a group of experts – not those who had answered the questionnaire. After the briefing the lay group completed the questionnaire a second time. The study showed that while there were some significant disagreements between the lay and expert panel results in the first round, the number of differences were greatly reduced in the second questionnaire round. This result can be interpreted in different ways. The authors of the study concluded that lack of expertise does not prevent lay people giving thoughtful consideration to scientific issues, and that lay people can be trusted to make rational and sensible choices if they are properly informed. The writers of the Landbank report have drawn a rather different conclusion from the same evidence, which is that experts can be relied upon to make the same choices as lay people who have been given the same information, and that therefore it is not necessary to consult the general public, given the extra difficulties involved in eliciting their opinions.

c) Experts can be ahead of public and political opinion, as was the case in the 1960s over the issue of atmospheric testing of nuclear weapons Public opinion, at least in the Western democracies, was generally in favour of such tests, and it was only the campaigning by certain leading scientists which led to a test ban.

d) Experts are citizens, too, and frequently have a highly developed sense of social and environmental responsibility.

The complexity of the ecological systems makes it very difficult to reach convergent estimates of the environmental impacts on an objective bases. Quite inevitably, attitudes and values enter the process. Because of individuality, each expert may be assumed to have individual patterns, notional damage curves for environmental impacts of different interventions, based on which the valuation is made (Wilson, 1997–1998). Because of the different damage curves, it is not possible to reach a consensus between experts. There will always be disagreement about the absolute values of environmental harms.

Two typical damage curves are shown schematically in Figure 8. The first curve on the left has no threshold and damage increases rapidly at lower levels of the intervention. This type of the damage concept would be typical of a campaigner (egalitarian). An expert who is a specialist in the problems caused by the intervention could perhaps share a fairly similar view. The second curve on the right has a considerable threshold and damage is increases less rapidly even at higher levels of the intervention. This type of the damage view might be typical of an expert whose speciality is other than this intervention. Damage curves vary with regard to the threshold and to the rate assumed for the growth of the harm relative to the increase in the intervention. Extremes are a very rapidly increasing damage without any threshold, which is typical of an extremely egalitarian view, and a "zero-effect" curve with an infinite threshold, typical of an extremely individualistic view, which will see no damage at any level of the intervention. There is unlikely to be consensus among the representatives of these views.

(Wilson, 1997–1998) has found, based on tests with 100 experts from universities, that familiarity of the experts with the impacts of interventions influence the damage curves, the tendency being that interventions with greater familiarity are ranked higher. A problem with the Delphi technique is, however, the giving of feedback about damage curves. A solution to the problem could be to survey the familiarity beforehand, and customise questionnaires according to the results, so that the experts would concentrate on the interventions corresponding to their expertise. This would mean, however, that a method of combining the ratings of different expert groups into a common basis should be available. At the moment there is no such method.

Important considerations when selecting a data collection method are whether the experts should be interviewed face-to-face or by post, as a panel or individually, and as named or anonymous participants. A relevant data acquisition method is needed in order to reduce possible errors stemming, for instance, from cognitive factors, stress and fatigue. The layout of the questionnaire should be considered from the point of view that, for instance, positioning of the interventions in the questionnaire may influence the ranking (Wilson, 1997–1998), particularly for those interventions with which the respondent is less familiar. And last but not least, the task instructions should be clear from the very beginning of the process so that all judges understand the method exactly in the same way.

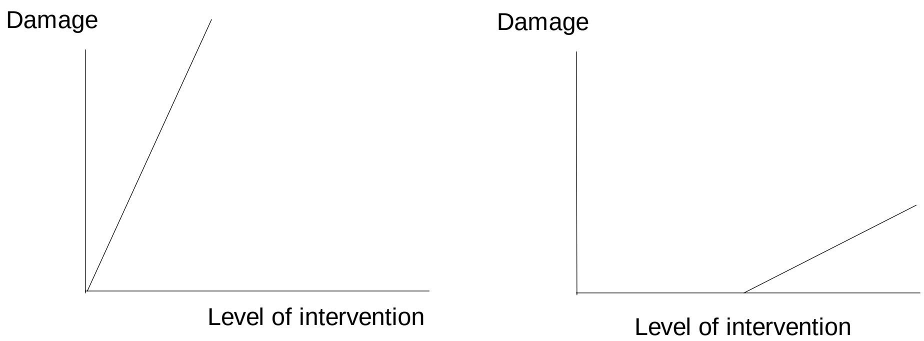

Figure 8. Schematic illustration of the possible differences between damage curves.

7. Delphi I study

7.1 Goal and scope

In the SIHTI II research programme, a project by Neste Corporation was accomplished during 1995, in which the Delphi method was tested for valuation of environmental loads. (Wilson and Jones, 1996). In addition to methodological development, the aim was to apply the results of the project to product development and to the comparison of alternative fuels.

The aim was to develop a method based on purely environmental rather than economic grounds – the basis of most of the other methods developed till then. An important advantage of the Delphi method was thought to be the good ability of a group of scientists with a good understanding of environmental problems to judge the relative significance of a number of interventions. That would produce the "true environmental significance" of an intervention compared with other interventions.

An important challenge was to have a suitable method for consulting a panel of experts, so that the participants could express their opinions honestly and free from all external pressures. This is a difficult task because, on the base of present knowledge, judgements of the relative significance of various interventions could easily be controversial. It is therefore important not to expose the panellists to public controversy or censure.

The project ("Delphi I") comprised the work of expert panels, some statistical analysis of the results, and application of the results to heating fuels. Delphi I did not include an analysis of the significance of the results or an examination of the possibilities for wider application and generalisation of the method.

Sixteen environmental experts from Finland, Sweden and Norway were involved in the project. Thus repeating the work as comprehensively in the near future would probably not be possible. The material produced in the project is utilised in this study for methodological development.

7.2 Set-up of panels

The process of assembling the panels was started by consulting a few known environmental experts, senior scientists of universities or of a national environmental protection agency. Invitations were then sent to 35 experts from the three Nordic countries, Finland, Sweden and Norway, to participate in the panel.

The initial contact letter and questionnaire comprised an invitation to participate in the evaluation of the environmental impacts of a collection of interventions. Neither at this stage nor later were the experts told about the reason for the actual evaluation process, about the particular case to which the interventions related, or about where the results of this evaluation were going to be used. They were told only that the evaluation would be used as source material to produce weights for the aggregate life cycle inventories of different interventions in order to compare life-cycle impacts of products. A £300.00 honorarium was offered. A short description of the Delphi method and a questionnaire with three tasks were also included.

Sixteen of the invited experts agreed to take part. Two panels, each of 8 members, one for control purposes, were formed. A short half of the experts were from universities, another short half from research institutes and the rest from government agencies. An auditor was chosen to ensure that the sample of experts was selected fairly and that the members had the requisite scientific standing.

7.3 Main phases

The main flow of the Delphi process in the Delphi I study is shown in Figure 9. It includes the selection of the panel, three questionnaire rounds and the final computing of valuation indexes. The process is described briefly in the following.

7.3.1 Initial questionnaire

The tasks in the initial questionnaire were as follows:

1) To rank 23 interventions in the order of importance according to their impact on the environment of Northern Europe (Finland, Sweden, Norway) and on the basis of the desirability of making a 1 per cent cut in the current annual flow of intervention. The criteria to be used were:

   a) urgency of reduction over the next 10–15 years
   b) the practical problems of reduction were to be ignored
   c) equal application across all sources (mobile, stationary etc.)
   d) independence of reductions in any other intervention
   e) and simultaneous protection of
      (i) human health and welfare
      (ii) the sustainability of ecological systems
      (iii) the sustainable use of non-renewable sources.

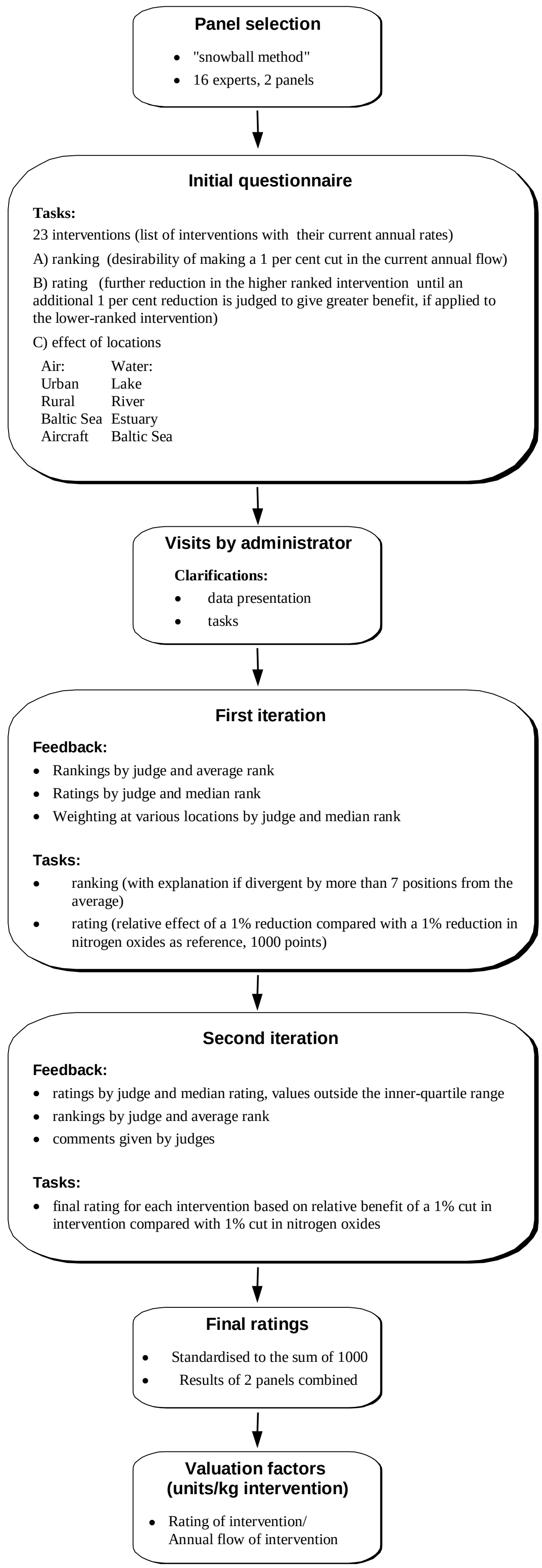

Figure 9. The main flow of the Delphi process in the Delphi I study.

2) To judge the relative impacts of 23 interventions by applying a further 1 per cent cut until a ”break-even point” is reached. The size of reduction in the first intervention should be equivalent (in terms of environmental benefits) to a 1 per cent cut in the second intervention; further reduction in the higher-ranked intervention is made until a point is reached at which an additional 1 per cent reduction is judged to give greater benefit, if applied to the lower-ranked intervention ”how much to cut an intervention before applying a 1% cut in the next”. At maximum, 100% of the higher level is eliminated before any cut to the next-ranked…

3) To consider the effects of location on some of the impacts

a) Air emissions:

(i) Urban (index of 1 000 to represent the environmental impact)
(ii) Rural
(iii) Baltic Sea
(iv) Aircraft

b) Water discharges:

(i) Lake (index of 1000)
(ii) River
(iii) Estuary
(iv) Baltic Sea
(v) Atlantic Ocean.

As background information, a list of interventions to be considered and their current annual rates was provided. The interventions, which comprised air emissions, water discharges, and resources, were given in the following order (Figure 10.):

Air emissions	I. Particulates	Q. Oils and greases
A. Ammonia	J. Pesticides	R. Phenols
B. Benzene	K. PM10	S. Phosphorus as P
C. Carbon dioxide	L. Sulphur dioxide	T. Suspended Solids
D. Carbon monoxide	M. VOCs	**Resources**
E. Heavy metals	**Water discharges**	U. Oil in ground
F. Methane	N. BOD+COD	V. P in ground
G. Nitrogen oxides	O. Heavy metals	W. Phosphorus
H. Nitrous oxide (N2O)	P. Nitrogen	X. Soil erosion

Figure 10. Intervention list for the initial questionnaire.

7.3.2 First iteration

Feedback from the initial questionnaire consisted of

1) ranking of the interventions by judge and average rank
2) rating of interventions by judge and median rank
3) weighting emissions at various locations (for air and water emissions) by judge and median rank.

Task instruction for the first iteration included

1) ranking

   As in the initial questionnaire. In addition, an explanation was requested if the ranking differed by more than 7 positions from the average rank (in the initial questionnaire).

2) rating

   Relative effect of a 1% reduction compared with a 1% reduction in nitrogen oxides, which is given an index of 1 000.

3) guidance

   a) independence of the reduction of interventions
   b) ignore the practicality of reduction
   c) originates in the 3 northern countries or elsewhere on behalf of those countries.

7.3.3 Second iteration

Feedback from the first iteration consisted of

1) data for each intervention in a table:

   a) ratings by judge (based on relative benefit as compared with a cut of 1% in nitrogen oxide) and median rating and lower and upper quartiles, values outside the interquartile range in bold face,
   b) rankings by judge and average rank
   c) and the comments made by any of the judges.

Task instruction for the second iteration included

1) rating

   Final rating for each intervention. Relative effect of a 1% reduction compared with a 1% reduction in nitrogen oxides, which is given an index of 1 000.

2) guidance

   a) relative benefit of a 1% cut in intervention compared with 1% cut in nitrogen oxides, not on a kg-to-kg basis
   b) emissions arising in the three Nordic countries
   c) all sources, anthropogenic activities included
   d) independence
   e) practical problems disregarded.

7.3.4 Computing the final ratings

Final ratings were standardised to a constant sum of 1 000, distributed between the 23 interventions to reflect the relative benefit of a 1% reduction in each. Thus the following ratings were obtained for the 23 interventions (Table 3):

Table 3. The final ratings from the Delphi I study.

Intervention	**Rating**
Carbon dioxide	189
Nitrogen oxides	113
Sulphur dioxide	72
Nitrogen as N	71
Phosphorus as P	68
VOCs	51
PM10	50
BOD+COD	46
Ammonia	38
Oil in ground	37
Nitrous oxide (N2O)	33
Heavy metals, air	31
Methane	27
Heavy metals, water	26
Benzene	24
Oils and greases	23
Carbon monoxide	22
Particulates	20
Pesticides	20
Soil erosion	15
P in ground	10
Phenols	8
Suspended solids	7
Total	1 000

A valuation factor was obtained for each intervention by dividing the rating of the intervention by its annual flow.

8. Statistical analysis of the Delphi procedure

8.1 General

The following analysis of the Delphi procedure application to the rating of the environmental harmfulness of interventions is based on a real Delphi study by Neste Corporation. The data used in the calculations are taken from the report of that study (Wilson and Jones, 1996). The figures given for Finland in the table on page A1.7 of the report are used for the annual intervention volumes. The data for Norway and Sweden are not utilised, because of many missing values. The missing value for PM_{10} for Finland is assumed to be 10% of the particulate emissions. For soil erosion the value used is 10 550 ha/a , which is an estimated rate of growth of other than bio-productive land in Finland during 1983–1993 (WRI, 1997).

Only the overall indexes are analysed here. In the Delphi I study, data were collected also for the effects of locations in order to take account of the differences in the impacts of the interventions in different types of environment. For air emissions the locations considered were urban, rural, Baltic Sea and aircraft, and for water discharges the locations were lake, river, estuary, Baltic Sea and Atlantic Ocean. The data were collected only in the first round of the study, because the results were found satisfactory enough so that further iteration was not necessary. Therefore, the results of the effects of locations are not treated here any further.

8.2 Problem of indexing

8.2.1 Logic of the rating

Mathematically the environmental harm (H_i) of an intervention (i) can be considered, independently on how its measured, as a product of the volume (V_i) and the (virtual) environmental harmfulness (H_i''') of an intervention, i.e.

$$H_i = H_i''' V_i \, . \tag{7}$$

The environmental harmfulness of an intervention normally depends in a complex way on the context of the intervention, that is the space, time, state of the environment and the volume of the intervention, but also on the evaluation approach (value hierarchy) applied. It is extremely difficult to build up an objective and correct picture of the environmental harms of different kinds of emissions. Therefore, a meaningful assessment of environmental harm as well as that of the beneficiality of optional abatement actions require special knowledge.

Environmental beneficiality (B_i''') can be understood as the gained reduction of environmental harm per unit of reduced intervention, which, when taking into account the above, means that

$$B_i''' = \frac{-\Delta H_i}{-\Delta V_i} \approx \frac{-H_i'''\Delta V_i}{-\Delta V_i} = H_i''' \,. \qquad (8)$$

Here, specific environmental harmfulness is considered not to change significantly, that is $\Delta H_i''' = 0$. This assumption, which has been employed practically by all environmental valuation applications based on the Delphi technique so far, means that the environmental harm depends linearly on the volume of the intervention. In reality, however, the harm to the environment is a non-linear function of the intervention volume.

Yet, it is normal that there are regions where the specific environmental harmfulness is rather constant and, thus, the harm function close to linear. For instance, for a logistic type of dependency between the volume and the harm of an intervention, there are two reasonably linear regions of the harm functions, one at low values of the intervention volumes and one at high values, before the saturation range. Elsewhere the function is clearly non-linear, which means that the specific harmfulness (harm per volume) is changing according to the volume of the intervention. This is illustrated in Figure 11. It is important to note that in the Delphi procedure the harmfulness indexes of the interventions are calculated on the assumption that linearity is in effect. The Delphi procedure is not an exception in this respect though. The same assumption presently underlies practically every LCA valuation method.

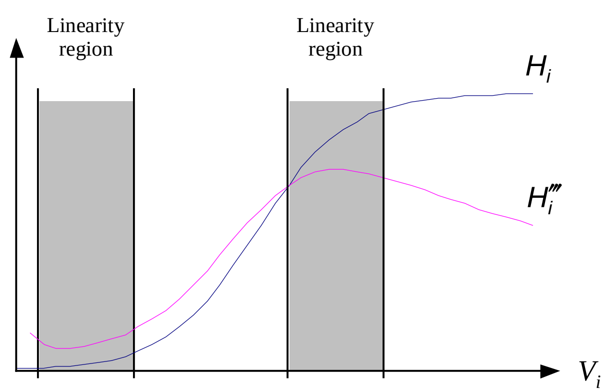

Figure 11. Illustration of the linearity regions of a logistic environmental harm function.

Environmental benefits gained by reducing intervention *i* and intervention *i+1* are equal when

$$B_i'''\Delta V_i = -\Delta H_i = B_{i+1}'''\Delta V_{i+1} = -\Delta H_{i+1} \tag{9}$$

Let us then assume that $\Delta V_i = pV_i^o$ and $\Delta V_{i+1} = pV_{i+1}^o$ at a point where an equal relative reduction (p) in each intervention gives the same environmental benefit. At this very point it must hold for the volumes and the harmfulness of the interventions that

$$V_i^o H_i''' = V_{i+1}^o H_{i+1}''' \tag{10}$$

In other words, an equal relative reduction in two interventions can give the same absolute environmental benefit only when the absolute harm levels of the interventions are the same. Since in the Delphi procedure the present harm level of the higher-ranked (*i*) intervention is higher than that of the lower-ranked (*i+1*), the higher-ranked intervention must be reduced in order to reach the equilibrium.

In the first round of the Delphi I study, the judges were asked to assess an equilibrium point, or "break-even point" as it was called in the instructions, between two subsequently ranked interventions (higher-ranked *i* and lower-ranked *i+1*). This was formulated as "*we will make further cuts of 1 per cent in the higher-ranked intervention until a point is reached when an additional 1 per cent reduction is judged to give greater environmental benefit if applied to the lower-ranked intervention rather than the higher-ranked one* " (Wilson and Jones, 1996, p. A1.3). The task was summarised as follows: "*In summary, therefore, the procedure is to decide how much you would wish to cut an intervention before applying a 1% cut to the next intervention in the ranking list. If the cut is switched to the lower-ranked intervention straight away, then a 1 is entered in the "Ratio" column. If the switch occurs at, say, the 5% level, then a value of 5 is entered in the grid. If there is no level at all at which a switch would take place, that means that the emissions of the higher intervention should be totally eliminated before any cut is applied to the next-ranked intervention – then a value of 100 would be entered.*"(Wilson and Jones, 1996, p. A1.4). These instructions clearly tell one to assess the required reduction of the higher-ranked intervention to reach the harm level of the lower-ranked one. So, to reach the equilibrium point, one estimates that the higher-ranked intervention should be reduced, say, by p_i per cent, whilst the lower-ranked one should not be reduced at all. At the equilibrium, then, a 1 per cent reduction in either of the interventions will give the same environmental benefit. If one chooses here to reduce the higher-ranked intervention by an additional 1%, meaning p_i +1 per cent total reduction, it becomes slightly more beneficial to make the next 1% cut in the lower-ranked intervention rather than in the higher-ranked one. Thus we can write the benefit equilibrium in terms of volumes as

$$V_i^o = (1 - p_i)V_i \tag{11a}$$

$$V_{i+1}^o = V_{i+1} \tag{11b}$$

Assigning this to (4) gives

$$(1 - p_i)V_i \; H_i'''' = V_{i+1} H_{i+1}'''' \tag{12}$$

from which we get, when we denote $(1 - p_i)$ with $\frac{1}{r_{i,i+1}}$, an important equation between the subsequently ranked interventions,

$$H_{i+1}'''' = \frac{V_i}{r_{i,i+1} V_{i+1}} H_i'''' . \tag{13}$$

One might call this the basic equation of the Delphi procedure, because it determines the eventual harmfulness indexes of all interventions. The rating factor ($r_{i,i+1}$) in this equation is the ratio between the environmental benefits of a 1% cut, as we can see when we write

$$\frac{\Delta_{1\%} H_i}{\Delta_{1\%} H_{i+1}} = \overbrace{\frac{1\% V_i \; H_i''''}{1\% V_{i+1} H_{i+1}''''}}^{\text{By definition}} = \overbrace{\frac{r_{i,i+1} V_i \; H_i''''}{V_i \; H_i''''}}^{\text{Based on (7)}} = r_{i,i+1} . \tag{14}$$

This factor ($r_{i,i+1}$) is not equal to the relative reduction (p_i) required for the higher-ranked intervention to reach the harm level of the lower-ranked intervention, and, therefore, cannot be obtained from it, unlike the task description on the first round of the Delphi I study misleadingly suggested: "*Assume, therefore, that a further 1% cut in the annual rate is available and you can choose which intervention to apply the cut. Judge n has two choices: s/he can either make a further cut of 1 per cent in calcium sulphate, giving a total cut of 2% with no reduction in rock salt, **OR** s/he can apply the 1 per cent cut to rock salt, so that both calcium sulphate and rock salt are cut by 1%. If the second choice is made, this implies that a cut of 1% in each of the interventions has the same environmental impact, and the procedure ends at this point. Expert X did not make this choice however, and applied the additional cut to calcium sulphate. A further cut of 1% is now made available, so that the choice is now a 3% reduction in calcium sulphate with no reduction in rock salt, or a 2% reduction in calcium sulphate plus a 1% cut in rock salt. Judge n chose the latter, implying that a 1% cut in calcium sulphate has about twice the environmental benefit of a 1 % cut in rock salt. A value of 2 is therefore entered in the column labelled 'Ratio' in the recording grid against calcium sulphate. This ends this particular step.*" (Wilson and Jones, 1996, p. A1.4). The point of equal harm levels was not what was meant by the break-even point in the Delphi I study though. Instead, it was the break-even point of the marginal environmental benefits which was looked for.

There are two logically different approaches to the equilibrium of environmental beneficiality between two interventions. These are illustrated in Figure 12. One, which is discussed above, could be called the principle of the ’same harm level’ and the other the principle of the ’same benefit’. Which one of these principles was actually meant in the Delphi I study was not quite clear from the instructions of the first query round and this has possibly affected the results. Three sentences of the above quotation tell about this uncertainty. “A further cut of 1% is now made available, so that the choice is now either a 3% reduction in calcium sulphate with no reduction in rock salt, or a 2% reduction in calcium sulphate plus a 1% cut in rock salt. Judge n chose the latter, implying that a 1% cut in calcium sulphate has about twice the environmental benefit of a 1% cut in rock salt. A value of 2 is therefore entered in the column labelled 'Ratio' in the recording grid against calcium sulphate ” (Wilson and Jones, 1996, p. A1.3). The first sentence refers to the 'same harm level' and the latter two sentences to the 'same benefit' or marginal equilibrium, which was also what was assumed when interpreting the rating results in the study. It is obvious that when the magnitudes of the problems differ greatly, for instance due to a different order of the volumes of the interventions, then the relative change required in the larger intervention to bring it to the same level as the smaller one (Figure 12a) is considerably larger than what is needed for equal marginal changes of the harm (Figure 12b).

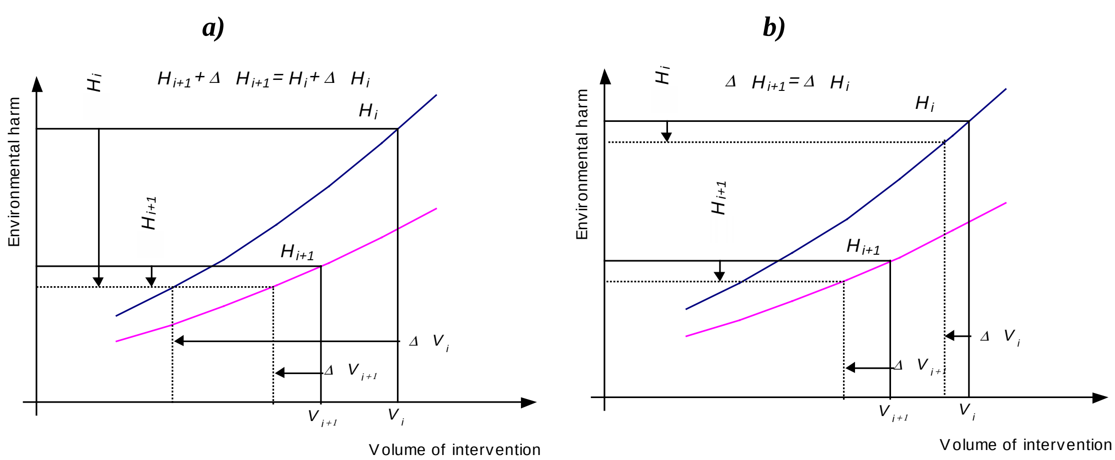

Figure 12. Two different approaches to the break-even point of two environmental interventions. a) Equal level of environmental harm, b) Equal reduction of environmental harm level.

Nevertheless, the marginal benefit equilibrium can also be put in the basic format of equation (7). Considering that the given ratio ($r_{i,i+1}$) between subsequently ranked interventions is actually the ratio between the potential environmental benefits for a one per cent, or in general, for a p per cent reduction in each of the compared interventions, one can write the benefit balance as

$$pV_i H_i'''' = r_{i,i+1}\, pV_{i+1} H_{i+1}''''. \tag{15}$$

This is easily transformed into equation (7). Thus the ratio, $r_{i,i+1}$, is indeed the ratio of the environmental benefits gained when both interventions are reduced relatively by the same amount. It is also the ratio of the total environmental harms in the measures of each judge. This was not quite clear in the first round of the Delphi I study, which caused some confusion. The possibility of confusion was also recognised during the Delphi I study. Still, even though the first round results were less important to the eventual indexes, they were important to the assessment of the consensus and, thus, to the choices made in subsequent steps of the Delphi process. Therefore, to avoid confusion, the task instructions for the first round need to be developed for future applications. In the following we study the effect which the choice of the approach to the break-even point has on the specific harmfulness indexes.

From equation (7) one can obtain a recursion for a series of specific harmfulness indexes $\{H_1'''', H_2'''', H_3'''',...\}$, so that

$$H_{i+1}'''' = \frac{V_i}{r_{i,i+1} V_{i+1}} H_i'''', i = 1,2,3,... \Leftrightarrow H_i'''' = \frac{V_1}{V_i} d_i H_1'''', i = 2,3,... \tag{16}$$

where $d_i = \dfrac{1}{\prod_{j=1}^{i-1} r_{j,j+1}}$. (17)

We may call d_i a 'size factor' on an implicit environmental harm scale, because from the right part of (10) we get for the relative (to the first-ranked) total harm

$$\frac{H_i'''' V_i}{H_1'''' V_1} = \frac{H_i}{H_1} = d_i, i = 2,3,... \tag{18}$$

Implicity enters the equation because we don't know the magnitude of the base of the relative scale, which in this case is the first-ranked intervention.

To use the recursion to build up an index statistics it is necessary to fix the series at some point, common for all judges. Fixing can be done in a number of ways. In the Delphi I study, total nitrogen oxide was selected as a basis because some consensus was found in its ranking. A preliminary *K*-entropy analysis also supports this finding. This selection of basis corresponds to the assumption that the total environmental harm of the nitrogen oxide emissions is equal for each judge. In practise, it means that the total harm of the NO_x emissions is given a fixed value of 1000 for every judge.

If we assume that $r_{j,j+1}$ is interpreted rather as the required reduction in the higher-ranked intervention to reach the harm level of the lower-ranked one than as the ratio between the benefits of equal relative cuts in both interventions, we can replace $r_{j,j+1}$ with $\frac{1}{1-\frac{r_{j,j+1}}{100}}$ in (11), and, after dividing the equation thus obtained by the original, we obtain in the total NO_x base the ratio of the indexes based on the alternative logical approaches to the break-even point of environmental harmfulness,

$$\frac{H_i'''^{\,b}}{H_i'''^{\,a}} = \frac{d_i^{\,b}}{d_i^{\,a}} = \frac{\prod_{j=1}^{i-1}(1-\frac{r_{j,j+1}}{100})\prod_{j=1}^{i-1} r_{j,j+1}}{\prod_{j=1}^{i_{NO_x}-1}(1-\frac{r_{j,j+1}}{100})\prod_{j=1}^{i_{NO_x}-1} r_{j,j+1}} . \qquad (19)$$

Here, superscript ^{a} refers to the default approach, which is based on equal reduction of the environmental harm level, and superscript ^{b} to the alternative approach, which is based on an equal level of environmental harm. In the former case, the value in the column labelled 'Ratio' in the recording grid is interpreted as the ratio of the environmental benefits gained by equal relative reductions (1%) of the higher- and lower-ranked interventions (default), and in the latter case the 'Ratio' is interpreted as the percentage of reduction of the higher-ranked intervention to bring it onto the same level as the lower-ranked intervention (alternative). The ratios of the average indexes in the total NO_x base obtained according to these two different approaches (alternative index per default index) in the first round of the Delphi I study are shown in
Figure 13.

Figure 13 shows that the ratio of the average indexes vary from less than 0.001 to about 3. Typically, the index calculated according to the 'same harm level' approach is lower than the index obtained according to the 'same benefit' approach. This is so because the rating factor of the 'same harm level' approach is smaller than that of the 'same benefit' approach, except when the value of 'Ratio' is close to one or a hundred (Figure 14). This evens out large 'jumps' in the index series which would occur at high values of $r_{j,j+1}$ in the 'same benefit' approach. Thus the indexes of those interventions which are ranked higher than NO_x get smaller, and the indexes of those interventions which are ranked lower than NO_x usually get bigger in the 'same harm level' approach than in the 'same benefit' approach.

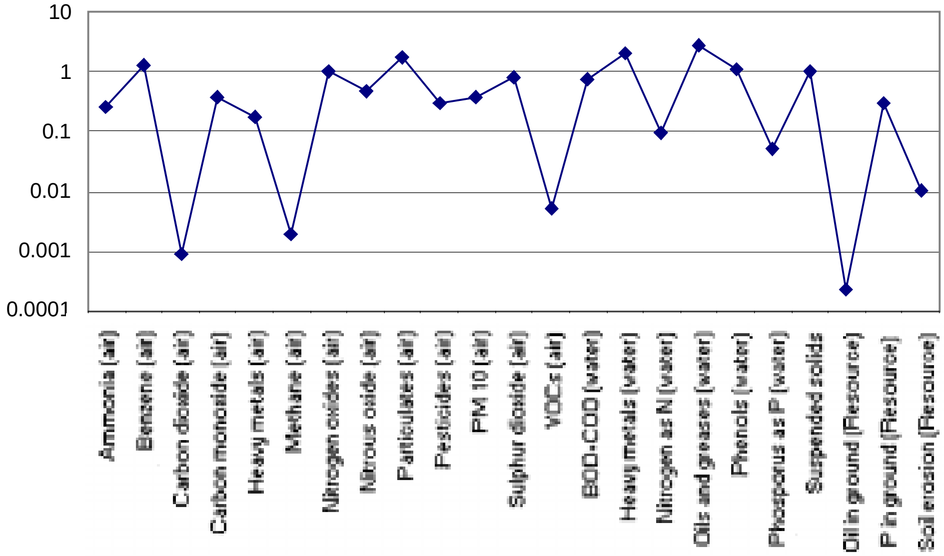

Figure 13. Ratios of the average indexes calculated according to the 'same harm level' approach to the respective average indexes calculated according to the 'same benefit' approach. The basis of the indexes is the total harm level of the NO_x emissions.

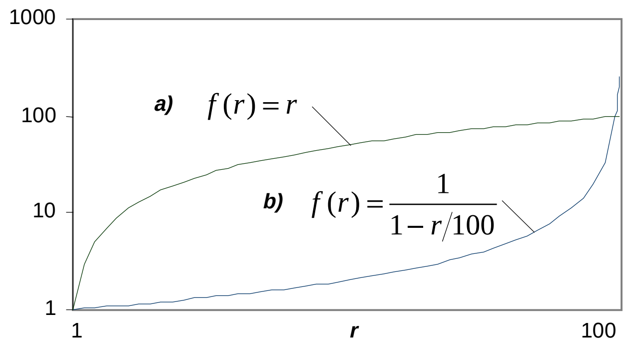

Figure 14. Rating factor ($f(r)$) when the value in the recording grid (r) is **a)** *interpreted as the ratio of the benefits of an equal 1% cut in the interventions ('same benefit' approach) and* **b)** *as the required percentage reduction in the higher-ranked intervention to reach the harm level of the lower-ranked intervention ('same harm level' approach).*

Thus, for many interventions, possible confusions in the interpretation of the rating task would have caused substantial errors in the recorded data and, consequently, would have mislead the conclusions drawn from them according to the default interpretation, such as the positions of individual judges with respect to the other judges in the rating of different interventions.

8.2.2 Basis of the indexes

In order to compute the environmental harmfulness indexes for the interventions, the specific harmfulness of the first-ranked intervention (H_1) must be determined in some uniform way for each judge. There is no proven way to do it though. Therefore, one must rely on some postulate, based on which the index system will be fixed, assuming that the postulate is true. A fundamental assumption underlying any such postulate is, however, that environmental harm can be expressed as a single quantity. This is a rather bold assumption considering all the different perspectives from which the importance of the environment may be assessed, even if the overall aspect was limited to ecology, as it was in the Delphi I study. Questions about the importance of the future and the present, dead and living nature, and humans and other species are just a few examples of the complexity of the problem. Nevertheless, the assumption that all the different aspects can be converted into a single quantity is necessary for the Delphi procedure.

In the Delphi I study the first postulate for the index base was that the total environmental harm of the NO_x emissions is the same for each judge on an implicit harm scale. The measures of the harm scale are unknown, which means that measurements cannot be made in absolute terms. This does not cause a problem though, because indexes are used for comparisons, i.e. relatively, and thus it is not necessary to know their absolute values. The selection of the total NO_x emissions as the basis of the indexes in the Delphi I study was justified by the finding that this was the intervention for which there was closest agreement in both panels on the first round of the study. This postulate will be called the 'total of NO_x' postulate.

Eventually, however, the postulate was changed from 'total of NO_x' to the 'total of all interventions' for the final results of the Delphi I study. This postulate supposes that the total environmental harm of all interventions is the same for each judge on an implicit harm scale. Its idea is that each judge has a similar comprehension of the magnitude of the aggregate total harm which the studied interventions together inflict on the environment. This is a justified assumption, because environmental experts may be anticipated to be similarly aware of the dimensions of all environmental problems and thus to have a common basis to assess the relative importance of each intervention. Since this postulate was introduced at the end of the Delphi I study in order to standardise the indexes, it will be analysed separately from the other three postulates under the title standardisation of the final indexes.

The departure postulate chosen here to study the effect of the index basis is that the specific harmfulness of the first-ranked intervention (H_1''') is the same on an implicit harmfulness scale for each judge. The idea of this postulate, which will be called the 'first-ranked' postulate, is that the judges rank the interventions based on their specific harmfulness rather than on the total harm they are causing. This is a justified assumption for two reasons. Firstly, the environmental harm depends directly on the potential of an intervention to cause problems, and this potential is often more certainly known than the total effects, which usually stem from several simultaneous interventions. Secondly, the instructions given to the judges in the Delphi I study set no explicit priorities for the approach in this respect. It is true though that the ranking and rating procedures in the Delphi I study favoured the total harm approach in the iteration rounds, which is why we also study other possible postulates employing the total harm approach.

To determine the effects of the base selection and functioning of the base postulates, in the following we will study indexes based on our departure postulate, on the two postulates of the Delphi I study, and on a fourth series of indexes based on another postulate about the total environmental harm. The postulate of this fourth basis is that the total environmental harm of the first-ranked intervention is the same for each judge on an implicit harm scale. The justification of this postulate, which will be called the 'total of the first-ranked' postulate, is in the specialisation of the experts. Many of them are extremely familiar with a certain group of interventions and their environmental effects, but do not necessarily know as much about interventions outside this particular group. Accordingly, they could be expected to focus on their special interventions. For each expert these interventions obviously represent the greatest environmental harm. The idea would then be to regard these greatest harms as being equal in the magnitude.

The key postulates of the four bases studied are summed up in Table 4. The kind of conformity assumed between the judges' thinking is the separating factor between them. Each postulate can be justified to a certain extent from the expertise point of departure, but none of them is hardly correct. The selected postulates are not the only ones possible though. There are several other postulates which could be equally well justified. Such postulates might include, for instance, that the environmental harm of sulphur dioxide emissions, or some other individual intervention, is conceived similarly by each judge. Still another possible assumption would be that the judges share the same view about the total environmental harm of water emissions. However, none of the postulates could be proven correct, and, therefore, extensive study of the variety of possible postulates would hardly bring much additional information to the problem of the base selection. Even the apparent convergence of the index distributions could not be taken as evidence of the postulate's correctness, because of the possibility that the assignment required by the base postulate might cause the convergence even though the postulate per se were not correct. On the other hand, a wide dispersion of indexes would not necessarily mean that the postulate is completely incorrect, because judges might think similarly about the environmental harm of the reference intervention, but still disagree about other interventions' harms. For these reasons we limit the analyse to the four postulates given in Table 4, which we believe will enlighten perhaps the most difficult problem of the Delphi procedure, namely the selection of the basis for the index statistics.

Table 4. The key postulates for the four different bases of the harmfulness indexes.

Base	**Assignment**	**Key postulate**
'first-ranked'	Index of 1st ranked = 1000	Specific harmfulness (per unit of intervention) of the 1st ranked intervention is the same for each judge.
'total of the first-ranked'	Total of 1st ranked = 1000	Total harm of the 1st ranked intervention is the same for each judge.
'total of NO_x'	Total NO_x = 1000	Total harm of NO_x is the same for each judge.
'total of all'	Total all interventions = 1000	Aggregate total harm of all interventions is the same for each judge.

In practice our departure postulate means that the specific harmfulness of each first-ranked intervention is given a value of 1000. The indexes calculated for each intervention on this basis are given in detail in Appendix A. The corresponding coefficients of dispersion are shown in Table 5. These coefficients are normalised standard deviation parameters, which have been obtained by dividing the actual standard deviation by the average of the distribution. The indexes are spread over a wide range. The coefficient of dispersion ranges from 1.34 to 3.99 over all study rounds. If we interpret this variation according to normal distribution, it means that the 95% confidence interval is in the range from ±2.6 to ±7.8 times the average of the distribution. In other words, the results are highly uncertain. The 95% confidence interval for the population mean, i.e. the index figure, is within 0.66 (the best case) and 1.96 (the worst case) times the observed sample mean. The spread of the indexes is somewhat narrower in the second and third rounds than in the first round, but the uncertainty is not essentially reduced. The converging tendency in the distributions is not consistent either. While the results of the second round seem to be considerably more compact than those of the first round (ratio of the average dispersion coefficient is 2.59/3.37), there is a slight increase in the average dispersion coefficient between the second and third rounds (2.87/2.59). The conclusion is that the indexes based on the 'first-ranked' postulate, i.e. that the specific harmfulness of the first-ranked intervention is the same for each judge, are highly uncertain and, therefore, the relevance of this postulate must be doubted.

It should be noted, however, that the postulate employed in the Delphi I study was quite different and was explicitly offered to the judges during the second and third rounds.

This may have driven the results towards an agreement with that particular postulate, and, consequently, towards a disagreement with other postulates.

Table 5. Coefficients of dispersion (ratio of the standard deviation to the average) of the harmfulness indexes in different rounds. Base: First-ranked = 1 000 for each of the 16 judges.

	Coefficient of dispersion		
Intervention	**1st Round**	**2nd Round**	**3rd Round**
Ammonia (air)	3.91	2.19	2.17
Benzene (air)	3.74	2.61	2.13
Carbon dioxide (air)	1.53	1.46	1.78
Carbon monoxide (air)	2.62	2.74	3.64
Heavy metals (air)	3.84	1.87	2.30
Methane (air)	2.37	2.75	3.03
Nitrogen oxides (air)	2.13	2.16	2.39
Nitrous oxide (air)	3.39	2.05	2.58
Particulates (air)	3.88	2.39	2.82
Pesticides (air)	3.95	1.97	3.31
PM 10 (air)	2.97	3.53	3.76
Sulphur dioxide (air)	2.48	2.83	3.10
VOCs (air)	3.60	1.93	2.07
BOD+COD (water)	3.93	3.53	3.57
Heavy metals (water)	3.93	2.63	3.16
Nitrogen as N (water)	3.04	2.40	2.95
Oils and greases (water)	3.96	3.26	3.74
Phenols (water)	3.99	1.75	1.34
Phosphorus as P (water)	3.92	2.51	3.28
Suspended solids	3.99	3.75	3.94
Oil in ground (Resource)	2.43	3.45	3.85
P in ground (Resource)	3.99	2.74	1.70
Soil erosion (Resource)	3.87	3.13	3.38
Minimum	1.53	1.46	1.34
Maximum	3.99	3.75	3.94
Average	3.37	2.59	2.87

Based on the wide dispersion of the indexes in the 'first-ranked' base it might be anticipated that the selection of the basis of the indexes would not essentially change the spread of the index distributions. A judge's index series is transformed, for example, from the first-ranked basis into the total NO_x basis by diving each term by the size factor of NO_x in the first-ranked basis and the volume of the first-ranked intervention. Thus the change of the basis affects the distribution of an intervention (X) among the judges (1–16) so that

$$\overbrace{H'''_{X,1}, H'''_{X,2}, \ldots, H'''_{X,16}}^{Base: 1^{st}\text{ specific}=1000} \rightarrow \overbrace{\{\frac{H'''_{X,1}}{V_{1,1}d_{NO_x,1}}, \frac{H'''_{X,2}}{V_{1,2}d_{NO_x,2}}, \ldots, \frac{H'''_{X,16}}{V_{1,16}d_{NO_x,16}}\}}^{Base:\text{ Total }NO_x=1000}. \quad (20)$$

Because H'''_1 is constant (=1000) for each judge in the first-ranked basis, we can write the new distribution also as

$$\overbrace{H'''_{X,1}, H'''_{X,2}, \ldots, H'''_{X,16}}^{Base:\text{ Total }NO_x=1000} = \{\frac{d_{X,1}}{V_{1,1}d_{NO_x,1}}, \frac{d_{X,2}}{V_{1,2}d_{NO_x,2}}, \ldots, \frac{d_{X,16}}{V_{1,16}d_{NO_x,16}}\}H'''_1 \quad (21)$$

Since all the indexes, including that of NO_x, are distributed over a wide range in the first-ranked basis, as the dispersion coefficients in Table 5 show, it is possible, although quite unlikely, that the selection of some other basis had major convergence effects on the distributions, other than that of the new basis. To study this possibility two other sets of distributions were produced, one with the total NO_x emissions as the base, as in the Delphi I study, and the other with the total of the first-ranked intervention as the base. The base value is 1000 in both cases. In the former case this value is assigned to the total harm of NO_x emissions and in the latter case to the total harm of each first-ranked intervention. The distributions of the index values in these cases are reported in detail in appendixes B and C.

Table 6 gives an overview of the development of the distributions of the indexes based on the 'total of the first-ranked' postulate, i.e. that the total environmental harm of the first-ranked intervention is the same for each judge on an implicit harm scale. Again the spread of the index values is wide for all interventions. The distributions are, however, significantly narrower than for the 'first-ranked' base. The dispersion coefficient varies from 0.62 to 2.81. The average dispersion coefficient is 40 to 64 per cent lower than in the 'first-ranked' base. The converging tendency in the distributions is consistent. From the first to the second round, the reduction of the average dispersion coefficient is about 42% and from the second to the third round about 12%. After the third round (second iteration round) the coefficient of dispersion is less than one for about half of the interventions. However, the uncertainty of the indexes is high, the 95% confidence interval being from ±2.2 to ±4.3 times the average of the distribution. The conclusion is that the indexes based on the 'total of the first-ranked' postulate are less dispersed than those calculated on the 'first-ranked' postulate. Nevertheless, they are very uncertain and, therefore, this postulate must be questioned as well.

Table 6. Coefficients of dispersion (ratio of the standard deviation to the average) of the harmfulness indexes in different rounds. Base: Total of the first-ranked = 1 000 for each of the 16 judges.

	Coefficient of dispersion		
Intervention	**1st Round**	**2nd Round**	**3rd Round**
Ammonia (air)	1.87	1.25	0.93
Benzene (air)	2.33	1.30	1.16
Carbon dioxide (air)	1.10	0.68	0.64
Carbon monoxide (air)	2.06	0.93	1.00
Heavy metals (air)	1.94	1.14	1.02
Methane (air)	1.81	1.41	1.11
Nitrogen oxides (air)	1.17	0.69	0.62
Nitrous oxide (air)	2.01	1.19	1.07
Particulates (air)	1.89	1.33	1.07
Pesticides (air)	1.91	1.12	0.94
PM 10 (air)	1.56	1.00	0.80
Sulphur dioxide (air)	1.55	0.93	0.75
VOCs (air)	1.94	1.03	0.77
BOD+COD (water)	2.12	1.34	1.18
Heavy metals (water)	2.38	1.11	0.98
Nitrogen as N (water)	1.52	0.95	0.76
Oils and greases (water)	2.46	1.11	0.87
Phenols (water)	2.51	1.41	1.70
Phosphorus as P (water)	2.41	0.98	0.80
Suspended solids	2.79	1.65	1.49
Oil in ground (Resource)	1.58	1.44	1.29
P in ground (Resource)	2.81	1.40	1.39
Soil erosion (Resource)	2.40	1.54	1.28
Minimum	1.10	0.68	0.62
Maximum	2.81	1.65	1.70
Average	2.01	1.17	1.03

The third postulate analysed is the 'total of NO_x' postulate, which was employed in the Delphi I study. Its principle idea is that the total environmental harm of the NO_x emissions is the same for each judge on an implicit harm scale. In the second and the third round the judges were offered this postulate by asking them to give their rates relative to the total NO_x emissions, which were given a value of 1 000 in the recording grid. Table 7 shows the resulting development of the index distributions. It appears that the average dispersion of the indexes in the first round is narrower than for the 'first-ranked' base but wider than for the 'total of the first-ranked' base. A full comparison of the dispersion coefficients in the first round is presented in Table 8. Dispersion reduces when shifting from the 'first-ranked' base to the 'total of NO_x' base, except for a few interventions, such as carbon dioxide emission. An interesting finding in Table 8 is that the basis of the 'total of the first-ranked' gives the narrowest average dispersion in the first round, about 29% narrower than the 'total of NO_x' base. This suggests that the widest consensus among the judges was not perhaps about the environmental harm of the NO_x emissions but about the harm of the first-ranked intervention.

Based on this observation, the first-ranked intervention might have been an alternative to NO_x emissions as a reference for the iteration rounds. In practice this would have meant that, instead of NO_x emissions, the fixed value (1 000) would have been pre-recorded for the first-ranked intervention according to each judge, and the judges would have been asked to give their ratings for other interventions relative to that. However, the final problem with this approach would have been how to bring the obviously different conceptions of the judges about the environmental harm into a common scale. This problem would occur with all index statistics which employ variable interventions as a basis, and would require a transformation of the harms of the different base interventions into a common scale, for instance, into a monetary one. Such an assessment would rather inevitably bring in other than ecological values, which were tried to be avoided in the Delphi I study. From this point of view, the 'total of NO_x' base, which enabled measurement of the environmental harms relative to a constant reference without additional transformations, seemed an appropriate choice.

Table 7. Coefficients of dispersion (ratio of the standard deviation to the average) of the harmfulness indexes in different rounds. Base: Total of the NO_x emissions = 1 000 for each of the 16 judges.

	Coefficient of dispersion		
Intervention	**1st Round**	**2nd Round**	**3rd Round**
Ammonia (air)	3.40	0.68	0.61
Benzene (air)	2.03	1.07	0.79
Carbon dioxide (air)	3.72	2.40	2.44
Carbon monoxide (air)	3.57	3.58	3.69
Heavy metals (air)	3.73	0.72	0.72
Methane (air)	3.97	3.57	3.64
Nitrogen oxides (air)	0.00	0.00	0.00
Nitrous oxide (air)	2.11	3.06	3.04
Particulates (air)	1.77	1.32	1.27
Pesticides (air)	2.54	0.88	0.74
PM 10 (air)	2.52	1.29	1.32
Sulphur dioxide (air)	1.73	1.40	1.32
VOCs (air)	3.45	1.53	1.63
BOD+COD (water)	2.62	1.05	0.91
Heavy metals (water)	1.82	0.79	0.75
Nitrogen as N (water)	3.86	1.62	1.38
Oils and greases (water)	1.72	1.66	1.69
Phenols (water)	2.87	1.18	1.51
Phosphorus as P (water)	3.55	1.72	1.41
Suspended solids	3.33	2.50	2.33
Oil in ground (Resource)	4.00	3.76	3.85
P in ground (Resource)	2.80	1.07	1.23
Soil erosion (Resource)	3.95	1.26	1.45
Minimum	0.00	0.00	0.00
Maximum	4.00	3.76	3.85
Average	2.83	1.66	1.64

Table 8. Comparison of the coefficients of dispersion (ratio of the standard deviation to the average) of the harmfulness indexes in the first round.

Intervention	1st Rank specific	NO_x	Δ from 1st Rank	1st Rank Total	Δ from NO_x
Ammonia (air)	3.91	3.40	–13.0%	1.87	–45.0%
Benzene (air)	3.74	2.03	–45.7%	2.33	14.8%
Carbon dioxide (air)	1.53	3.72	143.1%	1.10	–70.4%
Carbon monoxide (air)	2.62	3.57	36.3%	2.06	–42.3%
Heavy metals (air)	3.84	3.73	–2.9%	1.94	–48.0%
Methane (air)	2.37	3.97	67.5%	1.81	–54.4%
Nitrogen oxides (air)	2.13	0	–100.0%	1.17	
Nitrous oxide (air)	3.39	2.11	–37.8%	2.01	–4.7%
Particulates (air)	3.88	1.77	–54.4%	1.89	6.8%
Pesticides (air)	3.95	2.54	–35.7%	1.91	–24.8%
PM 10 (air)	2.97	2.52	–15.2%	1.56	–38.1%
Sulphur dioxide (air)	2.48	1.73	–30.2%	1.55	–10.4%
VOCs (air)	3.60	3.45	–4.2%	1.94	–43.8%
BOD+COD (water)	3.93	2.62	–33.3%	2.12	–19.1%
Heavy metals (water)	3.93	1.82	–53.7%	2.38	30.8%
Nitrogen as N (water)	3.04	3.86	27.0%	1.52	–60.6%
Oils and greases (water)	3.96	1.72	–56.6%	2.46	43.0%
Phenols (water)	3.99	2.87	–28.1%	2.51	–12.5%
Phosphorus as P (water)	3.92	3.55	–9.4%	2.41	–32.1%
Suspended solids	3.99	3.33	–16.5%	2.79	–16.2%
Oil in ground (Resource)	2.43	4.00	64.6%	1.58	–60.5%
P in ground (Resource)	3.99	2.80	–29.8%	2.81	0.4%
Soil erosion (Resource)	3.87	3.95	2.1%	2.40	–39.2%
Average	3.37	2.83	–16.0%	2.01	–29.1%

Figure 15 gives an overview of the development of the dispersion coefficients in the 'total of NO_x' base. In the second round, the first iteration round, the average dispersion of the index values is reduced by about 40%. The converging tendency does not continue in the third round though. The average dispersion coefficient is practically the same in the third round as it is in the second round. For some interventions the consensus seems to be slightly increasing and for some others decreasing. The eventual uncertainty of the indexes is high, the 95% confidence interval ranging from zero (for NO_x emissions, because they are the base) to ±8.5 times the average of the distribution. The average of the 95% confidence interval is ±4.36 times the average of the distribution. The conclusion is that the indexes based on the 'total of NO_x' postulate are less dispersed than those calculated on the 'first-ranked' postulate, but, nevertheless, they are very uncertain. The dispersion of the indexes reduced significantly in the first iteration round, but in the second iteration round, the search for consensus among the experts was not too successful. Perhaps additional iteration rounds would have helped to reduce the dispersion of the indexes. These were not carried out in the Delphi I study though, and the opinions of the experts remained quite far from each other.

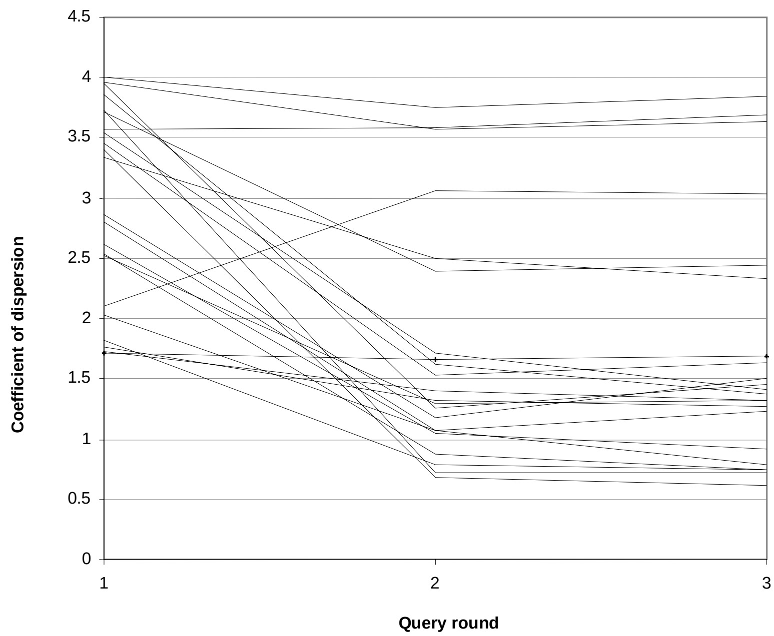

Figure 15. Development of the index dispersions in the Delphi process.

The reduction in the dispersion of the index distributions between first and second rounds was obviously partly due to improved task instructions, which reduced the possibility of misunderstanding of the rating method. Unlike the instructions of the first round, which did not make it quite clear which figures were looked for, the instructions of the second round were unambiguous and clear. The judges were asked to consider the relative effect of a 1% reduction in each intervention, compared with a 1% reduction in nitrogen oxides, which was given an index of 1 000 (Wilson and Jones, 1996, p. B1.10). This should have left no doubt about the figures meant. It is impossible to estimate, however, how much the improved task instructions really affected the index distributions in the second round and what was the actual effect of the Delphi technique.

8.2.3 Standardisation of the final indexes

Because of a wide dispersion of the index values in the 'total of NO_x' base, the final indexes were computed in the Delphi I study as 'standardised'. Standardisation meant in practice that the base of the indexes was changed from the 'total of NO_x' to the total environmental harm of all interventions. This was done by giving the total harm a fixed value of 1 000 and calculating the new indexes for each intervention in this basis as

$$H_i^{\text{Standard}} = 1000 \frac{H_i}{\sum_{j=1}^{23} H_i} . \tag{22}$$

We cannot show that the postulate of the standardised indexes, i.e. that the aggregate harm of all interventions is the same for each judge, would be less correct than any of the other base postulates studied above. It can be justified from the expertise point of view, because one can expect that experts have a scientific and up-to-date comprehension on the magnitude of the environmental problems, which would be the best guarantee for an objective assessment. In an ideal case, where the scientific foundation and the knowledge of the experts would be similar, the conceptions could be close to each other. But the conceptions of each expert cannot be proven to be similar. Thus we cannot show that this postulate would be more correct than any of the other postulates.

To explore the logic of the base transformation we start from the initial form of rating, which defined the benefits of 1% cuts in the interventions on a relative scale so that

$$-\Delta_{1\%} H_{NO_x} = h_{NO_x} = 1000 \tag{23}$$

was the benefit of a 1% cut in annual emissions of nitrogen oxides on that scale, and

$$-\Delta_{1\%} H_i = h_i \tag{24}$$

was the benefit of a 1% cut in any of the other interventions on the same scale. Now, because

$$\frac{-\Delta_{1\%} H_i}{-\Delta_{1\%} H_{NO_x}} = \frac{1\% H_i}{1\% H_{NO_x}} = \frac{H_i}{H_{NO_x}} = \frac{h_i}{h_{NO_x}} , \tag{25}$$

we can express the harm of an individual intervention as

$$H_i = \frac{H_{NO_x}}{h_{NO_x}} h_i , \tag{26}$$

and, thus, for the total harm of all interventions we can write

$$\sum_{i=1}^{23} H_i = \frac{H_{NO_x}}{h_{NO_x}} \sum_{i=1}^{23} h_i \ . \qquad (27)$$

Accordingly, the share of an individual intervention in the total harm is

$$\frac{H_i}{\sum_{i=1}^{23} H_i} = \frac{\frac{H_{NO_x}}{h_{NO_x}} h_i}{\frac{H_{NO_x}}{h_{NO_x}} \sum_{i=1}^{23} h_i} = \frac{h_i}{\sum_{i=1}^{23} h_i} \qquad (28)$$

If we then assume that all judges think exactly the same about the total environmental harm, we can give it a fixed value, for instance 1000, as in the Delphi I study, and, based on that, express the harm of each individual intervention as

$$H_i = \frac{h_i}{\sum_{i=1}^{23} h_i} (\sum_{i=1}^{23} H_i) \xrightarrow{\sum_{i=1}^{23} H_i = 1000} \frac{h_i}{\sum_{i=1}^{23} h_i} \cdot 1000 \qquad (29)$$

However, as mentioned above in the context of the 'total of the first-ranked base', it is very likely that there are differences in the conceptions of the judges about the magnitude of the total environmental harm. Should this be the case, then neglecting these differences will bias the indexes to some extent. The magnitude of this bias is impossible to estimate though, since no attempt was made in the Delphi I study to determine the possible differences in the conceptions of the judges about the magnitude of the total environmental harm.

Nevertheless, the transformation of the basis was made and the final indexes were calculated in the Delphi I study on the assumption that each judge comprehends the magnitude of the total environmental harm similarly. As a result of the base transformation, the dispersion of indexes was reduced, as Table 9 shows. A reduction of over 40% occurred in greenhouse gas emissions, carbon monoxide, VOCs, sulphur dioxide emissions, water releases of oils and greases, and oil as a resource. Dispersion of the indexes of nitrogen and phosphorus emissions to water, suspended solids, as well as air emissions of particulates, pesticides and PM_{10} were reduced by 20 to 40 %. For soil erosion and water emissions of phenols and heavy metals, the dispersion of indexes declined by less than 20%. The indexes of phosphorus resource, air emissions of benzene, ammonia and heavy metals, and BOD+COD water emissions became more dispersed, the latter by over 50%.

Table 9. Comparison of the coefficients of dispersion (ratio of the standard deviation to the average) of the harmfulness indexes in the third round between the 'total of NO_x' base and the 'total of all interventions' base (basis of the final indexes in the Delphi I study).

Intervention	**Total of NO_x**	**Total of all**	**Δ from total of NO_x**
Ammonia (air)	0.61	0.74	21.3%
Benzene (air)	0.79	0.90	13.9%
Carbon dioxide (air)	2.44	1.45	–40.6%
Carbon monoxide (air)	3.69	1.19	–67.8%
Heavy metals (air)	0.72	0.89	23.6%
Methane (air)	3.64	1.20	–67.0%
Nitrogen oxides (air)	0.00	0.65	
Nitrous oxide (air)	3.04	0.76	–75.0%
Particulates (air)	1.27	0.95	–25.2%
Pesticides (air)	0.74	0.58	–21.6%
PM 10 (air)	1.32	0.90	–31.8%
Sulphur dioxide (air)	1.32	0.77	–41.7%
VOCs (air)	1.63	0.69	–57.7%
BOD+COD (water)	0.91	1.37	50.5%
Heavy metals (water)	0.75	0.72	–4.0%
Nitrogen as N (water)	1.38	0.86	-37.7%
Oils and greases (water)	1.69	0.99	–41.4%
Phenols (water)	1.51	1.35	–10.6%
Phosphorus as P (water)	1.41	0.95	–32.6%
Suspended solids	2.33	1.64	–29.6%
Oil in ground (Resource)	3.85	1.78	–53.8%
P in ground (Resource)	1.23	1.26	2.4%
Soil erosion (Resource)	1.45	1.28	–11.7%
Minimum	0.00	0.58	
Maximum	3.85	1.78	–53.8%
Average	1.64	1.05	–36.0%

Typically, those indexes exhibiting reduced dispersion were widely dispersed and those indexes for which dispersion was increased were narrowly dispersed in the ’total of NO_x’ base. This is clear from Figure 16, which shows the dependency between the maximum-to-minimum ratio of the indexes in the ’total of NO_x’ base, and the reduction of dispersion through standardisation. One reason for this was that the total sum of the ratings was considerably higher for those judges who ranked NO_x emissions further from the top than that for those judges who ranked NOx emissions close to the top. The former often represented high values in the widely dispersed indexes, whereas the latter represented low values. Thus, high index values were reduced more in the standardisation than were low values, which reduced the dispersion. For those indexes exhibiting increased dispersion, the situation was the opposite. They were narrowly dispersed in the ’total of NO_x’ base and judges with a high total rating value often represented low values of the distribution. High index values, in turn, were represented by judges with a low total rating value. Thus, high index values were reduced less in the standardisation than were low values, which resulted in the increase of the dispersion.

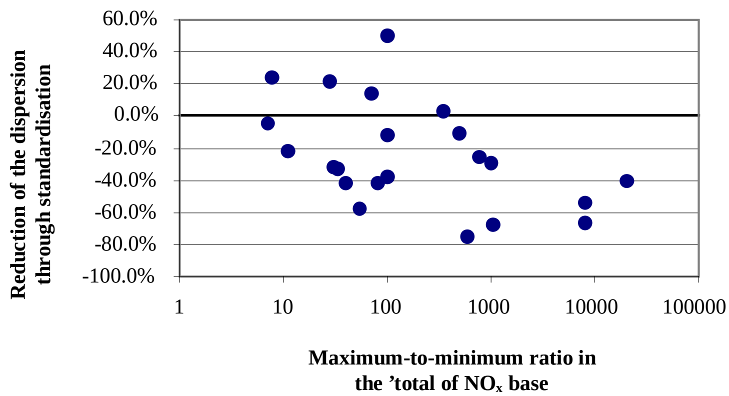

Figure 16. Dependency between the maximum-to-minimum ratio of the indexes in the ’total of NO_x’ base, and the reduction of the index dispersion through standardisation in the Delphi I study.

Examples of these cases are given in Figure 17. There, carbon dioxide emissions is the case where the dispersion of indexes was greatly reduced, and BOD+COD is the opposite case where the dispersion was greatly increased. In the figures, the relative total rating is the total rating of a judge divided by the average of the total ratings among the judges. The relative index is the index given by a judge divided by the average index. Both quantities are calculated for the ’total of NO_x’ base. The index dispersion of carbon dioxide emissions is greatly reduced when the two high index values are divided by the respective high rating totals and the rest of the indexes by their related low rating totals. Dispersion of the indexes of BOD+COD emissions is increased in the standardisation when three high index values are divided with their related low rating totals and the low index values with their slightly higher rating totals.

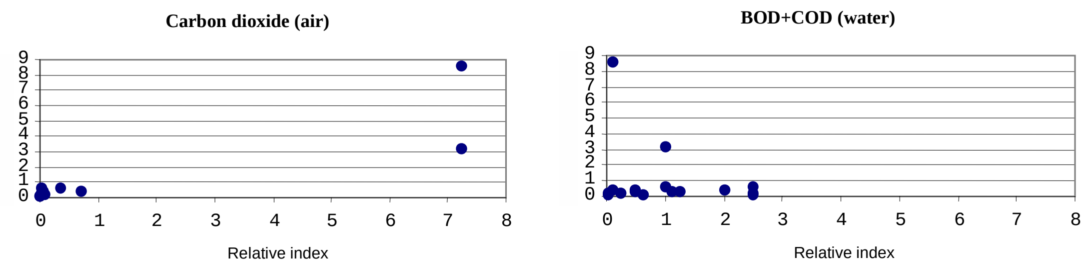

Figure 17. Examples of the statistics of the indexes and the rating totals in the 'total of NO_x' base in the Delphi I study. Values of both variables are relative to the corresponding means among the judges. The index dispersion of carbon dioxide emissions is greatly reduced and that of BOD+COD emissions increased in the standardisation (division of indexes by the rating totals).

A conclusion from the analysis of the statistics of the indexes and the rating totals in the 'total of NO_x' base is that the dispersion of indexes has been reduced by standardisation. The reduction in the average dispersion has been about 36%. It remains unclear, however, how much the convergence depended on the possibly more correct index basis employed in the standardisation and how much on the mechanical effects of standardisation procedure. The contributions of these two factors cannot be estimated, because the opinions of the judges about the dimensions of the environmental problems in the Nordic countries were not sought in the Delphi I study. In any case, standardisation represented a rather strong interference on the part of the conductors because it took place so late that it was no longer possible to involve the judges in the assessment of this action. Neither the justification nor the outcome of the standardisation were discussed with the experts involved in the study. Therefore, to some extent the final results reflect subjective choices of the conductors, which should be known when using the indexes in the LCA valuation.

How large the effect of the conductors' subjectivity might be as a whole is again difficult to assess. Nevertheless, a rough picture of it can be formed by comparing the final indexes to those of the 'total of NO_x' base. The latter were produced by the judges who obviously knew exactly the meaning of the indexes. They realised that they were estimating the proportions of harms using the total harm of NO_x emissions as a reference. Thus, we may assume that the influence of the conductors in those indexes was minimal in the framework of the method. The standardisation was carried out without influence from the judges. Thus, we may assume that the influence of the conductors in those indexes was maximal. If the judges had been asked to assess the proportions of the interventions using the total environmental harm as a reference, the results would have been obviously different.

Table 10. Comparison of the indexes (1/1 000 t, soil erosion 1/ha) of the 'total of NO_x'-base (after the third round) and the final indexes of the 'total of all interventions' base (final indexes in the Delphi I study). The latter have been reduced so that the index of the nitrogen oxide emissions is the same in both bases.

Intervention	**Total of NO_x**	**Total of all**	**Δ from total of NO_x**
Ammonia (air)	11.41	11.12	–2.6%
Benzene (air)	1.31	1.04	–20.8%
Carbon dioxide (air)	0.19	0.02	–87.9%
Carbon monoxide (air)	4.43	0.42	–90.6%
Heavy metals (air)	734.57	676.75	–7.9%
Methane (air)	10.92	1.00	–90.8%
Nitrogen oxides (air)	3.95	3.95	0.0%
Nitrous oxide (air)	52.65	12.96	–75.4%
Particulates (air)	5.15	3.26	–36.6%
Pesticides (air)	123.75	107.04	–13.5%
PM 10 (air)	129.05	82.13	–36.4%
Sulphur dioxide (air)	8.75	5.32	–39.2%
VOCs (air)	4.01	2.34	–41.8%
BOD+COD (water)	1.09	1.10	1.1%
Heavy metals (water)	2 326.22	1 985.97	–14.6%
Nitrogen as N (water)	42.41	31.12	–26.6%
Oils and greases (water)	4 453.04	3 310.60	–25.7%
Phenols (water)	42 100.00	39 249.67	–6.8%
Phosphorus as P (water)	955.18	685.78	–28.2%
Suspended solids	3.28	2.01	–38.7%
Oil in ground (Resource)	0.52	0.03	–93.3%
P in ground (Resource)	982.04	1 056.23	7.6%
Soil erosion (Resource)	16.29	12.67	–22.2%

If we reduce the standardised final indexes back to the NO_x reference – by multiplying each index by the ratio of the index of NOx in the 'total of NO_x' base – to that in the standardised basis, we obtain the comparison given in Table 10. This comparison, when used to assess the influence of the conductors, relies on the assumption that, had the judges really made the rating in the 'total of all' base and had they stuck to their opinions, the indexes of the interventions should have the same proportions as in the 'total of NO_x' base. Another necessary assumption is that the differences in the comprehension of the experts about the magnitude of the total environmental harm could have been correctly taken into account when making the index statistics. Both assumptions would, of course, materialise with a very low probability. Thus the comparison, if taken as a descriptor of the conductors' influence, represents a very rare extreme case.

A graphical comparison of the indexes is shown in Figure 18. As can be observed, standardisation has changed the mutual weights of the interventions dramatically. The indexes of oil in ground (resource), methane (air), carbon monoxide (air), carbon dioxide (air) and nitrous oxide (air) have been reduced by over 75%, the indexes of VOCs (air), sulphur dioxide (air), suspended solids, particulates (air) and PM_{10} (air) by 30 to 40 percent, the indexes of phosphorus as P (water), nitrogen as N (water), oils and greases (water), soil erosion (resource), benzene (air), heavy metals (water) and pesticides (air) by 10 to 30 per cent and the indexes of heavy metals (air), phenols (water) and ammonia (air) by less than 10%. The indexes of BOD+COD (water) and P in ground (resource) have been slightly increased. Consequently, we can conclude, under the reservations discussed above, that the influence of the conductors' decision to change the index basis for the final results has been very significant. The index profile differs essentially from the one produced by the experts with nitrogen oxide emissions as the reference.

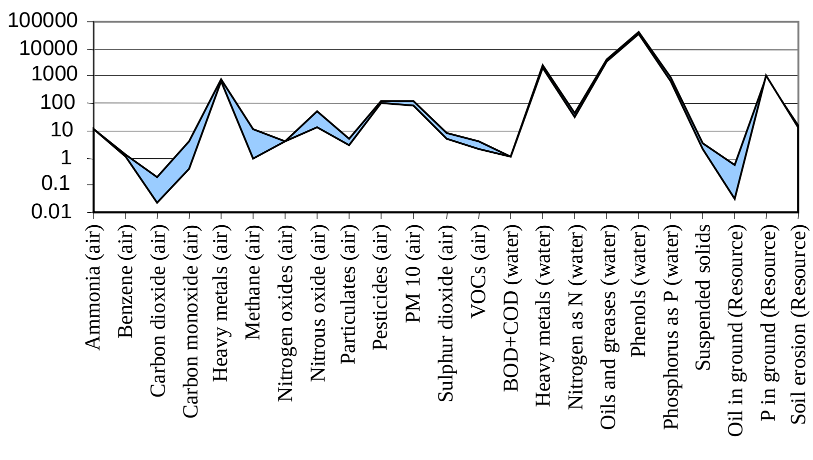

Figure 18. Differences (dark shade) between the indexes of the 'total of NO_x' base (after the third round) and the final indexes of the 'total of all interventions' base (final indexes in the Delphi I study, light shade). The latter have been reduced so that the index of the nitrogen oxide emissions is the same in both bases. Note the logarithmic scale.

8.3 Problem of uncertainty from an entropy point of view

8.3.1 Entropy as a measure of uncertainty

The Kolmogorov entropy (denoted by *K*) was initially developed by the Russian mathematician A. Kolmogorov to measure the chaotic tendency of a system. Before introducing this quantity, however, it is useful to recall that the thermodynamic entropy *S* measures the disorder in a given system. A traditional example, for a system where *S* increases, is that of gas molecules that are initially confined to one half of a box, but are then suddenly allowed to fill the whole container. The disorder in this system increases because the molecules are no longer separated from the other half of the box and thus, have new possibilities to fill the container. This increase of disorder is coupled with an increase of our ignorance about the state of the system. Initially there were fewer places (and states of energy) that we needed to consider for the particles to be in (or have) than after the confinement was lifted. In other words, we used to know more about the positions and behaviour of the molecules.

According to Shannon (1948), entropy *S*, which can be expressed as

$$S = -\sum_i p_i \ln p_i \qquad (30)$$

where $\{p_i\}$ are the probabilities of finding the system in different states $\{i\}$, measures the information needed to locate the system in a certain state, i.e. *S* is a measure of ignorance.

8.3.2 Information capacity of a store

Figure 19a shows a system with two possible states. If the position of the points is unknown, a priori, and we learn that it is in the left box, we gain by definition information amounting to one bit. If we obtain this information, we save one question (with a possible answer, yes or no, which we would have needed to locate the point). Thus, the maximum information content of a system with two states is one bit. For a box with four possible states (Figure 19b), one needs two questions to locate the point, i. e. its maximum information content *I* is two bits. This can be written as the logarithm to the base two (ld) of the number of possible states :

$$I = ld\ 4 \qquad (31)$$

According to Figure 19c, this logarithmic relation between the maximum information content *I* and the number of states *N* is true in general, i.e.

$$I = ld\ N \qquad (32)$$

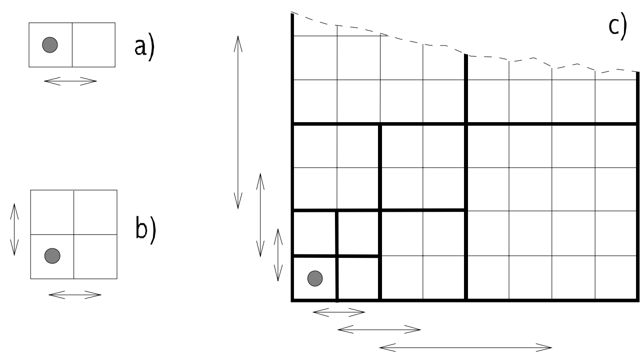

Figure 19. Information capacity of a store a) a box with two states. b) It takes two questions (and their answers) to locate a point in a system with four states: right or left? up or down ? c) In order to locate a point on a checkerboard with 64 =2^6 states, one needs six questions.

8.3.3 Information gain

Let us now calculate the average gain of information if one learns the outcome of statistical events. Suppose we toss a coin such that heads or tails occur with equal probabilities

$$p_1 = p_2 = 1/2 \tag{33}$$

The information I acquired by learning that the outcome of this experiment is, for instance, heads, amounts to 1 because there are two equally probable states, as in Figure 19a. This result can be expressed via the $\{p_i\}$ as

$$I = -\left(\frac{1}{2}\mathrm{ld}\frac{1}{2} + \frac{1}{2}\mathrm{ld}\frac{1}{2}\right) \text{or} \tag{34}$$

$$I = -\sum_i p_i \mathrm{ld} p_i \tag{35}$$

The latter equation can be generalised to situations where the p_i ‘s are different:

$$p_1 \neq p_2 = 1 - p_1 \tag{36}$$

It then gives the average gain of information if we toss a deformed coin many times. Let $p_1 = r/q$, where r and q are integers, and let us choose the number m of events such that mr/q is again an integer. The total number of distinct states which occur if one tosses a deformed coin m times is

$$N = \frac{m!}{(p_1 m)!(p_2 m)!} \tag{37}$$

where the permutations that correspond to a rearrangement of equal events have been eliminated by division (the sequences hht and hth with h = head and t = tail, where the h's have been interchanged, correspond to the same state). When m $\rightarrow \infty$, which corresponds to learning to know the precise values of p_1 and p_2, Stirling's formula (where e is the base of the natural logarithm) can be applied, and thus the yield for the average information gain (see also Figure 20.)

$$\begin{aligned} I &= \frac{1}{m}\mathrm{ld}N = \frac{1}{m}\mathrm{ld}\left[\left(\frac{m}{e}\right)^m \left(\frac{e}{p_1 m}\right)^{p_1 m} \left(\frac{e}{p_2 m}\right)^{p_2 m}\right] \\ &= -(p_1 \mathrm{ld} p_1 + p_2 \mathrm{ld} p_2) \end{aligned} \tag{38}$$

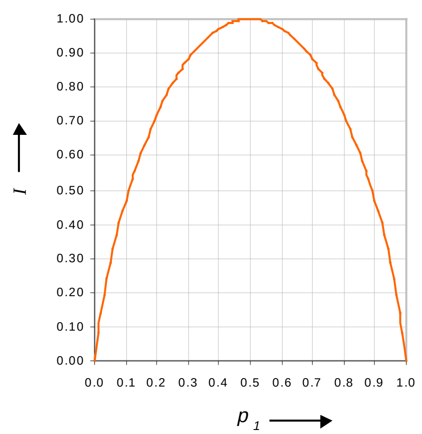

Figure 20. ***I***(p) *for an experiment with two possible outcomes. If* p_1 *= 0, it is sure that the outcome will be event 2, and no information is gained. The maximum information is acquired for* $p_1 = p_2$ *= 1/2, where the uncertainty of the outcome has its maximum and one learns most from the experiment.*

The generalisation of this is Shannon's result:

> If we, a priori, know only that 1...n states of a system occur with probabilities $\{p_i\}$, such that $\Sigma\, p_i = 1$, and we learn by measurement that the system actually occupied a certain state, then, if we repeat this measurement many times, we gain the average information

$$I = -\sum_i p_i \mathrm{ld} p_i \tag{39}$$

8.3.4 *K*-entropy

The above example from statistical mechanics shows that disorder is essentially a concept of information theory. It is therefore not surprising that the *K*-entropy can be defined by Shannon's formula in such a way that it becomes proportional to the information needed to predict the system in a particular future state with a certain precision, assuming that one knows the history of the system and the possible states that the system can enter. According to Schuster (1989), the quantity

$$K_n = -\sum_{i_0 \ldots i_n} p_{i_0 \ldots i_n} \ln p_{i_0 \ldots i_n} \tag{40}$$

is proportional to the information needed to locate the system in a special trajectory of states $i_0^* \ldots i_n^*$ with a certain precision (see Figure 21). Therefore, $K_{n+1} - K_n$ is the additional information needed to predict in which cell of the phase space the system will be if we know the cells where it used to be previously.

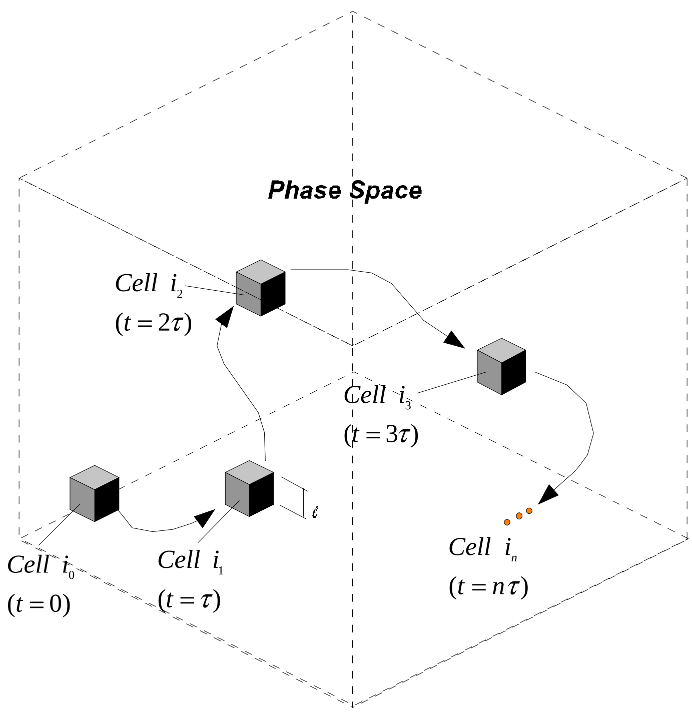

Figure 21. A state trajectory of a system in a phase space. The size of each cell is l^d (d is the number of dimensions in the space). $p_{i_0 \ldots i_n}$ is the joint probability of the trajectory, i.e. that system's state is inside i_0 at t = 0, in i_1 at t = τ, ... and eventually in i_n at t = $n\tau$.

The K-entropy is defined as the average rate of information loss:

$$K = \lim_{\tau\to 0}\lim_{l\to 0}\lim_{N\to\infty}\frac{1}{N\tau}\sum_{n=0}^{N-1}(K_{n+1}-K_n) \tag{41}$$

When this is applied to a case, such as predicting the behaviour of the ecological system under various interventions, with one discrete arbitrary time step only ($N = 1$, $\tau = 1$) and an arbitrary precision ($l > 0$), then

$$\begin{aligned}
K = K_1 - K_0 &= -\sum_{i_0\ldots i_1} p_{i_0\ldots i_1}\ln p_{i_0\ldots i_1} - (-\sum_{i_0} p_{i_0}\ln p_{i_0}) \\
&= -\sum_{i_0}\sum_{i_1} p_{i_0}p_{i_1}\ln p_{i_0}p_{i_1} + \sum_{i_0} p_{i_0}\ln p_{i_0} \\
&= -\sum_{i_0}\sum_{i_1} p_{i_0}p_{i_1}\ln p_{i_0} - \sum_{i_0}\sum_{i_1} p_{i_0}p_{i_1}\ln p_{i_1} + \sum_{i_0} p_{i_0}\ln p_{i_0} \\
&= -\sum_{i_0}\overbrace{(\sum_{i_1} p_{i_1})}^{=1} p_{i_0}\ln p_{i_0} - \sum_{i_1}\overbrace{(\sum p_{i_0})}^{=1} p_{i_1}\ln p_{i_1} + \sum_{i_0} p_{i_0}\ln p_{i_0} \\
&= -\sum_{i_1} p_{i_1}\ln p_{i_1}
\end{aligned} \tag{42}$$

Thus, K-entropy is the measure of information needed (a linear transformation of Shannon's information) to predict the future state of the system when the initial state is known. If there is no information on the system (complete ignorance), all possible states are predicted to be equally probable. Consequently, K-entropy reaches its maximum. If the system behaviour is precisely known (complete awareness, theoretical), only one state is predicted. K-entropy is zero because no additional information is necessary.

8.3.5 *K*-entropy of ranking the interventions

In the following we study the development of the consensus of the judges by means of entropy analysis of the rankings in order to see the extent to which the Delphi process affects the consensus, or in other words, changes the opinions of the individual judges. Another purpose is to assess the final degree of agreement to get a picture of the epistemological certainty of the valuation indexes produced in the case study. Entropy analysis is based on application of K-entropy to the ranking of interventions. This is done by dividing the ranks into five classes, as indicated in Table 11, and by calculating the distribution of the judges' votes into those rank groups for each intervention. Maximum entropy corresponds to complete disagreement about the rank of an intervention. Then, all rank groups are predicted to be equally probable for an intervention and the distribution of the judges' votes becomes maximally even. Minimum entropy follows from complete agreement. In that case, all the judges consider an intervention to belong to the same rank group, and the *K*-entropy becomes zero.

Table 11. Ranges of the rank groups for different resolutions. Five rank groups were used as a basis for the K-entropy analysis of consensus.

Number of rank groups	Rank groups
5	I=1...5, II=6...10, III=11...15,IV=16...20,V=21...23
8	I=1...3, II=4...6, III=7...9,IV=10...12,V=13...15,VI=16...18,VII=19...21,VIII=22...23

K-entropy does not necessarily reveal anything about the level of knowledge of the judges. It only measures the degree of agreement among them. It is quite possible, for instance, that there is a judge who has excellent knowledge on the effects of several interventions. Based on this knowledge this judge ranks a certain intervention into a certain rank group. At the same time, other judges may be less familiar with this intervention than other interventions, and thus rank that particular intervention differently, causing disagreement, which is then seen as high entropy. The Delphi method, however, relies on the average votes and the development of consensus among the judges, and does not try to assess their knowledge, which would, no doubt, be an impossible task. Thus *K*-entropy, which indicates the diversity of the votes, is a relevant indicator of the collective certainty of the method.

The *K*-entropy limits given in Table 12. are used to classify the degree of consensus. We say that the degree of consensus about the rank of an intervention is in a certain class if the *K*-entropy is lower than or equal to the limit given for that class and lower than the limit of the next higher class. Each limit *K*-entropy corresponds to a set of limit distributions of the judges votes so that each of these distributions give the same *K*-entropy. Hence, there is more than one limit distribution for each consensus class. For instance, the *K*-entropy limit of the highest consensus class, "very strong", can be reached with the required 14 votes in any of the five rank groups and the remaining 2 votes in any two of the other four rank groups. Therefore, the distributions given in Table 12 are examples of the limit distributions only.

The rankings are evaluated in groups in order to identify any major consensus trends that may exist in the expert judgements. Because of the complexity of the impact assessment and the consequent uncertainty of judgements, it may be assumed that the ranks of a certain intervention could fall inside a reasonable ranking range rather than into one particular rank, even if the judges were quite unanimous about their importance. Absolute values of *K*-entropies vary according to the ranges of groups. The theoretical extremes are, in this case, 2.8 for 23 groups, each containing one single rank, and 0 for one group covering all ranks. Neither of these is, however, meaningful from the point of view of detecting diversity of the judgements. In order to demonstrate how the group resolution affects the *K*-entropies, three resolutions are compared in Figure 22, one with five ranks, one with three ranks, and one with one rank per group, corresponding to five, eight, and 23 rank groups, respectively. The comparison is made for relative *K*-entropies, that is for ratios of the absolute *K*-entropy to the maximum *K*-entropy, the maximum entropies being 1.6 for the five group, 2.1 for the eight group, and 2.8 for the 23 group resolution. As can be observed from Figure 22, increasing group resolution dissolves the relative entropy variation. This is a logical effect of increasing demand for precision.

Table 12. K-entropy limits for the different degrees of consensus (five rank groups). The distributions given are examples of the limit distributions.

Distribution of judges votes in rank groups					***K*-entropy**		**Total**	**Agree**	**Disagree**	**Consensus**
I	**II**	**III**	**IV**	**V**	**Abs**	**Rel**				
	1	*14*	*1*		*0.46*	*0.29*	16	14	2	Very strong
1	*1*	*12*	*1*	*1*	*0.91*	*0.57*	16	12	4	Strong
1	2	10	2	1	*1.16*	*0.72*	16	10	6	Rather strong
2	2	8	2	2	*1.39*	*0.87*	16	8	8	Rather weak
2	3	6	3	2	*1.52*	*0.95*	16	6	10	Weak
3	3	4	3	3	*1.60*	*1.00*	16	4	12	Very weak

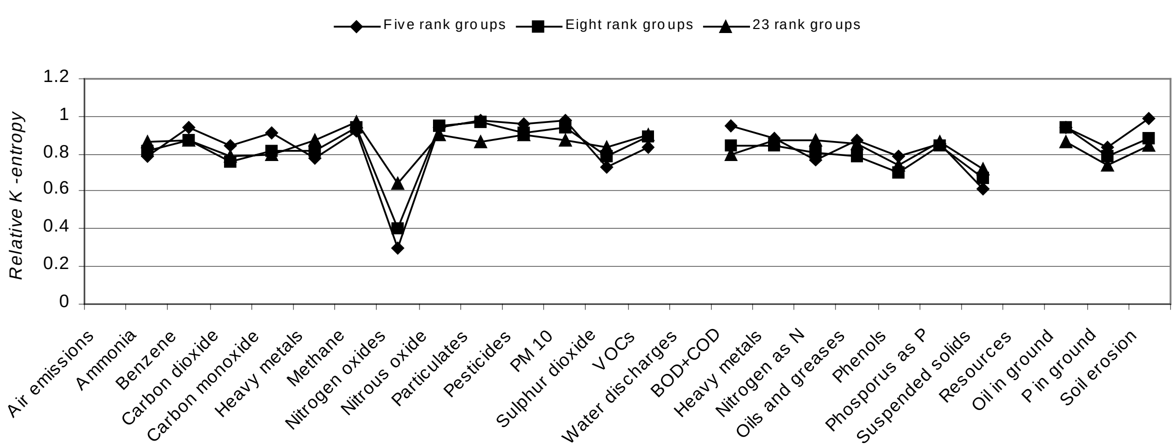

Figure 22. Comparison of relative K-entropies for different resolutions of five, eight and 23 rank groups.

However, there seems to be one intervention which stands out in both resolutions, i.e. nitrogen oxide. Its relative entropy is less than 0.5 (1.0 corresponds to complete disagreement) for the five and eight groups and less than 0.7 even for the 23 group resolution. This means that the judges have been fairly unanimous about its importance. For the rest of the interventions, the entropies range relatively from 0.6 to 1.0. It appears that traditional interventions, e.g. sulphur dioxide, heavy metals (to air), nitrogen (to water) and suspended solids have slightly lower entropies than the "newcomer" interventions. This may reflect the establishment of knowledge.

The actual *K*-entropy analysis is made using the five rank group resolution. Its detailed results are given in Appendix D.

Figure 23 shows a summary of the analysis of relative entropies. It appears that the entropies have generally declined between the first and second rounds. From the second to third rounds the trend in *K*-entropy is also downward, but not so clearly as from the first to the second rounds. This suggests that the tolerance of opinion changes was already used up in the first iteration round and, thus, the second iteration could not essentially increase the consensus, because the judges stuck to their previous opinions.

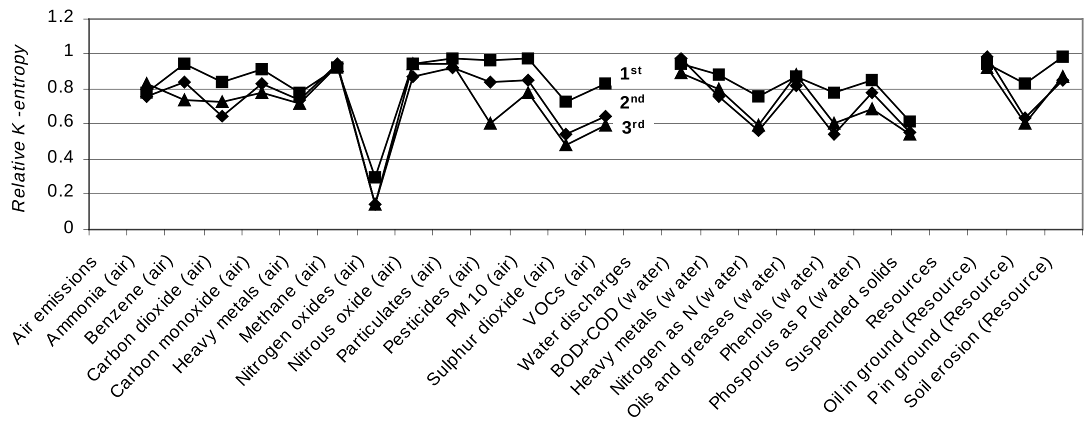

Figure 23. Relative K-entropies for ranking of interventions in the three rounds of the Delphi I study, five rank groups.

Figure 24 shows the relative entropy changes between the rounds in the order of the change between the first and the second round. The largest change in the first iteration (the second round) has occurred in nitrogen oxide emissions, over 50%. This finding is not surprising if we consider that in the iteration rounds NO_x emissions were used as a reference, to which all the other interventions were to be compared. Moreover, they were listed at the top of the recording grid according to the order of the average rankings from the initial round. They were thus pushed in a way to the top of the ranking in the first iteration. A more surprising finding might be that out of the top ten interventions in the entropy decline order, six interventions appeared in the top ten of the intervention list provided in the recording grid for the first iteration round. This list was in the order of the rankings from the first round, and these ranks were also accompanying the interventions on the list. These six interventions were Nitrogen oxides (air) (1. in entropy decline, 1. in the recording grid), Nitrogen as N (water) (3. in entropy decline, 3. in the recording grid), Sulphur dioxide (air) (4. in entropy decline, 2. in the recording grid), Carbon dioxide (air) (6. in entropy decline, 5. in the recording grid), VOCs (air) (7. in entropy decline, 6. in the recording grid), and PM 10 (air) (10. in entropy decline, 8. in the recording grid). This indicates that the method used for consensus search in the case study worked quite efficiently in the first iteration round for the top ten interventions from the first round. Also the average reduction of *K*-entropy tells the same story. For the first ten interventions in the recording grid of the first iteration the reduction of *K*-entropy was 14%. For the rest of the interventions, i.e. from the eleventh in the recording grid, the corresponding reduction was 10%, and for all interventions 11%. The fact that the convergence of the top interventions was about 40% stronger than that of the rest of the interventions suggests that both the feedback from the initial ranking and the way (rank order) it was communicated to the judges affected their decisions in the first iteration.

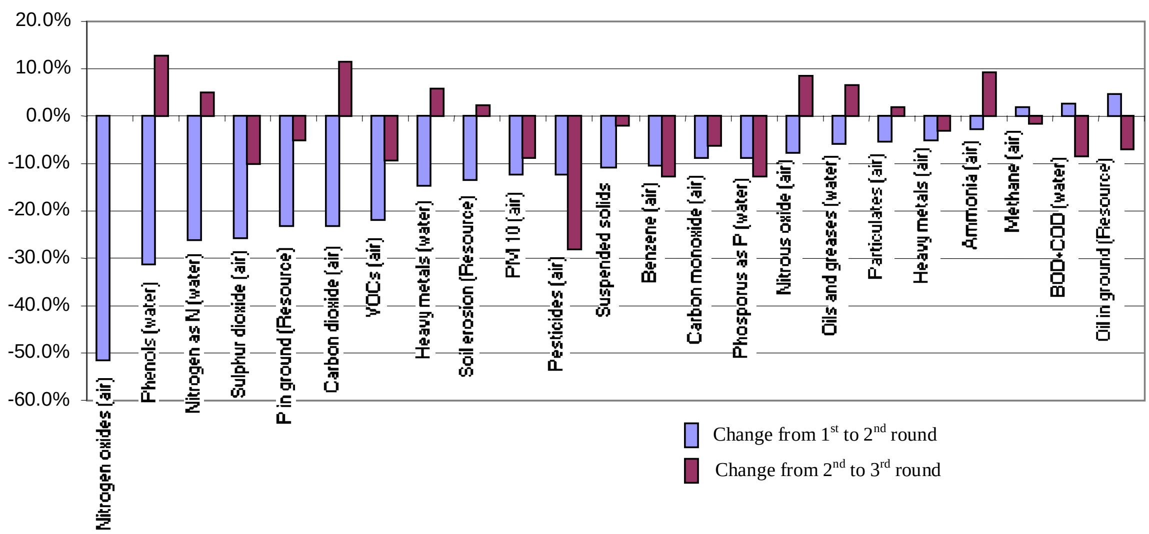

Figure 24. Relative changes of K-entropies for ranking of interventions from the first to the second round and from the second to the third round of the Delphi I study, five rank groups.

However, if Figure 24 is redrawn (Figure 25) to show the interventions in their rank order, rather than in order of their entropy change from the first and second rounds, a different interpretation emerges. The distribution is distinctly (inverted) U-shaped.

Although entropy changes are above average for the first few highest ranked interventions, as recognised above, so too are the lowest ranked ones, with the smallest changes in the middle of the range.

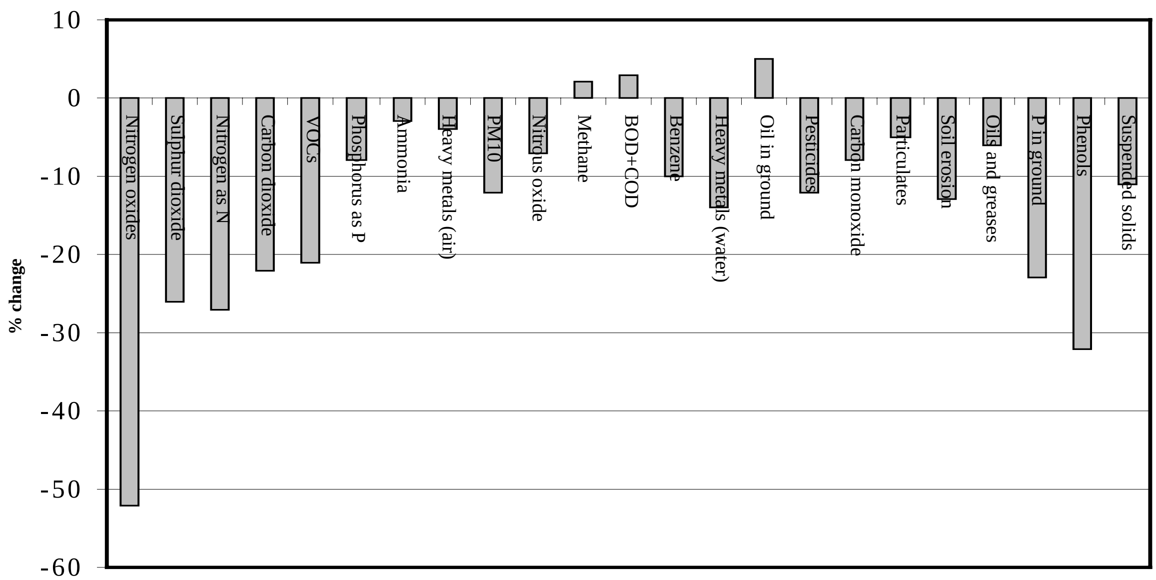

Figure 25. Relative changes of K-entropies for ranking of interventions from the first to the second round of the Delphi I study, five rank groups. (Redrawn Figure 24. 1st to 2nd round only.)

For higher-ranking interventions the minorities moved closer to the majorities, as they did for the lower-ranked interventions. In both cases both panels showed the same movements. For the middle-ranking interventions the panels tended to move in opposite directions, which led to the lack of entropy change observed above.

In the second iteration of the Delphi I study, consensus developed quite differently from the first iteration. The overall *K*-entropy declined by 3.0% and the first ten interventions in the recording grid of the second iteration by 1.5%, whereas the interventions further in the grid were reduced by over 4%. These figures indicate that the third round was not very effective in increasing consensus and that the feedback from the first iteration did not have a similar effect as in the second round. One explanation for this finding may be that the ranking of interventions fed back to judges was rather similar to the feedback of the first iteration, during which many judges who saw it necessary to change their rankings had already changed them. Among the top ten, eight of the interventions were the same as in the first iteration. Thus there was less reason to change the rankings anymore in the second iteration. On the other hand, judges who felt that there was no reason to change the rankings maintained that view in the second iteration as well.

Figure 26 shows the development of relative *K*-entropy in the main intervention categories, namely air emissions, water discharges and resources. The differences between the categories are small both in entropies and in their changes from round to round. The resource category has a slightly higher entropy in all rounds than do air and water emissions, the latter two are practically on the same entropy level throughout the process. This result is in line with the present situation in LCA valuation, where valuation of resources, especially fossil energy resources, is a disputed issue and common agreement on suitable valuation methods seems to be quite far away.

As a whole, *K*-entropy is reduced by about 17% in the whole process, which is not much considering that the initial level is quite high, relative *K*-entropy 0.843. The final average *K*-entropy, 0.72, corresponds to "rather strong" in our consensus classification. The average entropy does not, however, give a quite correct picture of the state of consensus. A more realistic picture can be seen in Table 13, which lists the consensus classes for different interventions. It shows that for ten interventions the consensus is rather strong or stronger, and for thirteen interventions rather weak or weaker. For seven interventions the consensus is very weak. The interventions about which the judges were most unanimous are nitrogen oxide emissions (very strong), and sulphur dioxide emissions and suspended solids (both strong). The emissions for which hardly any consensus was found (very weak) are methane, nitrous oxide, and particulate emissions to air, BOD+COD, oils and greases released to water, oil in ground and soil erosion.

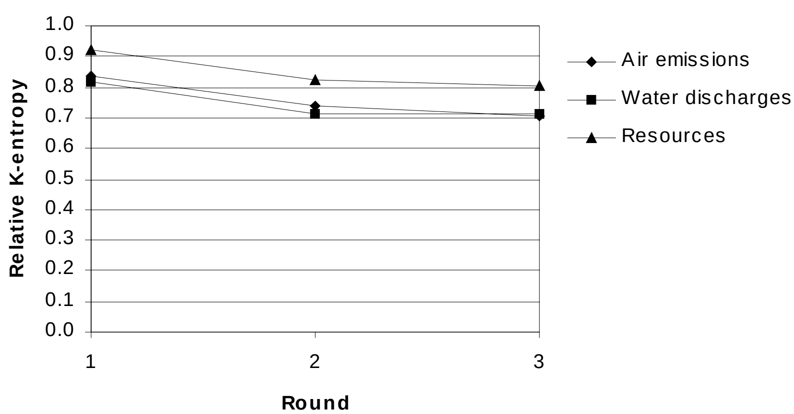

Figure 26. Development of average relative K-entropies for ranking of interventions in the main intervention categories of the Delphi I study, five rank groups.

Table 13. Development of consensus in the rankings of the Delphi I study, 5 rank groups.

Intervention	**Consensus**			
	1st round	**2nd round**	**3rd round**	**Δ1→2/Δ2→3**
Ammonia (air)	Rather weak	Rather weak	Rather weak	0/0
Benzene (air)	Very weak	Rather weak	Rather weak	+2/0
Carbon dioxide (air)	Rather weak	Rather strong	Rather weak	+1/-1
Carbon monoxide (air)	Very weak	Rather weak	Rather weak	+2/0
Heavy metals (air)	Rather weak	Rather weak	Rather strong	0/+1
Methane (air)	Very weak	Very weak	Very weak	0/0
Nitrogen oxides (air)	Strong	Very strong	Very strong	+1/0
Nitrous oxide (air)	Very weak	Very weak	Very weak	0/0
Particulates (air)	Very weak	Very weak	Very weak	0/0
Pesticides (air)	Very weak	Rather weak	Rather strong	+2/+1
PM 10 (air)	Very weak	Rather weak	Rather weak	+2/0
Sulphur dioxide (air)	Rather weak	Strong	Strong	+2/0
VOCs (air)	Rather weak	Rather strong	Rather strong	+1/0
BOD+COD (water)	Very weak	Very weak	Very weak	0/0
Heavy metals (water)	Very weak	Rather weak	Rather weak	+1/0
Nitrogen as N (water)	Rather weak	Strong	Rather strong	+2/-1
Oils and greases (water)	Very weak	Rather weak	Very weak	+2/-2
Phenols (water)	Rather weak	Strong	Rather strong	+2/-1
Phosphorus as P (water)	Rather weak	Rather weak	Rather strong	0/+1
Suspended solids	Rather strong	Strong	Strong	+1/0
Oil in ground (Resource)	Very weak	Very weak	Very weak	0/0
P in ground (Resource)	Rather weak	Rather strong	Rather strong	+1/0
Soil erosion (Resource)	Very weak	Rather weak	Very weak	+2/-2

9. Feedback from the Delphi I judges

Statistical analyses of the Delphi process were complemented by a feedback questionnaire and interviews with the judges in order to find out the factors which had affected the formation and expression of their opinions in the Delphi study, to assess the feasibility of the Delphi method for LCA valuation and to identify needs and possibilities of developing the method for future applications.

The questionnaire included four compartments:

1. Opinions and experiences of the Delphi process
2. Familiarity with the environmental impacts of various interventions
3. Criteria for evaluating interventions
4. Cultural theory typologies.

The structure of the questionnaire is given in Table 14. The questionnaire was complemented by personal interviews with the judges.

In the interviews, the following issues were also discussed with the experts:

– Possible causes for the controversiality of the answers of the judges ?

– How did the judges perceive the environmental impacts during the Delphi process (e.g. on emission level or on damage level) ?

– General aspects of the valuation of environmental impacts ?

– Economic considerations ?

A summary of the feedback from the judges in Delphi I is presented in chapter 9.1. Results concerning the familiarity of the judges with the environmental interventions (questionnaire part 2.) are in chapter 9.2. A correspondence analysis of the answers concerning the criteria for evaluating interventions (questionnaire part 3.) is in chapter 9.3 World views of the judges on the basis of the questionnaire are in chapter 9.4.

The verbal feedback of eight panellists to the questionnaire and in the interview are enclosed in Appendices A and B.

Table 14. Overview of the structure of the judge feedback questionnaire.

Compartment	Scale	Statements
1. Opinions and experiences of the Delphi process	strongly agree ….. strongly disagree	1.a. The task was interesting 1.b. It was difficult to understand the tasks 1.c. The background information provided for the evaluation was adequate 1.d. The process changed my view about the harmfulness of some environmental interventions 1.e. The process made me change my evaluation of some environmental interventions in the iteration 1.f. The final coefficients obtained are generally suitable for valuation in LCA applications in the Nordic countries 1.g. The list of interventions should be expanded (Please suggest additions if you agree) 1.h. The final overall ratings of heating oil from North Sea oil and rape seed oil were quite reasonable 1.i. I would be prepared to participate in a similar process in a few years time to update the ratings
2. Familiarity with the environmental impacts of various interventions	extremely familiar … not at all familiar	List of interventions included in the Delphi study
3. Criteria for evaluating interventions	An enclosed list of criteria (see Appendix A)	3.a. CO_2 3.b. NO_x 3.c. SO_2 3.d. VOC 3.e. Benzene 3.f. PM10 3.g. Heavy metals to air 3.h. Heavy metals to water 3.i. Nitrogen to water 3.j. Oil in ground as a resource 3.k. Other comments
4. Cultural theory typology	A four-field typology: Fatalist, Hierarchist, Individualist, Egalitarian, total of 10 points	
5. Comments		

9.1 Summary of feedback from the Delphi I judges (see also Appendixes B and C)

Generally, the experts found the task interesting and even as a learning experience. They found that their own evaluations were influenced by the opinions of their colleagues.

Most of the experts were used to concentrating on one problem at a time. They found it difficult to think about various environmental issues or harms as a whole at one time.

There were some difficulties in understanding the tasks. The tasks and instructions could be developed. Some argued that the weighting task was too difficult: “comparing the reduction of one emission to the reduction of another was impossible to understand…” When making an evaluation one has to think about the endpoints, also the final effects. This means making “models in the mind”. Conceptualisation of the problem for the valuation task would be helpful. Should the valuation task also be moved towards to the endpoints? Anyway, according to some experts: “It is not possible to put a value on emissions as such either”… “Putting the impact categories into an order would be ok.”

On the other hand, the interventions were seen as important in outlining the processes. One intervention has many impacts and consideration must be given to where it is occurring, what or who are exposed, etc. Having the task at the intervention level was also seen as good, because there were numbers available. The impact categories comprise more uncertainty, depending on the place, and impacts are caused on several levels (plants, animals etc.) anyway.

In the opinion of the experts the intervention list for the valuation task could not be expanded, as it was already quite heavy. On the other hand, many issues that could be important in some cases were missing.

Valuation should be made category by category; the global and local impacts should be separated. Perhaps it should also be done separately in the short or long term? Generally, decision-making has its own logic for national problems and global problems.

In some cases opinions clearly differed, especially in the case of CO2; some thought that it caused severe effects and that the effects of other interventions were minor. Others expressed the view that in the case of the Nordic countries the impacts were not of much importance. The criteria used by the experts depended on their individual specialisation: for example, an expert from the public health field considered the exposure of the population, while an expert on eutrophicative emission placed emphasis on that aspect. The experts were from quite different backgrounds and it is important to choose people carefully to get a good balance.

According to the experts it is impossible to base the valuation solely on environmental factors. The economic issues were not really considered, but it was still not possible to be totally “free” of them. It was suggested that a rough estimate of the costs of the actions should be included in the interview. In decision-making the perspective of costs is anyway important. It was stated that if economic valuation were possible, it would be good. And, for example, rape seed oil is largely an economic issue.

Delphi was regarded as a good technique for mapping the experts’ views in an orderly form. But no “truth” is produced. An “ideal view” may be reached with Delphi. The way in which that view is reached and the nature of the panel should be made transparent. The result of the process represents the collective view of a certain expert group, and it should treated as such! It would be dangerous to view the outcome as being irrefutably definitive. It was suggested that the views of lay people should also be sought at the “endpoints” level. Every group who participates in the production of environmental impacts – industry, the authorities, scientists, consumers – must also take part in their valuation.

Lack of knowledge was seen a big problem in valuation! There are also impact-related issues on which there is currently no information available. This means that the valuation of environmental problems cannot really be done discreetly. Whatever method is used, valuation will always remain a difficult task.

The results were regarded as being quite uncertain. The application should always be considered; the valuation cannot be made generally. But should only the views of experts be collected or should the perspective of actual decision-making also be included? Anyhow, LCA was seen as a good way of organising present knowledge.

The experts were not very enthusiastic about the inclusion of more argumentation in the task if that would entail more work. (Perhaps it would be possible if the task, comprising different interventions and impacts, could be divided between different experts.) Generally, all would be willing to participate again.

9.2 Familiarity of Delphi I judges with the interventions

It is unlikely that all the judges knew everything about all of the 23 interventions that they were asked to evaluate in Delphi I, but the depth of their familiarity was not measured at the time. They were therefore asked in the follow-up questionnaire to rate their familiarity with each intervention on a scale ranging from extremely familiar to not familiar at all (Question 2). The results of this assessment are shown in Table 15.

Table 15. Familiarity of judges with interventions.

		Mean	S.dev
2l	Sulphur dioxide	4.4	0.7
2e	Heavy metals	4.3	0.7
2c	Carbon dioxide	4.1	1.0
2g	Nitrogen oxides	4.0	0.9
2o	Heavy metals	4.0	0.7
2k	PM10	3.9	1.2
2m	VOCs	3.9	0.9
2I	Particulates	3.8	0.9
2p	Nitrogen as N	3.6	1.0
2a	Ammonia	3.6	1.1
2f	Methane	3.6	0.8
2d	Carbon monoxide	3.6	0.8
2j	Pesticides	3.6	1.1
2b	Benzene	3.4	1.1
2q	Oils and greases	3.4	0.5
2h	Nitrous oxide	3.4	1.0
2s	Phosphorus as P	3.2	1.0
2u	Oil in ground	3.2	1.0
2v	P in ground	3.1	0.7
2r	Phenols	3.1	0.9
2w	Soil erosion	3.0	0.7
2n	BOD+COD	2.9	1.1
2t	Suspended solids	2.8	0.9

n=9 Extremely familiar = 5 Not at all = 1

It is clear from the table that there is a considerable range in the level of familiarity with the interventions. What is less obvious is that the level of familiarity is correlated with the severity of the problems arising from the interventions. When the above means are compared with the total valuation points in Delphi I (Table 3.8, section 3, p. 37), a strong correlation is found ($r = 0.49$, $p = 0.026$). This effect can also be seen in Table 16 below.

Table 16. Rankings of interventions for various familiarity levels.

Percentage of comparisons where 1st intervention is ranked higher than 2n		**1st intervention**				
		Not at all familiar				Extremely familiar
2nd intervention		*1*	*2*	*3*	*4*	*5*
Not at all	*1*	-	68	61	81	91
	2		-	39	67	66
	3			-	68	66
	4				-	59
Extremely	*5*					-

The table is based upon 2 274 pairs of comparisons of the ranking of interventions made by the 9 judges in the final iteration of Delphi I. The easiest way to explain the table is by way of example. For those comparisons, for instance, where the judges were extremely familiar with an intervention, but not at all familiar with the other (top right hand corner of the table), 91% ranked the intervention with which they were extremely familiar (response 5) higher than the one with which they were not at all familiar (response 1). In 81 % of the cases in which a comparison was made between an intervention with a familiarity rating of 4 and the other with a rating of 1 (not at all familiar), the intervention with a familiarity rating of 4 is given a higher ranking. The table shows that in all but one case, the intervention with the higher familiarity rating tends to be ranked higher than one with a lower familiarity rating. The exception is where the familiarity ratings are 3 and 2 – only 39% ranked the intervention with the higher familiarity rating above the one with the lower rating. This is probably a statistical freak.

The conclusion to be drawn from the table is that ranking – and therefore rating – depends on familiarity, or vice versa. It could either be that judges are more familiar with more harmful interventions, or that they consider the interventions with which they are familiar with to be more harmful than those with which they are not. The direction of causality cannot be established here. The implication, though, is that one is likely to obtain more reliable rankings and ratings where the interventions are roughly equally familiar to a judge and, probably, that the higher the level of familiarity, the more reliable the ranking is. If, say, the ranking and rating tasks were only allowed where both interventions had a familiarity rating of at least 4, the number of comparisons in the sample would be reduced by roughly two-thirds. Thus it would be necessary, on average, to recruit three times as many judges to achieve the same number of responses overall. The responses would, however, be of significantly higher quality. It would also be necessary to ascertain the level of familiarity for each intervention before assigning the ranking and rating task.

9.3 Correspondence analysis of responses to the Delphi follow-up questionnaire

9.3.1 Method

Correspondence analysis is a method of representing data from a contingency table spatially, using a minimum number of dimensions, so that it can be examined for structure. The contingency table consists of stimuli as columns and attributes as rows. The technique seeks to represent stimuli and attributes in the same space. In our study the stimuli are 10 substances and the attributes are 20 statements. The contingency table is shown in Table 18, based on the responses of 9 experts.

There are several approaches which can be taken to correspondence analysis. The one used here is reciprocal averaging. The ’goodness position of fit’ of the model is expressed in terms of 'inertia' – the reluctance of each of the points in the solution space to move away from its current location. For our problem it was found that the table can be completely represented in 9 dimensions, but each dimension accounts for decreasing amounts of inertia, as follows (Table 17):

Table 17. Dimensions of the contingency table (Table 18) and their inertia according to correspondence analysis.

Dimension	Inertia (%)	Cumulative (%)
1	39.1	39.1
2	18.8	57.9
3	17.6	75.5
4	8.0	83.5
5	5.6	89.1
6	4.5	93.6
7	3.0	96.6
8	2.5	99.1
9	0.9	100.0

(The first dimension accounts for 39.1 % of total inertia, the second for 18.8 %, etc.)

It is probably not worthwhile going beyond, say, 5 dimensions, in which virtually 90 % of the total inertia is captured. Perhaps the most efficient solution has 3 dimensions, accounting for three-quarters of the total inertia.

Table 18. Contingency table.

		CO_2	NOx	SO_2	VOC	Ben-zene	PM_{10}	HM (Air)	HM (Wat)	N to wat	Oil grnd	Total
A	Has considerably damaged environmental objects in the past in Nordic regions	0	4	5	1	0	2	1	2	4	1	20
B	Is presently damaging environmental objects in Nordic countries	0	7	6	1	1	2	1	0	4	1	23
C	Its level is approaching the tolerable limit in Nordic countries, and will become a problem in the near future	0	1	0	2	0	1	1	0	1	0	6
D	Causes long-lasting effects and may lead to environmental problems in Nordic countries in the longer term	7	4	1	0	0	1	3	1	1	1	19
E	Has recently been recognised as a cause of environmental problems in Nordic countries	1	0	1	2	0	4	0	1	0	0	9
F	Is the major source of a particular environmental problem in Nordic countries	3	4	4	0	0	2	1	0	3	0	17
G	Is a minor source of a particular environmental problem in Nordic countries	0	1	0	5	5	0	4	2	1	2	20
H	Is under control in Nordic countries	1	0	4	1	3	0	3	4	1	2	19
I	Causes important effects according to recent findings	4	3	3	2	0	6	1	1	1	1	22
J	Has been receiving much publicity	6	3	3	2	0	3	1	1	2	0	21
K	Causes impacts whose consequences are severe and not tolerable	2	5	5	1	1	5	1	0	3	2	25
L	Causes impacts which are tolerable	1	0	0	0	2	0	1	1	1	0	6
M	Would be very expensive to reduce	6	4	3	2	0	0	0	1	2	1	19
N	Causes irreversible damage	4	4	3	0	0	2	1	1	2	3	20
O	Would lead to damage which is very expensive to restore	5	4	3	2	1	1	1	3	2	2	24
P	Causes damage which has only minor cost	0	0	0	0	0	0	0	0	1	0	1
Q	Has global impacts	7	3	3	1	0	2	1	2	0	2	21
R	Affects large areas	5	6	4	2	1	3	3	3	3	1	31
S	Affects the environment outside the Nordic countries	5	5	4	3	2	2	1	3	2	3	30
T	May endanger the environment of future generations	6	4	5	1	1	1	2	2	3	2	27
	Total	63	62	57	28	17	37	27	28	37	24	

Table 19 shows the co-ordinates of both the statements and the substances on each of the first five dimensions. Figure 27 show a plot of the substances (in black rectangles) and the statements, represented by their letters. The plot shows that PM10, for instance, is very closely related to statement E (recently recognised), benzene is close to G (minor source) and L (impacts which are tolerable), and H under control – which the heavy metals are also close to.

Table 19. Correspondence analysis.

Co-ordinates on first 5 dimensions

			Dimension		
Statement	**1**	**2**	**3**	**4**	**5**
A	85	52	49	21	17
B	86	52	53	30	14
C	64	11	25	80	6
D	85	100	21	45	47
E	88	0	0	12	35
F	100	66	43	36	44
G	0	50	27	46	12
H	29	78	35	0	28
I	93	43	17	22	35
J	92	65	22	41	36
K	89	47	37	20	28
L	4	85	38	29	98
M	89	87	28	47	0
N	90	80	33	20	15
O	72	79	28	29	9
P	90	45	100	100	100
Q	87	88	16	22	17
R	76	68	31	33	31
S	68	70	28	25	4
T	77	83	34	29	24
			Dimensions		
Substance	**1**	**2**	**3**	**4**	**5**
CO_2	92	100	0	65	76
NO_x	93	64	59	64	40
SO_2	89	58	66	11	47
VOC	56	11	1	100	0
BEN	0	57	56	32	95
PMT	100	0	4	17	99
HMA	49	67	37	70	100
HMW	56	77	22	0	57
NIT	84	50	100	82	85
OIL	68	79	38	7	2

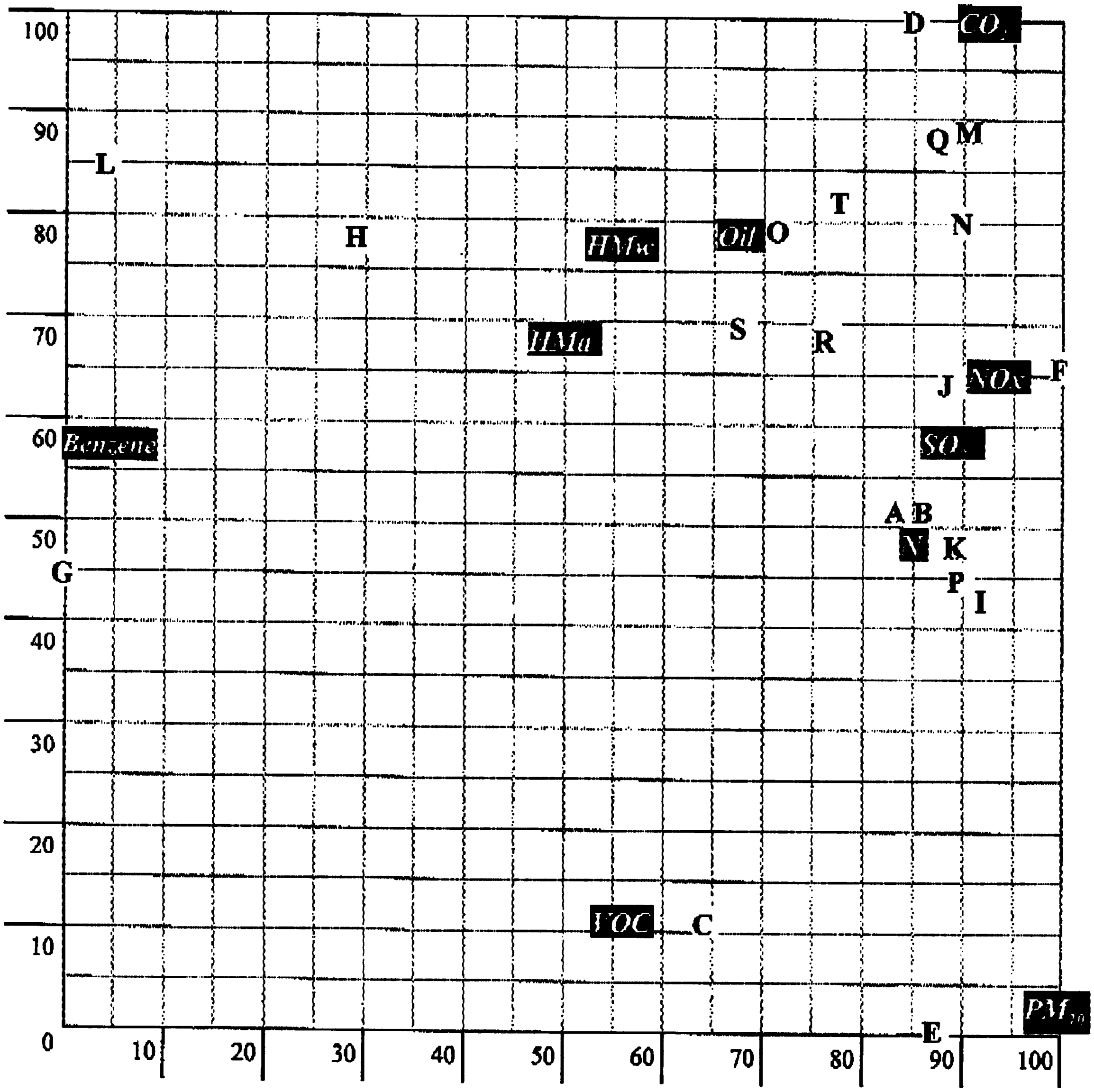

Figure 27. Plot of the first two dimensions.

9.3.2 Interpretation of the output from the correspondence analysis

The output can be interpreted by examining either the two-dimensional plots (as provided earlier) or the projections of the points on each dimension, as attached. The basic approach is to examine the statements at each end of the dimension and try to find a unifying connection. It is not always easy, because all the statements will appear on the dimension, even if they are not connected with it anyway. Here is my interpretation:

9.3.2.1 Dimension 1: Severity/size of problem

The location of interventions with respect to dimension "Severity/size of the problem" is shown in Figure 28. The dimension clearly reflects the severity of the problem, perhaps not unexpectedly. At the severe end there are: major source, important effects, irreversible damage, expensive, severe and intolerable... At the other end there are: minor source, impacts tolerable, under control ...

The co-ordinates correlate significantly (r = 0.57) with the total points for the intervention, as given in the final column of the Table 3.8 in the Neste report. PM_{10} , NO_x, CO_2 , SO_2 , and nitrogen to water are rated at the severe end, Oil as a resource, VOCs, heavy metals to water and air in the middle, while benzene is not seen to cause severe problems.

9.3.2.2 Dimension 2: Time

The location of interventions with respect to dimension "Time" is shown in Figure 29. At the top end of the dimension there are statements such as long-lasting/longer term, global impacts, expensive, future generations, irreversible, expensive to restore. At the other end the statements tend to reflect the past and present: recently recognised, approaching limit, recent findings. This dimension seems therefore to reflect time: the future at one end, the past / present at the other.

CO_2 is seen as a future problem, along with oil in the ground and heavy metals to a certain extent. PM_{10} and VOCs are seen as recent or past problems.

9.3.2.3 Dimension 3: Cost of damage

The location of interventions with respect to dimension "Severity/size of the problem" is shown in Figure 30. This dimension seems trivial, and depends very much on the response of one individual. The only person who has marked statement P (Causes damage which has only a minor cost), has done so for nitrogen to water. This dimension seems to be uni-polar - with everything else at the other end. The deviating response needs checking to make sure it is not a mistake.

9.3.2.4 Dimensions 4 and 5

It was difficult to find any unifying theme for these dimensions, which contribute only a very small amount to the total inertia of the system. Therefore, these dimensions were not studied further.

DIMENSION 1: Severity

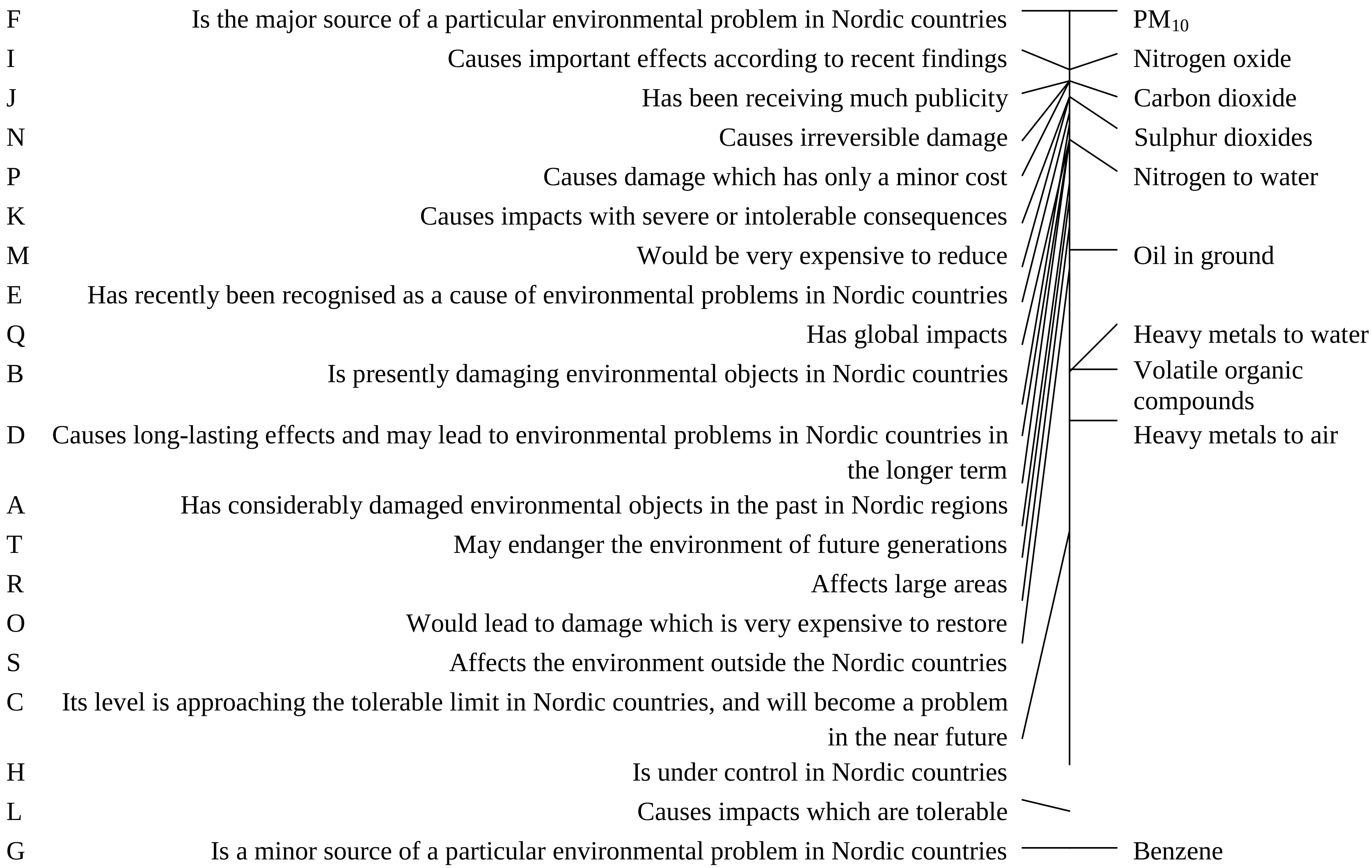

Figure 28. Location of interventions and statements with respect to dimension "Severity".

DIMENSION 2: Time/space

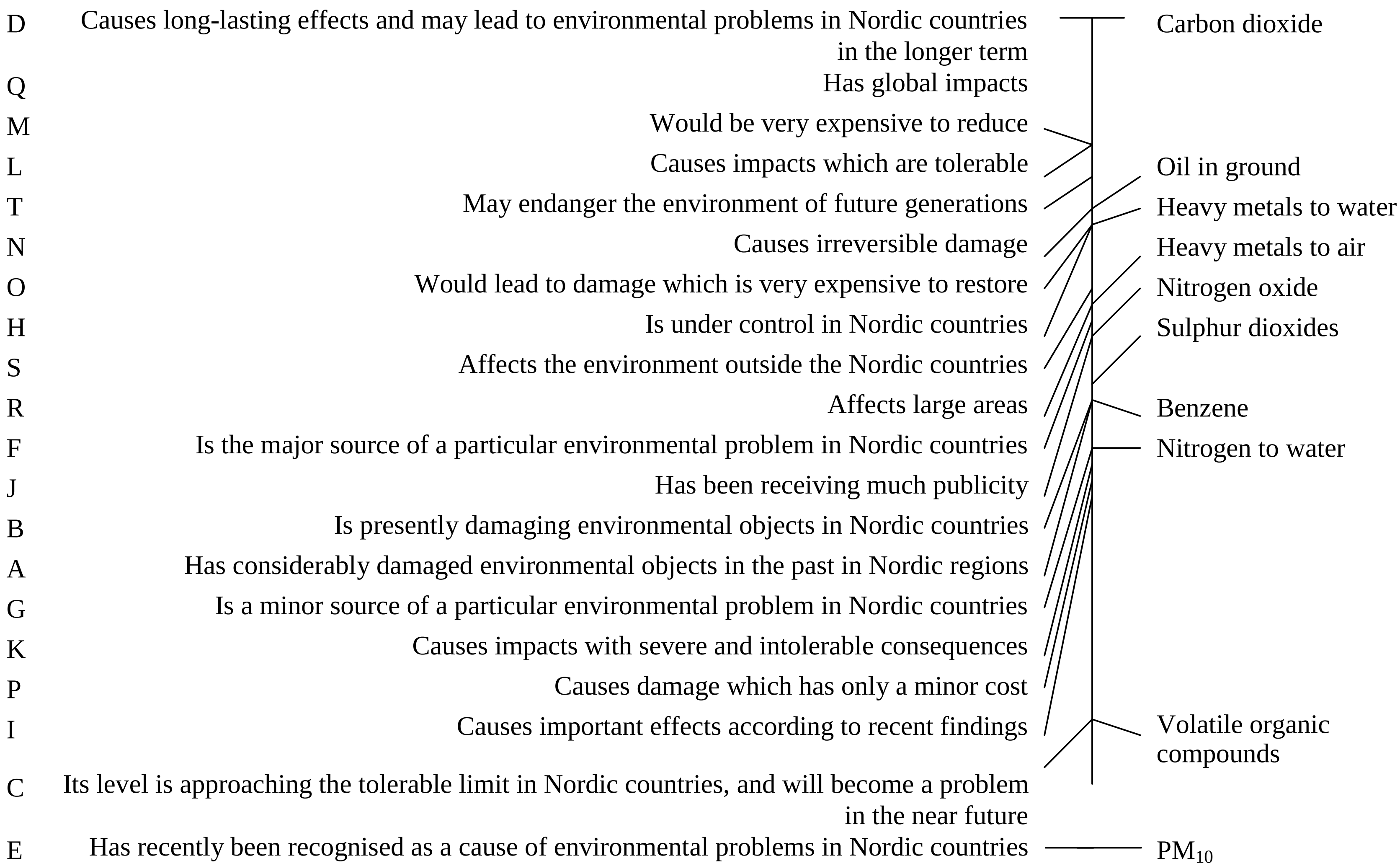

Figure 29. Location of interventions and statements with respect to dimension "Time/space".

DIMENSION 3: Cost

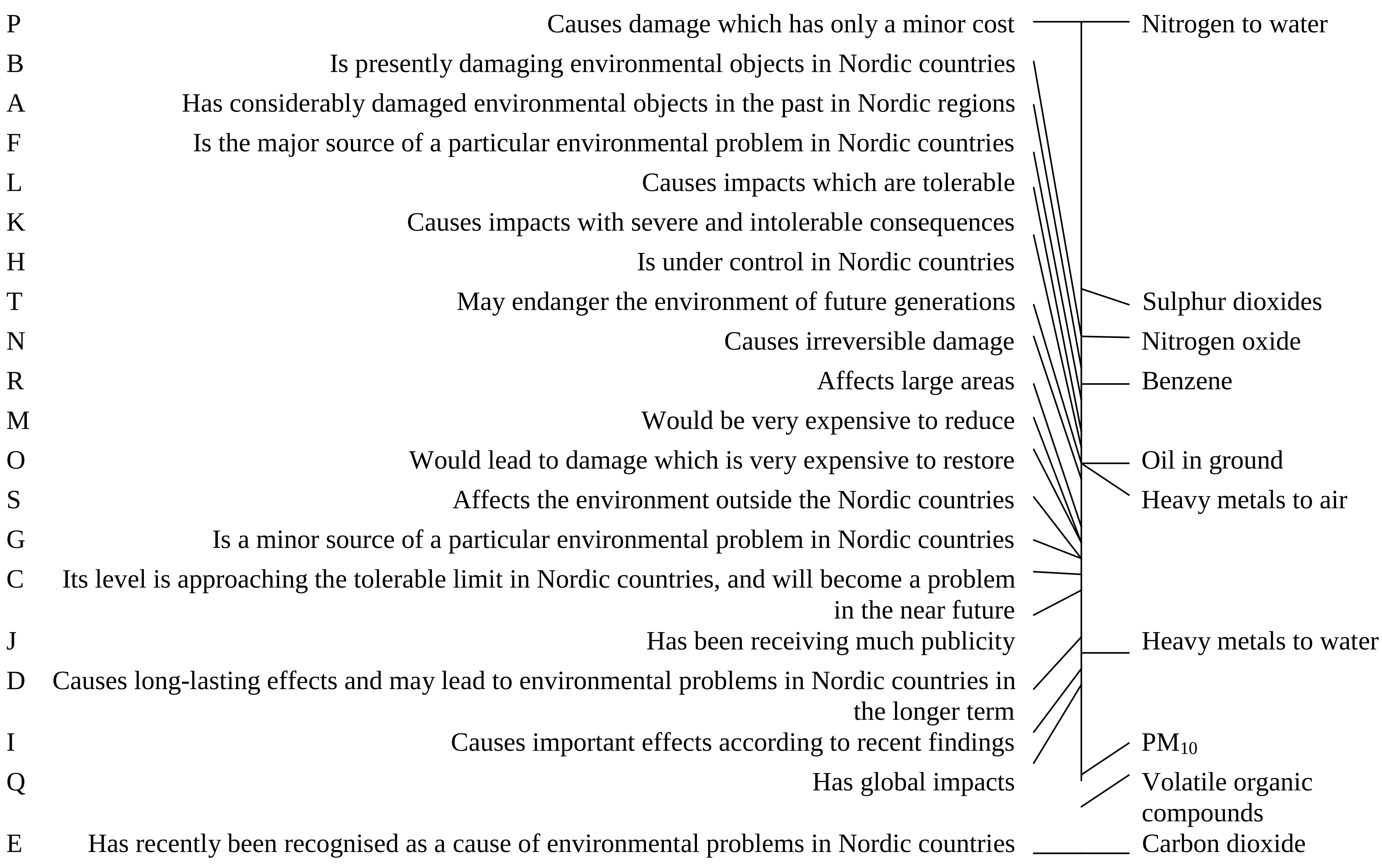

Figure 30. Location of interventions and statements with respect to dimension "Cost".

9.4 World view of the Delphi I judges

A typology of world views was described in section 4.1; the typologies outlined there being: fatalist, hierarchist, egalitarian and individualist. A number of sets of questionnaire scales have been used in the past to try and assign individuals into the typologies. The number of scale items used can range from 24 to 46, and are rather personal and intrusive in nature. As far as is known, none of these scales have been particularly successful thus far. All the studies have been conducted using lay persons rather than experts. It was thus thought that in order to obtain the world view of the experts who took part in Delphi I, it should only be necessary to show them the descriptions of the typologies and ask them to rate themselves on each one. This was done by asking them to allocate 10 ‘points’ in total in such a way that the points reflected their affinity with each typology.

The individual responses for the assignment of points are shown in Table 20, together with summary results for each typology. The judge numbers refer to those used in Delphi I.

Table 20. The individual responses of the judges for the assignment of points.

Judge	Fatalist	Hierarchist	Egalitarian	Individualist
A2	0	6	2	2
A3	0	8	2	0
A7	0	7	1	2
A8	1	2	4	3
B1	2	4	3	1
B4	2	4	4	0
B5	2	4	2	2
B6	0	6	1	3
B7	1	4	3	2
Mean	0.9	5.0	2.4	1.7
S.dev	0.9	1.8	1.1	1.1

Although no judge has allocated all 10 points to one typology, most have, as might be expected, allocated the majority of points to the Hierarchist category. The only exceptions are Judge A8, who gave most points to the Egalitarian category, and Judge B4, who gave equal points to Hierarchist and Egalitarian. Egalitarian attracted the second highest point allocation, ahead of individualist, with, perhaps surprisingly, more than half the judges giving Fatalist one or two points.

In Delphi I the judges were assigned to three categories on the basis of their observed ratings for carbon dioxide. The categories were called “Pro-active”, for those with very high ratings for carbon dioxide, “Mainstream”, for those with moderate ratings, and “No action” for those with low ratings. The categories were purely empirical and the thresholds for the categories were set arbitrarily.

A comparison of these categorisations with the world views obtained in the follow-up study shows an interesting phenomenon, as set out in Table 21 below. The ‘Pro-active” group tends to score relatively highly as Egalitarians, and the “Mainstream/ No action) group relatively lowly.

Although the number of cases is very low (n = 9) and the correlation between the Delphi I category and the Egalitarian points is a little hazy, the table does nevertheless suggest that there is a very good chance of establishing a relationship between world view and valuation factors. If this is the case, such a relationship could be invaluable in selecting and operating panels, and in establishing weightings which could be used to standardise valuation weights to make them more representative of different populations.

Table 21. A comparison of the Delphi I category and egalitarian world view.

Category	Judge	Standardised rating	Egalitarian points
Pro-active	A8	939	4
	B4	704	4
	B7	416	3
Mainstream	B6	157	1
	B1	91	3
	A3	78	2
	B5	63	2
	A7	61	1
No action	A2	12	2

10. Opinions of the interest groups

The acceptance of and possibilities of generalising the results from expert judgement based on the Delphi technique were analysed within groups dealing with environmental issues in decision-making or other activities of interest. The views of possible users of the results in companies, in the public sector and in citizen movements (Ministry of Environment, Finnish Association of Nature Conservation, etc.) were collected in the form of a workshop, utilising the method description produced in the study. The issues discussed in the workshop are presented in Appendix D.

The conclusions of the workshop are briefly presented in the following:

- **Responsibility questions**

Models need to be developed and used for the analysis and structuring of decision problems. However, models do not assume responsibility for the final decisions; it must be borne by someone, and it should be clear by whom.

- **Choosing the experts, world views**

There are two ways of choosing experts (or panellists) for valuation task: comprehensively, choosing the panellists with relevant attributes concerning knowledge and world view, or choosing certain panellists and describing their qualifications.

- **Context**

The context of a model should always be explained: for what kind of problem and in what kind of context it is relevant. The target area and time should be defined. Valuation weights free from the context are not possible, at least not yet.

- **Methods**

Models can structure the process and help understanding in decision-making situations. It is recommendable to use several models in parallel.

- **Learning process**

Understanding the decision problem and the causalities involved is more important than the resulting numbers. Expert knowledge is also a slice of reality.

11. Expanding valuation indexes for a larger number of interventions

11.1 Introduction

In a valuation task performed by a panel, it is possible to include only a very limited list of environmental interventions. Therefore, expanding the valuation indexes for further interventions may become necessary. Expansion can be based on another study, which means that a reference intervention common to both studies is used and weights for the missing interventions are taken from that other study. Further data may also be collected from experts, in which case, similarly, a suitable reference intervention is used to make comparisons of the missing interventions. Still another approach for expanding the scope of interventions would be to have a valuation performed for different intervention categories (see Appendix), firstly by “general experts”. Further valuation would then be made within the categories by choosing more specialised experts for each category, Wilson (1997–1998).

In the Delphi I study the index values for additional interventions were derived by using generic fate and exposure models developed by Mackay. These models allow comparative valuations. Also interventions giving rise to similar problems may be compared on the basis of a valued reference intervention (Wilson and Jones, 1996).

11.2 Use of Mackay models in the Delphi I study

Mackay models were used in the Delphi I study to estimate relative concentrations for “similar substances” in the standard environment, and to compare their impact to that of a reference substance, valued in the Delphi exercise. According to the Mackay models, the impact of a substance on the environment is caused by the exposure of organisms to that substance, in all environmental compartments, and by the toxicity of the substance to the organisms.

Mackay models are used to predict the transport and fate of a chemical substance in the environment. The environment is divided into the “compartments” of air, water, soil and bottom sediment. The model predicts how a release in a particular compartment will move to other compartments.

The models require the following data about the physical system being modelled (see also Table 22):

a) areas and depths of compartments
b) temperatures of the systems
c) densities and the fraction of the compartment composed of organic carbon.
d) physical properties of the substance being modelled, such as molecular mass, melting point, vapour pressure and solubility in water
e) partition coefficients between compartments
f) rates of flow of the water, air and sediment compartments
g) degradation rates of the substance in each compartment (half-lives)
h) velocities of each of the transports between the media being studied.

According to the application performed in Delphi I, the concentrations of the substance in each compartment were compared with limits for exposure. *The hazard potential ratio*, which is defined as the ratio of the predicted concentration in the environment to the no-effect concentration, was calculated for each substance for different combinations of compartments (air, water, soil, bottom sediment), pathway (inhalation, ingestions, freshwater, seawater) and assessment criterion (reference concentration, reference dose, cancer unit risk, cancer slope factor, TLV, acute LEC, chronic LEC, etc.). The criteria were taken from the human toxicity limits and standards as specified by USEPA in the IRIS (Integrated Risk Information System) and HSDB databases. The water quality limit values were those used in Michigan.

In order to obtain relative hazard potentials for different substances, the maximum hazard potential ratio of the calculated combinations for each substance was chosen as the value to be compared to the maximum ratio of benzene. The ratio for benzene was given a value of 1. The valuation factors for the substances could then be calculated on the basis of these relative hazard potentials by multiplying them by benzene's valuation indexes produced in the Delphi I study. The resulting valuation factors are given in Table 22.

11.3 Updating of Mackay models

In the Delphi I study Mackay models were applied to Nordic countries, for the total Nordic region, and the values for the areas were largely derived from those used for Minnesota, as studied by Mackay. Consequently, the application was based on quite rough estimates, and afterwards it has been updated by Prof. Jaakko Paasivirta from the University of Jyväskylä. In the updating study, Southern Finland was used as a representative area with a corresponding drainage area of the water system going down to Selkämeri in south-west Finland. More accurate data was collected on the photo- and biochemical degradation, especially of easily degradable substances, as well as on temperature dependencies, steam pressure, and water solubility. A PC program developed for the purpose, FATEMOD was used for the computation. Environmental data on the target area used both in Delphi I and in the updating study as input to the model are presented in Table 22. Additional data on the physical and chemical characteristics of substances, such as steam pressure of liquid, solubility in water and reaction half-lifetimes were also included in the updated model.

Table 22. Data used for different environmental compartments in the Mackay model calculations in Delphi I and the updating study by Paasivirta.

Factor	Air		Water		Soil		Sediment [a]	
	Delphi I	Updated	Delphi I	Updated	Delphi I	Updated	Delphi I	Updated
Study area, Delphi I: Nordic countries Updated: part of south-west Finland								
Area, m^2	1,11E+09	3,64E+10	8,20E+07	3,34E+09	1,02E+09	3,30E+10	8,20E+07	3,34E+09
Height, m	1000	1000	10	2	0,2	0,1	0,05	0,02
Volume, m^3	1,11E+12	3,64E+13	8,20E+08	6,68E+09	2,04E+08	3,30E+09	4,10E+06	6,68E+07
Advection residence time, h	—	100	—	5010	—	1,00E+10		200000
Advection flow, m^3/h		3,64E+08		1,33E+06		0		334,1
Organic carbon fraction, abs.	—	—	—	—	0,02	0,05	0,04	0,06
Density, kg/m^3	1,185	1,185413	1000	1000	2400	2400	2400	2300
System temperature, °C	25	5						
Rainfall, mm/year	?	700						

[a] bottom, not suspended

In the updated calculations the model output was the concentrations of the studied substances in a stationary situation in the representative area of south-west Finland at 5 ^{o}C and under a constant emission of 10 000 kg/hour. The obtained concentrations were then transformed into so-called Cr values, which are extrapolated concentrations for an emission intensity of 1 000 kg/km^2,h. This was done by multiplying by a factor of 3 636, which is the ratio of 1 000 kg/km^2,h to the emission intensity corresponding to the reference flow used in the model calculations (0.275 kg/km^2,h = 10 000 kg/h/36 358 km^2). The Cr values in the air, water and soil are given in Table 23.

Table 23. Comparison of Cr concentrations in the air, water and soil as produced in Delphi I and in the updating study by Paasivirta. The Cr concentrations are output concentrations of the FATEMOD model extrapolated to an emission intensity of 1 000 kg/km², h.

	Air mg/m^3		**Water** mg/l		**Soil** µg/g	
	Delphi I	**Updated**	**Delphi I**	**Updated**	**Delphi I**	**Updated**
Benzene	142.44	77.52	0.06	0.78	0.10	1.94
Toluene	23.94	30.21	0.07	0.22	0.26	2.26
Xylenes	23.94	14.76–30.20	0.08	0.13–0.39	0.67	1.46–3.35
Acetaldehyde	23.46	29.90	0.87	15.52	0.19	0.90
Formaldehyde	8.00	22.39	18.60	0.29	223.00	0.06
1.3-Butadiene	7.04	6.73	0.02	0.002	0.02	0.004
Methanol	73.11	80.54	25.90	105.34	92.60	9.30
MTBE	141.02	57.96	0.0007	6.68	0.02	3.59
Anthracene	23.94	42.80	0.0000003	37.92	0.003	642,84

Table 24 includes the PNEC (Predicted No-Effect Concentration) values used in the updating. These data are collected from chemical handbooks, reference values for the workplace and the environment, WHO, the Finnish Ministry of Social Affairs and Health, and toxicological studies (Paasivirta, 1998). The PNEC values for air are estimated so that they are one tenth of the corresponding reference values for the indoor air for humans. The PNEC values for water and soil, when not found in published form, are assessed as one tenth of the LOEC concentration in water and in soil for the most sensitive organisms. When the LOEC was not known, the PNEC was estimated as 1% of LC50 (or EC50) (ibid.). The relative risk coefficients ($Rr(i)$, i refer to different emissions) are calculated as

$$Rr(i) = Cr(i)/PNEC(i) \tag{43}$$

A comparison of the PNEC values used in the updating study and in Delphi I for computing the maximum hazard potential ratio (several alternatives) is presented in Table 25.

Table 24. PNEC values used in the updated calculations by Paasivirta.

Substance	**PNEC in air** mg/m^3	**PNEC in water** mg/l	**PNEC in soil** µg/g
Anthracene	0.02	0.0019	900
Acetaldehyde	2.5	0.308	
Benzene	0.2	0.005	200
Butadiene	1.1	0.3	0.05
Formaldehyde	0.06	0.02	
Methanol	26	60	
MTBE	18	40	
Toluene	18.8	0.013	
o-Xylene	44	0.013	
m-Xylene	44	0.037	
p-Xylene	44	0,02	

Table 25. Comparison of the PNEC values used in the updating study by Paasivirta (1997) and in Delphi I (Wilson and Jones, 1996, several alternatives) for computing the maximum hazard potential ratio.

Emission	**Air**			**Water**						
	PNEC mg/m^3	Inhalation reference concentration [a] mg/m^3	Inhalation unit risk [a] mg/m^3	PNEC mg/m^3	Oral slope factor [b] mg/m^3	Oral reference dose [b] mg/m^3	TLV (US) mg/m^3	TLV (Finland) mg/m^3	Ambient water quality criteria W and F cons	WHO 93 mg/m^3
Anthracene	0.02	0	2.2E–09	1.9	0	10 500	0.2	0	0.0028	0
Acetaldehyde	2.5	0.009	8.3E–09	308	0	0	45	90	0	0
Benzene	0.2	0	2.8E–07	5	1 015	0	32	32	0.66	10
Butadiene	1.1	0	1.3E–08	300	0	0	4.4	0	0	0
Formaldehyde	0.06	0	0	20	0	7 000	0.37	0	0	900
Methanol	26	0	0	60 000	0	17 500	262	0	0	0
MTBE	18	3	0	40 000	0	0	144	0	0	0
Toluene	18.8	0.4	0	13	0	7 000	188	0	1 430	700
o-Xylene	44	0	0	13	0	70 000	434	0	-	500
m-Xylene	44	0	0	37	0	70 000	434	0	-	500
p-Xylene	44	0	0	20	0	70 000	434	0	-	500
Source	Paasivirta (1997)	Wilson and Jones (1996)	Wilson and Jones (1996)	Paasivirta (1997)	Wilson and Jones (1996)	Wilson and Jones (1996)	Wilson and Jones (1996)	Wilson and Jones (1996)	Wilson and Jones (1996)	WHO (1993)

[a] allowable concentration in air, when the daily intake of air is assumed as 20 m^3/day, for a person of 70 kg weight

[b] allowable concentration in water, when the daily intake of water is assumed as 2 l, for a person of 70 kg weight

11.4 Updated valuation coefficients

Table 26 presents the relative risk coefficients and valuation indexes for additional emissions based on Mackay models and the PNEC values in Delphi I and the updating study. Large changes in concentrations have occurred to the water compartment but also to soil. The changes in concentrations to air have changed quite moderately. In the case of a certain emission the changes to air, water and soil have been different. The valuation factors obtained depend on the data used for modelling and on the no-effect values (PNEC) used for the emissions to air, water and soil.

When comparing the updated results to the results of Delphi I, in the case of the emissions to air, the emission concentration has doubled for some emissions (benzene, formaldehyde or anthracene). For some other emissions (MTBE) the rate has halved. The rest have remained about the same. In water emissions, in most cases the concentration rate has increased by a factor of 10–20. In the case of formaldehyde and butadiene the concentration has decreased. Emission concentrations to soil have increased in the case of butadiene, toluene, xylenes, acetaldehyde, MTBE and anthracene, in most cases by a factor of 5–20. Rates of formaldehyde, butadiene and methanol have decreased considerably.

Table 26. Relative risk coefficients and valuation indexes for additional emissions based on Mackay models and PNEC values in Delphi I and the updating study by Paasivirta.

	Relative risk coefficients, Rr(i)			**Valuation indexes**		
		Updated			**Updated**	
	Delphi I	PNEC from Delphi I	PNEC from Updated	**Delphi I**	PNEC from Delphi I	PNEC from Updated
Anthracene	0.0060	0.264	51.50	2	86	16 788
Acetaldehyde	0.2598	43.456	0.13	85	14 167	42
Benzene	1	1.000	1.00	326	326	326
Butadiene	9.9201	0.099	0.02	3 234	32	5
Formaldehyde	0.5234	7.641	0.96	171	2491	314
Methanol	0.0014	0.001	0.01	0	0.2	3
MTBE	0.0049	0.002	0.01	1	1	3
Toluene	0.003	0.009	0.04	1	3	14
o-Xylene	0.0003	0.0001	0.08	0.001	0.03	25
m-Xylene	0.0003	0.00004	0.01	0.001	0.01	3
p-Xylene	0.0003	0.0001	0.03	0.001	0.02	9

11.5 Sources for no-effect values

Calculating relative risk coefficients (as referred to in the updated calculations) or hazard potential ratios (in Delphi I) could be based on different guideline values or limiting values. Using different comparative values will probably result in different relative values for substances.

In the work of the World Health Organisation (WHO) the term *guideline values* means the numerical values for maximum concentrations presently considered adequate for the protection of public health in the view of the experts. These are given with explanatory texts and their purpose is to guide risk management decision-making of risk management. In principle, *limiting values* may not be exceeded.

Guideline values are defined by the WHO and by its working groups. In 1987 the “Air Quality Guidelines” was published by the WHO for 28 pollutants, of which 12 were organic and 16 inorganic. The values have been given as maximum concentrations. There is, however, a need to amend the values and the correctives may be available by the end of 90’sThe values given by the WHO have basically been defined for various substances that do not cause cancer but are otherwise detrimental to health and comfort. Additionally, the substances’ cancer risks for the population have been evaluated.

Other sources for guideline and limiting values could be found, for example, from the published summary of the assessments of the US authorities (Calabrese and Kenyon 1991). IARC, the international cancer research centre, has been classifying the substances according to their known or assumed potential to cause cancer.

The following bases have been used by the WHO:

- IARC group classification
- Risk estimate based on carcinogenic endpoint
- Toxicological endpoint
- Sensory effects or annoyance reaction
- Ecological effects.

Guideline and limiting values are also given in EU directives. They are based on dose-response studies carried out by the WHO. The EU directives are mainly concerned with the conventional pollutants (SO2, NO2, O3 etc.) Additionally, national reference values may be established for various substances.

12. Comparison of results from the Neste study, expert valuations with valuation data from other sources

In the following we compare the valuation factors produced in the Delphi I study to two Finnish studies, a German Delphi study, Euro-Barometer and the Swedish EPS valuation method. The comparisons are made in three different ways. Firstly, relative contributions of different impact categories to the total penalty point score are compared, including all the methods and using the total intervention quantities of the target area of each method as the basis of the total score calculations. Secondly, the relative contributions of different emissions to the total penalty points given by the Finnish valuation methods are compared with those obtained by applying the Delphi I indexes. The total scores, which are normalised to 100 for the comparison, are calculated for the Finnish valuation methods using the total annual emissions in Finland, and, for Delphi I, using the sum of the total annual emissions in Finland, Sweden and Norway. The third comparison is between the valuation results of different methods for rape seed oil and light fuel oil, which were the LCA cases in the Delphi I study.

It must be noted that the results of the different valuation methods are not fully comparable. Assumptions had to be made to obtain the following figures. Not all of the methods comprise the same intervention list, for instance. Therefore, not all of the inventoried interventions are valued here because of missing valuation coefficients.

The first of the compared valuation methods (SYKE) was developed in a Finnish study, which was carried out by the Finnish Environment Institute (Seppälä, 1997). Valuation of impact categories was performed by different expert groups. The experts were from the Finnish Environmental Institute and from other organisations representing different skills regarding environmental issues. The categories weighted were:

- Climate change
- Acidification
- Ozone formation
- Ecotoxicity
- Eutrophication
- Oxygen depletion
- Impacts on biodiversity
- Consumption of fossil fuels.

In the second Finnish study (Statistics Finland) by Statistics Finland (Puolamaa et al., 1996), the opinions of different stakeholders about the weights of impact categories were collected by means of the AHP method. Stakeholders from manufacturing (the largest group), traffic, agriculture, environmental NGO's, public administration, politicians, research and the media were involved. The valuation factors of the Swedish EPS are based on the cost to society of protecting biodiversity, OECD market prices, the cost to society of reducing excess deaths caused by various risks, and peoples' willingness-to-pay to avoid diseases, suffering and irritation.

The impact category values for the Delphi results could be roughly derived by aggregating the weightings for various emissions. The contribution of emissions in several impact categories, which is the case, for instance, for NO_x emissions, is managed according to data in Seppälä's (1997) report, based on the total emissions in Finland.

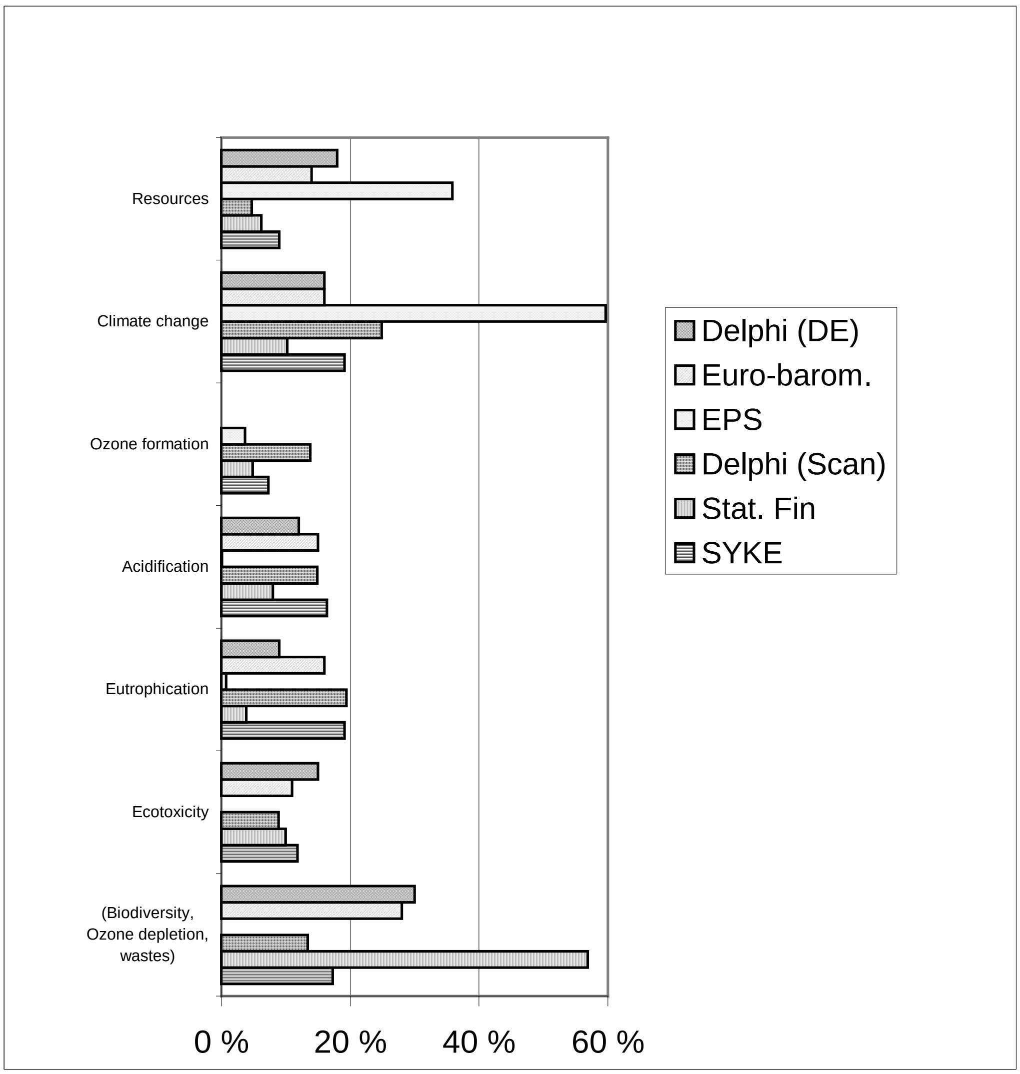

Figure 31. Comparison of the relative contributions of different impact categories to the total scores of different valuation methods. The results are based on the total intervention quantities of the target areas of each method.

In the Figure 31 the relative importance of different impact categories in the total penalty point score are compared using the total intervention quantities of the target area of each method as the basis for the total score calculations. The scores for SYKE, Statistics Finland and EPS are based on the total interventions that occurred in Finland. The results of Delphi (Scan) are based the interventions in Finland, Sweden and Norway. The Delphi (DE) applies to the German area and the Euro-barometer to the European area.

The Swedish EPS method puts great emphasis on the use of fossil fuel and other resources and on the emissions causing climate change. In the study by SYKE (Seppälä, 1997) eutrophication, climate change, acidification, and biodiversity receive the biggest scores, with resource consumption and ecotoxicity receiving somewhat less. The formation of tropospheric ozone has the smallest weight. The study of Statistics Finland weights biodiversity and ozone depletion most highly. Then come climate change, ecotoxicity and acidification. The results based on Delphi I indexes stress climate change and eutrophication most, then comes acidification and ozone formation. The Delphi study performed in Germany weights consumption of resources, ozone depletion, the greenhouse effect, ecotoxicity and wastes quite evenly. Eutrophication is stressed relatively less. In the Eurobarometer ozone depletion, global warming and eutrophication receive the highest scores, waste and ecotoxicity somewhat less. The global impact on climate change also gets a relatively high weight in all the methods compared, but it clearly dominates in the case of the EPS method. The areal impact, acidification, receives quite evenly a considerable weight. In the case of local impact, eutrophication, the weight is relatively high in most of the methods, but in few cases, like EPS and the method of Statistics Finland, the weight if relatively low.

In Figure 32 the relative contributions of different emissions to the total penalty points given by the Finnish valuation methods for the total annual emissions roughly compare with those obtained applying the Delphi I indexes. This comparison comprises only the interventions and impact classes included in all three of the studies. Additionally, there were other environmental interventions or impact classes included in those studies, such as biodiversity or ozone depletion, etc.

The total scores of SYKE and Statistics Finland are calculated using the total annual emissions in Finland, and the Delphi I scores using the sum of the total annual emissions in Finland, Sweden and Norway. For the comparison each total score of the compared intervention is normalised to a value of 100, i.e. the sum of the normalised contributions of the interventions included in … is 100 for each valuation method.

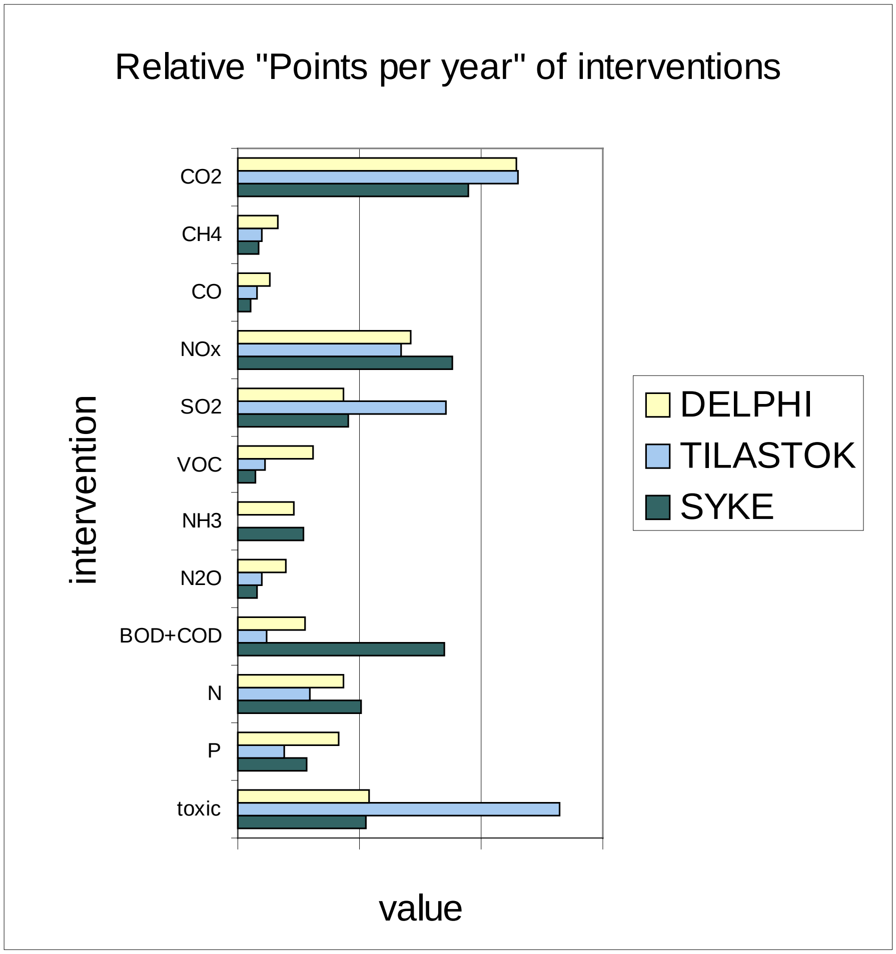

Figure 32. Comparison of the relative contributions of different emissions to the total penalty points given by the Finnish valuation methods (SYKE and Statistics Finland), and Delphi I indexes for the total annual emissions in Finland. The total score basis for Delphi I is the sum of each emission in Finland, Sweden and Norway.

In the case of the interventions compared, CO_2 and NO_x are predominantly stressed in all three studies. Eutrophicative emissions are more weighted in the SYKE method, SO_2 and toxic emissions have the largest contribution in the study of Statistics Finland. The total results of Statistics Finland cover more impact categories than those of Delphi I and SYKE, for instance, emissions causing ozone depletion and radiation. Biodiversity and ozone depletion, for instance, which were not included in this comparison, are scored relatively highly in the method of Statistics Finland and thus reduce the importance of other interventions.

Figure 33 compares the total penalty points given by the Delphi I, Statistics Finland (TILASTO), SYKE and EPS methods for the life cycle inventories of rape seed oil and light fuel oil, which were the LCA cases of the Delphi I study. The comparison comprises the interventions compared in Figure 32, except for the toxic emissions. For the comparison, the valuation factors of each method are harmonised using CO_2

emissions as the basis, meaning that the CO_2 factor has been given a value of 100 in each method and the other factors are reduced to this basis by multiplying them by the ratio of 100 to the original value of the CO_2 factor.

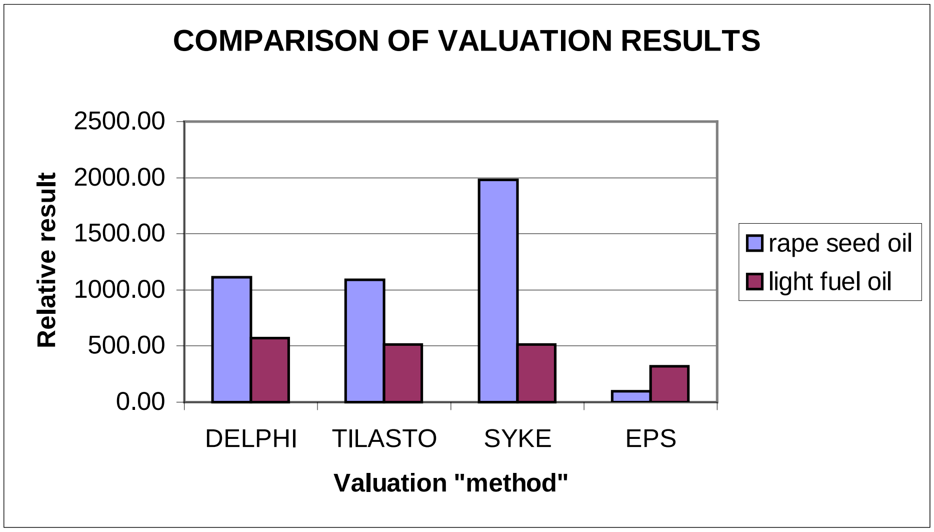

Figure 33. Comparison of valuation results for rape seed oil and light fuel oil on the basis of various valuation methods.

The results of Statistics Finland and Delphi I seem very similar. The ratio between the total scores of the compared systems is about two. The SYKE method gives a value of four to the same ratio, and EPS about 0.25. In the case of all three methods the eutrophicative emissions of the rape seed oil production chain contribute relatively much to the total result. The EPS puts great emphasis on CO2 emissions, and according to the inventory results the CO2 balance is in favour of the rape seed oil.

The similarity between Statistics Finland and Delphi I might be due to the similarity in the origin of the coefficients, because both methods employed Finnish experts in the development of the valuation factors. But, since the SYKE method, which shows quite different scores, was, to a large extent, also based on the same sources, this explanation must be questioned.

13. Expert judgement with the Delphi procedure in LCA valuation

13.1 Key findings

The possible bases for valuing environmental problems are environmental targets, monetary terms or panel judgements. Different techniques or communication procedures may be utilised, for example in the case of panel judgements, to elicit information about the judgements. Transparency and certainty are essential qualities for any acceptable and trusted valuation method.

An ideal feature of expert judgement is that it relies on a group of experienced scientists with a good understanding of environmental problems, who are the most knowledgeable and capable members of society to judge the relative significance of interventions. This should, in principle, enable the method to reveal the real ecological harmfulness of each intervention and the solution could be based on purely environmental rather than economic grounds. The experts will also have the latest, and possibly still unpublished, information available for their judgement.

On the other hand, panel judgement is a social process, which makes it a subjective, even with experts. Depending on the quality of the panellists, judgements will somehow be affected and inspired by facts and scientific information. But the panellists retain the freedom to deviate from scientific standards and knowledge, and instead, to take greater account of the uncertainties and logic of risk management, for example by following a precautionary principle.

The evaluation of the expert judgement method based on the Delphi technique, as developed in the Delphi I study, suggests that both the transparency and certainty criteria may be only partially accomplished by such a method. As for the technical procedure, the method is well documented and transparency is good. Most importantly, the argumentation of the judgements should be increased. Also it was possible to identify some points at which the weighting tasks could be developed and transparency further improved.

In the Delphi II study an attempt was made to clarify the criteria of the experts by means of a list of statements in the case of each intervention. The statements were concerned with the severity or tolerability, the areal allocation and the areal extensiveness, the temporal dimension (present, future, reversibility), the actuality and the economic factors of the environmental interventions (and the problems caused by them). A correspondence analysis of the results showed few main dimensions explaining the valuations. The strongest dimension “severity” was composed of the following criteria, in the order of importance: The intervention is the main source of the specific environmental problem in Nordic countries; It causes significant effects according to the latest research; It has gained much publicity; It causes irreversible changes. The criteria used also seemed to depend on the field of expertise. The temporal dimension was the second strongest, and damage in the future also gained relatively much weight.

The comparison and weighting of the different environmental problems in the task in Delphi I proved to be rather difficult evenfor the environmental experts. Several of the experts involved were specialists in a certain environmental issue. The task formulation could also be developed.

When using panel judgement in valuation, it is an important issue to consider whose values should be respected: those of citizens, experts, politicians, industry or the media. Furthermore, the rights of all countries, species and generations should be included. And the weight of different interest groups should be incorporated in a just manner. For example, the judgements on environmental issues of lay people should be properly studied in comparison with the judgements of experts.

In a valuation task performed by a panel, it is possible to include only a very limited list of environmental interventions. One possibility for expansion of the valuation indexes for further interventions is to use generic fate and exposure models developed by Mackay. The value for a new intervention is obtained by comparing the effect with a valued reference intervention giving rise to similar problems. This was done in Delphi I, and the values were updated in Delphi II. The valuation factors obtained depend on the data of the target area, such as the system temperature, which is used for modelling, and the no-effect values (PNEC), which are used for the emissions in air, water and soil. The calculation of relative risk coefficients (as they are referred to in the updated calculations) or hazard potential ratios (in Delphi I) could be based on different guideline values or limiting values. Using different comparative values will probably result in different relative values for substances.

Comparing a few different valuation methods shows that the different methods have some different weightings. In some methods global issues, such as resource use and climate change, are clearly most stressed, while in some other methods local issues, such as eutrophication, gain more weigh. The weights obtained in Delphi I for different environmental impacts settle somewhere between the two extremes.

According to the interviews with the experts and a group of different stakeholders it is not possible or recommendable, at least not yet, to establish “general” weighting factors. The specific problem and the application context should always be considered, and the target area and time should be defined. It is also questionable whether it is possible to base valuation purely on environmental grounds. On the other hand, information on economic issues is also lacking.

The demands set on the qualities of a valuation method or procedure also depend on the decision-making situation, and, for example, on whether the results are meant for internal use, for instance, in product development or for external use, such as marketing purposes. The development and use of models are necessary for the analysis and structuring of decision problems. Anyway, models do not assume responsibility for the final decisions; it must be borne by someone, and it should be clear by whom.

The basis selected for the index statistics significantly affects these statistics and hence also the resulting ratings, which are arithmetic means of the indexes given by individual judges. This basis can be selected in many ways, each justified by a postulate stating some kind of conformity between the judges' thinking about the dimensions and magnitudes of environmental harms. The problem of the basis selection is how to bring the obviously different conceptions of the judges about environmental harms into a common scale.

The final indexes of the Delphi I study were computed on the 'total of all interventions' basis. Compared to the 'total of NO_x' base, which was first used, the dispersion of many indexes was substantially reduced. For instance, a reduction of over 40% occurred in green house gas emissions, carbon monoxide, VOCs, sulphur dioxide emissions, water releases of oils and greases, and oil as a resource. However, it was impossible to estimate how much the convergence depended on the possibly more correct index basis and how much on the mechanical effects of the standardisation procedure. The effects on the index values were also big. For example, indexes of oil in ground (resource), methane (air), carbon monoxide (air), carbon dioxide (air) and nitrous oxide (air) were reduced by over 75%. The whole index profile differed essentially from the one produced by the judges with nitrogen oxide emissions as a reference.

Even in the final statistics the dispersion of the indexes was wide. The coefficient of dispersion (standard deviation per mean) varied from 0.58 in the best case to 1.78 in the worst case. Thus the factual uncertainty of the majority of indexes is high.

K-entropy analysis showed that the main increase in the consensus on the ranking of the interventions occurred in the first iteration round. The second iteration could not essentially increase consensus. The largest entropy reduction in the first iteration round occurred in nitrogen oxide emissions, over 50%. This finding is not surprising, considering that NO_x emissions were used as a reference, with which all other interventions were to be compared. A more surprising finding was that out of the top ten interventions in the entropy decline order, six interventions appeared in the top ten of the intervention list provided for the first iteration round. This list was in the order of the rankings from the first round, and these ranks also accompanied the interventions on the list. This indicates that the communication technique had worked quite efficiently for consensus search in the first iteration round. The fact that the convergence of the top interventions was about 40% stronger than that of the rest of the interventions suggests that both the feedback from the initial round and the way (rank order) in which it was communicated to the judges have affected the judges' decisions.

The differences between the main intervention categories, namely air emissions, water discharges and resources, were small both in *K*-entropies and in their changes from round to round. The resource category had a slightly higher entropy in all rounds than air and water emissions; the latter two were practically on the same entropy level throughout the process. This result is in line with the present situation in LCA valuation, where the valuation of resources, especially fossil energy resources, is a disputed issue and common agreement on suitable valuation methods seems to be quite far away.

As a whole, *K*-entropy was reduced by about 17% during the process, which is not much considering that the initial level was quite high, relative *K*-entropy 0.843. According to the selected classification, for ten interventions the final consensus was rather strong or stronger, and for thirteen interventions rather weak or weaker. For seven interventions the consensus was very weak. The interventions about which the judges were most unanimous were nitrogen oxide emissions (very strong), and sulphur dioxide emissions and suspended solids (both strong). The emissions for which hardly any consensus was found (very weak) were methane, nitrous oxide, and particulate emissions to air, BOD+COD, oils and greases released to water, oil in ground and soil erosion.

13.2 Development issues

A development issue suggested for improving the transparency of weightings between the various environmental interventions was to include an argumentation procedure. this could be accomplished by creating a databank, into which arguments and facts about the particular environmental intervention and impact could be collected. It could include, for example, facts concerning the management of the problem or the possibilities of compensating for the impacts. (Kuusi 1997–1998)

In Delphi I the panel had to weight 23 different interventions, which was probably too many. A judge should only consider the interventions with which s/he is sufficiently familiar. One possibility for task formulation would be the valuation of intervention classes (comprised of the interventions that cause a certain impact type) instead of interventions. In the weighting between the different intervention classes, environmental "generalists" could be used. The weighting of interventions within the impact classes could be accomplished by more specialised experts for each category (Wilson 1997–1998).

There are two ways of choosing experts (or panellists) for a valuation task: comprehensively, choosing the panellists with relevant attributes concerning knowledge and world view, or choosing certain panellists and describing their qualifications.

One possibility would be to set up of a register of experts in the geographical area to be considered. The experts could be taken from universities, consultancies, industry, learned institutions, etc. The register could contain information of the expert's qualifications, familiarity with interventions and world view. This register would allow the selection and weighting of a panel or panels to perform the valuation procedure, using either the Delphi technique as the communication system or some other device. (Wilson 1997–1998).

The typology on "views on nature" based on cultural theory was experimented in the Delphi II study and the results seemed to be able to explain the weightings to some extent. This kind of typology could be perhaps further developed and utilised in describing the quality of the panellists.

The study showed that transparency and certainty, which are essential qualities for an acceptable and trusted valuation method, are only partially accomplished by the expert judgement method in the format in which it was developed in the analysed case. As for the technical procedure, the method is well documented and transparency is good. Argumentation of the judgements, however, should be increased. The quality of the valuation indexes is explicitly available, but their certainty is very low for most interventions. The opinions of the experts differ greatly. How much this depends on different values and how much on differences in knowledge etc. is impossible to assess. Also, how much the technique used and the statistical processing of the experts' answers may have influenced the eventual scores of different interventions is difficult to assess. The structure of the valuation process and the statistical methods should be standardised to be able to reach reliable and comparable results.

The application of expert judgement to LCA valuation is a new idea, and the method is still very much under development and far from maturity. Nevertheless, utilisation of expert knowledge can be a significant addition to model approaches to ecological impact assessment, which, because of the chaotic behaviour of ecosystems, are limited and uncertain in predicting the ecological consequences of interventions to the environment. This should be taken into account when considering the results of the evaluation of the case study, which was the third of its kind in Europe.

References

Ahbe, S., Braunschweig, A. and Muller-Wenk, R. 1990. Methodik für Ökobilanzen auf der Basis Ökologischer Optimierung. Schriftenreihe Umwelt Nr. 133. BUW AL, Bern, Switzerland.

Annema, J. 1992. Methodology for the evaluation of potential action to reduce the environmental impact of chemical substances. In SETAC-Europe: Life-Cycle Assessment. SETAC-Europe, Brussels, Belgium. Pp. 73–80.

Braunschweig et al. 1994. Evaluation und Weiternetwicklung von Bewertungsmethoden für Ökobilanzen – Erste Ergebnisse. IWÖ – HSG. St. Gallen.

Calabrese, E.J. and Kenyon, E.M. 1991. Air toxics and risk assessment. Toxicology and environmental health series. Chelsea, Mich., Lewis. 662 p. ISBN 0-87371-165-3.

Christiansen, K., Grove, A., Hansen, L.E., Hoffman, L., Jensen, A.A., Pommer, K. and Schmidt, A. 1990. Miljøvurdering av PVC og udvalgte alternative materialer. Miljøprojekt 131, Miljøstyrelsen, Copenhagen, Denmark.

Costanza, R. et al. 1997. The value of the world’s ecosystem services and natural capital. Nature, Vol. 387, 15 May 1997.

Douglas, M. and Wildavsky, A. 1982. Risk and culture: an essay on the selection of technical and environmental dangers. Berkeley, CA University of California press cop. 1982.

Finnveden, G. 1996. Valuation Methods within the Framework of Life Cycle Assessment.

Finnveden, G. and Lindfors, L.-G. 1997. Life-cycle Impact Assessment and Interpretation. LCANET Theme report. Stockholm, 1997.

Goedkoop, M., 1995. The Eco-indicator 95. PRé Consultants, Amersfoort (NL).

Graedel, T.E., Allenby, B.R. and Comrie, P.R. 1995. Matrix Approaches to Abridged Life Cycle Assessment. Environmental Science and Technology, 29, pp. 134A–139A.

Guinée, J. 1994. Review of Classification and Characterisation Methodologies. In Udo de Haes, H.A., Schaltegger, S. and Hofstetter, P. (Eds.). First Working Document on Life-Cycle Impact Assessment Methodology. SETAC-Europe Workshop Report, ETH, Zurich, Switzerland. Pp. 51–60.

ISO 1997a. Environmental management – Life cycle assessment – Principles and framework (ISO 14040:1997).

ISO 1997b. Environmental management – Life cycle assessment – Life cycle impact assessment (ISO/CD 14042.3:1998).

Kalisvaart, S.H. and Remmerswaal, J.A.M. 1994. The MET-points method: A new single figure environmental performance indicator. In Udo de Haes, H.A., Schaltegger, S. and Hofstetter, P. (Eds.): First Working Document on Life-Cycle Impact Assessment Methodology. SETAC-Europe Workshop Report, ETH, Zurich, Switzerland. Pp. 143–148.

Kortman, J.G.M., Lindeijer, E.W., Sas, H. and Sprengers, M. 1994. Towards a single indicator for emissions – an exercise in aggregating environmental effects. Interfaculty Department of Environmental Sciences, University of Amsterdam, Amsterdam, The Netherlands.

Kuusi O. 1997–1998. VATT Economical Research Centre of Finland. Participation in the Delphi II.

Linfors, L.-G., Christiansen, K., Hoffman, L., Virtanen, Y., Junttila, V., Hanssen, O.-J., Rönning, A., Ekvall, T. and Finnveden, G. 1995. Nordic Guidelines on Life-Cycle Assessment. Nord 1995:20, Nordic Council of Ministers, Copenhagen, Denmark.

Paasivirta, J. 1998. University of Jyväskylä. Personal communication.

Pearce, D.W. 1993. Economic Values and the Natural World. Earthscan Publishers, London, UK.

Puolamaa et al. 1996. Index of Environmental Friendliness. A Methodological Study. Statistics Finland. Helsinki.

Schmitz, S., Oels, H.-J., Tiedemann, A. et al. 1994. Eco-balance for Drink Packaging. Comparative Investigation of the Environmental Effects of the Various Packaging Systems for Fresh Milk and Beer. German Federal Environmental Office, III 2.5. Revised Edition, June 7.

Schuster, H.G. 1989. Deterministic Chaos. An introduction. VCH Verlagsgesellschaft, Weinheim, Germany.

Schwarz, M. and Thompson, M. 1990. Divided We Stand: Redefining Politics, Technology and Social Choice. New York Harvester Wheatsheaf.

Seppälä, J. 1997. Decision analysis as a tool for life cycle impact assessment. Finnish Environment Institute. The Finnish Environment 123.

Shannon, C.E. 1948. A mathematical theory of communications. Bell Syst. Tech. J., Vol. 27, pp. 379–623.

Stahl, B, Walx, R. and Böhm, E. 1997. Anleitung zur Bevertung in Ökobilanzen. Fraunhofer Institut für Innovationsforschung. UmweltWirtschaftsForum, 1997. Springer Verlag.

Steen, B. and Ryding, S.-O. 1992. The EPS enviro-accounting method. An application of environmental accounting principles for evaluation and valuation of environmental impact in product design. IVL Report B 1080. IVL, Göteborg, Sweden.

Turner, R.K., Pearce, D., Bateman, I. 1994. Environmental economics, an elementary introduction. Harvester Wheatsheaf.

Udo de Haes, H. A. 1996. Towards a Methodology for Life Cycle Impact Assessment. SETAC Europe. Brussels.

Wilson, R. 1997–1998. Landbank Environmental Research and Consulting. Consultancy for the Delphi II.

Wilson, R. and Jones, B. 1996. Evaluating Environmental Interventions in Finland, Sweden and Norway. Landbank Environmental Research and Consulting, London, UK.

Wilson, R. and Jones, B. 1994. The Phosphate Report. Landbank Environmental Research and Consulting, London, UK.

Wilson, R. and Jones, B. 1995. The Swedish Phosphate Report. Landbank Environmental Research and Consulting, London, UK.

WRI. 1997. World Resources 1996–97, World Resource Institute

Additional literature on the subject

Benarie, M. 1988. Delphi and Delphi-like Approaches with Special Regard to Environmental Standard Setting. Technological forecasting and social change 33. Elsevier Science Publishing Co. Inc. Pp. 149–158.

Bengtsson, M. 1998. Värderingsmetoder i LCA. Centrum för produktrelaterad miljöanalys. Rapport 1998:1.

Bilitewski et al. 1998. CHAINET Definition Document Draft. European Network on Chain Analysis for Environmental Decision Support.

Curley, S., Browne, G. and Benson, G. 1995. Arguments in the Practical Reasoning Underlying Constructed Probability Responses. Journal of Behavioral Decision Making, Vol. 8, pp. 1–20.

Fava, J., Consoli, F., Denison, K., Mohin, T. and Vigon, B. 1993. A Conceptual Framework for Life-Cycle Impact Assessment. SETAC. Pensacola, Florida, USA.

Holland, A. 1997. The foundations of environmental decision making 2. Int. J. Environment and Pollution, Vol. 7, No. 4 (Special Issue), pp. 483–496.

Lindeijer, 1997. Results try-out Japanese/Dutch LCA valuation questionnaire 1996. IVAM Environmental Research.

Linstone, H. and Murray, T. 1975. The Delphi Method, Techniques and Applications. Addison-Wesley Publishing Company.

Marttunen and Hämäläinen 1994. Päätösanalyysihaastattelu ympäristövaikutusten arviointi-menettelyssä. Vesitalous 3/1994.

Matala, H. 1986. Immateriaaliset arvot monitavoitteisessa päätöksenteossa [Intangible values in multicriteria decisionmaking]. VTT Research Notes 413. Espoo. 143 p. + app. 2 p. ISBN 951-38-2614-7.

Miettinen, P. and Hämäläinen, R. 1996. How to benefit from Decision Analysis in Environmental Life Cycle Assessment. European Journal of Operational Research, Vol. 102. Elsevier Science. Pp. 279–294.

Määttä, Y. 1996. Ympäristökysymysten rationaalinen arvottaminen: analyysi ihmisten ympäristötietoisuuden yhteiskunnallisuudesta. Helsingin yliopisto 1996, Helsinki.

Powell, J., Pearce, D. and Brisson, I. 1995 Valuation for Life-Cycle Assessment of Waste Management Options. CSERGE, School of Environmental Sciencies, University of East Anglia, Norwich.

Rieke, R. and Sillars, M.O. 1984. Argumentation and the Decision Making Process. Second Edition. Scott, Foresman and Company. Glenview Illinois.

Sackman, H. 1975. Delphi Critique. Expert Opinion, Forecasting, and Group Process. Lexington Books.

Stahl, B., Waltz, R. and Böhm, E., 1997. Anleitug zur Bewertung in Ökobilanzen. Umwelt Wirtschafts-Forum, Vol. 5, Juni 1997. Springer-Verlag. Pp. 83–88.

Timmermans 1993. The Impact of Task Complexity on Information Use in Multi-attribute Decision Making. Journal of Behavioral Decision Making, Vol. 6, pp. 95–111.

Varis 1989. The Analysis of Preferences in Complex Environmental Judgements – A Focus on the Analytic Hierarchy Process. Journal of Environmental Management, Vol. 28. Academic Press Limited. Pp. 283–294.

Volkwein et al. 1996. The Valuation Step Within LCA, Part II: A Formalized Method of Priorisation by Expert Panels. International Journal of LCA, Vol. 4, pp. 182–192.

Winterfeldt, D. and Edwards, W. 1986. Decision Analysis and Behavioral Research. Cambridge University Press, Cambridge.

List of statements for use with Question 3

A Has considerably damaged environmental objects (water bodies, forests, air bodies) in the past in Nordic regions (*Please give examples*).

B Is presently damaging environmental objects in Nordic countries (*Please give examples*).

C Its level is approaching the tolerable limit in Nordic countries, and will become a problem in the near future.

D Causes long-lasting effects and may lead to environmental problems in Nordic countries in the longer term.

E Has recently been recognised as a cause of environmental problems in Nordic countries.

F Is the major source of a particular environmental problem in Nordic countries.

G Is a minor source of a particular environmental problem in Nordic countries.

H Is under control in Nordic countries.

I Causes important effects according to recent findings.

J Has been receiving much publicity (recently?).

K Causes impacts whose consequences are severe and not tolerable.

L Causes impacts which are tolerable.

M Would be very expensive to reduce.

N Causes irreversible damage.

O Would lead to damage which is very expensive to restore.

P Causes damage which has only minor cost.

Q Has global impacts.

R Affects large areas.

S Affects the environment outside the Nordic countries.

T May endanger the environment of future generations.

APPENDIX B

Comments of the experts to the questionnaire

1.a The task was interesting

- *The task allowed one to learn different options on difficult issues and also how to reach consensus on those issues.*
- *The task was also useful, because it made think the different environmental impacts, to acquire some more information on them etc.*
- *Also new information was distributed, for ex. PM10, had recently appeared as a problem*
- *The time consumed for the process was about 2 days and additionally time for thinking. The first round was the most difficult, the others were easier.*

1.b It was difficult to understand the tasks

- *Difficult to keep only 2 interventions in mind at a time.*
- *In the beginning the formulation was difficult to understand, perhaps because the approach was so new.*
- *The instructions of the first questionnaire were not quite clear*
- *The task description in the formula could also be developed.*
- *The visit in the course of the process was good. The visits were too late…*

1.c The background information provided for the evaluation was adequate

- *In the background material maybe some preferences to D-method…*
- *The numbers provided adequate information.*
- *It was adequate, because giving of more information could have impacted the results of evaluation*
- *The task description in the formula first sent was too long, which made it more difficult to understand. It could have been written more short and clear.*
- *The visits by interviewer were too late to help the understanding of tasks.*

1.d The process changed my view about the harmfulness of some environmental interventions

- *The information on the total emission amounts affected.*
- *Affected radically the evaluation.*
- *I had to find more information about the harmfulness of some interventions, which were not so familiar for me, and it changed my views in some degree. Also the process and the comments made by the others, made me re-evaluate my thoughts or some of them. In one's own work the one concentrates mainly on the specific harms, now also other harms came into consideration.*
- *Had been self thinking much of valuation … Ready answers concerning philosophy etc.., estimating and valuing damages helps to understand „what is big and what is small".*

1.e The process made me change my evaluation of some environmental interventions in the iteration

- *In the course of the process the weighting changed, the initial weight for CO2, very strongly weighted, had been right.*

- *Others had not been thinking the whole so much, „I was stable”.*

1.f The final coefficients obtained are generally suitable for valuation in LCA applications in Nordic countries

- *The results are too uncertain, that they could be applied more widely or generally.*
- *I think, that the coefficients are not suitable, if other applications than energy production are in question. For example, construction applications are very different from energy use. It is not possible to have general coefficients for all applications.*
- *The emissions and the impacts have to be on hand, and they have to be considered in each valuation situation. The transparent valuation method could be for help*
- *A method that would take into consideration different other impacts…also other than impacts to the environment should be connected.*
- *The views about the harmfulness of the interventions are based on something, which has been adopted earlier. It is impossible to base only on environmental factors*
- *This mainly adds information of attitudes of certain social groups.*

1.g The list of interventions should be expanded (Please suggest additions if you agree)

- *The list needed depends on the application, for ex. damages for landscape may be important in some cases.*
- *Resource use is application specific and there are several other resources, that should be included.*
- *The interventions were heavily stressed in the list vs. the questions to aesthetic issues. This kind of list may well be manipulative…*
- *In a task like this the intervention list should rather be more limited, also, if there would be less interventions, it would be possible to compare them this way, directly.*
- *The matrix becomes too complex if you increase the number of variables.*
- *Various types of land use and impacts on water resources should be added. …perhaps take away some others.*
- *A list of questions concerning food or energy, the issues which are easy to understand.*

1.h The final overall ratings of heating oil from North Sea oil and rape seed oil were quite reasonable

- *The result was dominated by few factors, trivial, it would be more interesting to compare more uncertain issues, for example nuclear power vs. oil etc. or natural gas vs. oil, „traffic issues”, also, where strong affective factors involved.*
- *When comparing rape seed oil and fossil oil impacts on human health and life during the life-cycle should be considered, for ex. the deaths at mines etc. should be taken into account. In agricultural production there is no dramatic…*
- *It depends whose environment you want to protect, future generations…?*

1.i I would be prepared to participate in a similar process in a few years time to update the ratings

- *Generally yes.*

3.a CO2

- *Impacts to Nordic circumstances, not much importance*

- *Potential for many deleterious effects, agreement that greenhouse effect due to human activities.*
- *Relates to the core of the industrial society – energy production. Although one is not advocating nuclear power, on the base of what is known, the risks relating to nuclear power are minor than those relating to fossil…*
- *The warming of ocean for 3–4 km, dynamic cycle is ceasing…oxygenation is ceasing, life will be destroyed, seas will become passive…*
- *The effects of others than CO2 are minor*
- *Big distribution in weighting, it was astonishing that some experts valued it so negligible*
- *An issue that divides. Not necessarily depend on publicity.*
- *Even though there is much uncertainty about the actual impacts of CO2, we should not take the risk, because the impacts may be large, irreversible, and they may severely affect the living conditions of future generations*

3.b NO_x

- *Respiratory effects in general urban population.*
- *Acidification … leaching of Al to lakes*
- *Impacts on acidification, eutrophication and tropospheric ozone production, which are all problems in Nordic countries*
- *Formation of oxidants, affects N_2O production and aerosol formation.*
- *Extra deposition of NO_x- also positive effects – fish production, wood prod – CO_2 binding. Problem only in limited areas of Baltic Sea…All do not want to take the positive impacts into account.*

3.c SO2

- *Respiratory effects in general urban population.*
- *Is believed to have some negative effect on forest health and some regional soil lake qualities.*
- *As NOx*
- *Is almost under control in Nordic countries. The emissions, coming outside Nordic countries, may be a problem in some areas.*
- *A problem of regional scale, extinction of species, climate impact trough aerosol formation*

3.d VOC

- *Minor problems*

3.e Benzene

- *Minor problems*

3.f PM10

- *Respiratory effects in general urban population.*
- *Health concern, increasing mortality to cities, 1–5 % excess mortality: local problem*
- *Maybe it should be PM2.5 though the measured value has been PM10 until these days.*

3.g Heavy metals to air

- *Minor problem, local*

3.h Heavy metals to water

- *Minor problem in terms of direct toxicity, causes exposure trough food*

3.i Nitrogen to water

- *Release of N from fertilisers to groundwater, a potentially large problem locally regionally*
- *N emission, when there is lot, may cause P, K, and Mg deficiency… damage for forests, acidification of forest soil, same with ammonia also… risks are big*
- *Oxygen deficiency in bottom waters, is a major source*
- *Eutrophication of lakes*

3.j Oil in ground as a resource

- *May affect living conditions of future generations*

3.k Other comments

- *The statement L had value content, responsibility for other people.*
- *„People for one issue", for example PM10 causes 1–5% extra risk, compare with car accidents, smoking, about which there is certain information, 90% from lung cancer etc…*
- *CO2 was the only global, lot of local, regional…*
- *The majority of the issues are anyhow small things in relation to health.*
- *NOx, SO2 and CO2 important, about others it is difficult to say, more local…*
- *It was difficult to classify the interventions under the statements in the list, because most of the impacts are something between severe and tolerable. Statements in the questionnaire were too extreme. There should be the possibility for somewhere between..*

Feedback of the experts in the interviews

Possible causes for the controversiality of the answers of the experts ?

- *The complexity of the formula*
- *Knowledge on the issue, if enough knowledge, it could make the issue less important or other way round, could make it more important*
- *The points given by separate experts did not differ so much from each others*
- *People from different backgrounds, importance of choosing people to get some balance…*
- *On the list there were interventions, which are known to higher or to lower degree. Many impacts are still not known, unclear etc. For ex. in the case of CO2 it is easy to have different views, uncertainty about the impacts etc.*
- *Haven't done this type of exercise earlier. Many scientists does not consider resource use. Work in narrow specialised field.*

How did the experts perceive the environmental impacts during the Delphi-process (for ex. on emission level or on damage level) ?

- *The impacts were considered in the level in which the people are exposed…how many people are involved. For ex. Many emissions from the traffic, 80% of Finnish people are exposed, or pesticides, mainly only farmers…*
- *Weighting task was too difficult, comparing the decreasing of one emission to an other, impossible to understand…Putting the impact categories to an order would be ok.*
- *Intervention categories (emissions affecting acidification, eutrophication, etc.) could be a good idea, weighting between the classes and within the classes.*
- *Structuring the problem according to impact categories*
- *The experts have been considering widely in their minds, how the traffic for example…*
- *Category by category, the global and local impacts should be separated… separately in short or long view ? National problems and global problems, in the decision making there is own logic for both…*
- *Also estimation of the costs of the actions should be included in the interview (for ex. 1 000 mk …10 000 000mk per tn). In the decision making the view on costs is important.*
- *The view of serving the decision making process; the present production-energy production structure, what problems it is producing what kind of damages are being caused, view on the costs*
- *Only the experts views or also the relation to the real decision making ?*
- *Level is on the impacts, for example acidification. Al is gill toxic, kills fishes for oxygen depletion. The more there is acidification, the bigger the effects…*
- *The interventions are important, the outlining the processes, one thing has many impacts… where it is happening, which or who are exposed…*
- *When making the valuation, one had the final harm in mind. The emission quantity has importance on the impact, harm caused. If the impact considerably minor or extensive, affected the valuation. Or if the impact harmful in Nordic Countries etc.*
- *It was good that the task was in intervention level, because there were numbers available. The impact categories comprise more uncertainty, depend on the place, impacts in several levels (plants, animals etc.).*
- *When making evaluation one has to think about the endpoints…"models in mind" conceptualisation of the problem… Move to the endpoints ? Not possible to value as emissions either…*

General aspects on valuation of environmental impacts ?

- *It is assumed, that the experts have enough knowledge, no arguments are needed.*
- *Delphi is a good technique for mapping the experts views in orderly form. But no „truth" is produced.*
- *This was a way of quantifying the comparison of rape seed oil and conventional oil*
- *There are differences in scientific facts about CO2 or soil erosion…*
- *Both the own knowledge base and values affects the weightings, for example valuing the aesthetic damages in relation to health damages.*
- *Valuation in the Delphi-process is representing a view of a certain expert group, is restricted to that ! May be used, when valuation is needed, restricted use. Danger for misuse, in marketing for consumers.*
- *Comparable methods, Saaty etc… The destination is not seen, not transparent at all. Delphi is between the Saaty method and a transparent one…*
- *More argumenting would need more work*
- *Real knowledge vs. assumption, only few has knowledge on every intervention.*
- *In analysing the results, it would be possible to weight with the quality of expertise and see how it affects the distribution of answers.*
- *Different experts… The one who makes the study or the experts themselves could evaluate the quality of expertise.*
- *Expertise vs. humanity*
- *Also impacts related, on which there is no information available*
- *Why the questions were laid like this, was the logic ok ?*
- *The foresight studies have also been criticised for not forecasting any real events… The Delphis believe in themselves, will carry out themselves, are committed to certain destination*
- *Questions by which the meaningless of the panel can be tested, for example „ethical distribution",*
- *For example, 10 experts from different areas: water- air experts and „affective" panel*
- *Valuing of environmental problems is not really possible discreetly, because there is not enough knowledge*
- *An „ideal view" may be reached with Delphi, it should be made transparent, how it was reached, also the panel was „like this"*
- *Valuing on monetary bases is even worse. There is a principle problem in that „everything is for sale"*
- *Valuing in monetary terms which is possible and leaving the other things separately.*
- *23 interventions, means already many things in mind.*
- *Who decides, where the comparison begins, what have been left out and why*
- *the industry: „there is not enough knowledge, that something could be done"… It is not a question of, that the industry would not know (an example of cigarettes). The decisions have to be made at some stage.*
- *Well being is also created by industrial activity…*
- *Result, a political consensus, such is the way in a society.*
- *Everyone has to participate into valuing environmental impacts who participate in producing them, industry, authorities, scientists, consumers…*
- *The issues discussed in the follow-up interview were not all considered in answering the Delphi questions. Would have demand more profound examination. Might have affected in the background…*
- *All the parties have legitimate rights to the views, which base on some concrete benefits (everything is based on that…). For example from the energy producer point of view the*

most important is to create profit, but on the longer view, the customers views are important.

- *When there are lot of people involved it is possible to get good results, consensus.*
- *It is difficult to have an opinion, when there is not enough information …. Results were good, on the base of present information … may change in future*
- *It is possible to steer the results…experts on heavy metals keep them important etc…*
- *In the case of most important emissions there is agreement…*
- *Delphi: important to find right people involved (people) in research, general people are more open to press etc…*
- *Valuation is difficult with help of any method.*
- *A method based on expertise is better than the one based on money; those also comprise very much uncertainty.*
- *Economical valuation – if it would be possible it would be good…*
- *Move questions to concern endpoints… Comprises forecasting future. Also future generations could be taken into account with endpoint analysis*
- *Expert's rules … as God's word ? „dangerous"*
- *Small things are small, big are big…10 fishes vs. 300 000 people starving in Africa…. For ex. poisonous substances…potential effects, but maybe do not hurt anybody anyway…*
- *„Results reflect what the scientist think"*
- *When solving environmental issues, trying to solve other problems same time. Government: Sustainability + employment etc…*
- *LCA to get comprehensive view… should follow everything that happens, (not only expand technical system…). You have to handle complexity etc (uncertainty etc) someway. LCA as a way of organising present knowledge*
- *Ask also from lay people in the level of „endpoints"*
- *Ranking different endpoints…differs according to culture.*
- *It affects, when some issue is in publicity.*

Economical considerations ?

- *Not considered the economical issues really, but it was not possible to be totally „free" of them*
- *Anyhow, for example rape seed oil is an economical issue for large extent*
- *Not even knowledge about the money, so that could consider the amounts. Energy is a big issue…*

1. Opinions of the interest groups

Workshop 28.8.1998 Valuation and decision making in the LCA context

Pekka Heikura, reporter, environemental issues, Finnish Association for Nature Conservation
Eva Heiskanen, Helsinki School of Economics and Business Administration
Antero Honkasalo, Ministry of environment (chairman, morning session)
Jarkko Hukkanen, UPM Ltd
Raimo Hämäläinen, Helsinki University of Technology
Pekka Järvinen, IVO Ltd (chairman, afternoon session)
Osmo Kuusi, VATT Economical Research Centre of Finland
Sari Kuvaja, environmental consultant
Torsti Loikkanen, VTT Industrial Environmental Economics
Helena Mälkki, VTT Industrial Environmental Economics
Jyri Seppälä, SYKE Finnish Environment Institute
Osmo Soininvaara, Finnish Parliament, The Green League of Finland
Sirpa Torkkeli, VTT Industrial Environmental Economics
Leena Vilkka, University of Helsinki
Yrjö Virtanen, Sirpa Torkkeli, VTT
Risto Willamo, University of Helsinki

1.1 Starting point

In the economic activity and decision making the balance between environment, technology and economics is an important aim, but social view points and ethics should also be included.

Environmental issues are important to different interest groups for different reasons. Anyhow, evaluation of the importance of environment is a complex and difficult problem, and the primary research of the environmental impacts is still very little. Accordingly, knowledge and understanding of the environmental processes may change essentially in the passage of time. The nature of the impact functions is very complex – anyhow, for the time being, for example, models with critical loads exist just for few interventions.

Problems with assessing environmental impacts rise, among others from the following factors:
- extension of effects in time and space
- missing knowledge of the impact chains
- uneven geographical distribution of the problems
- contradictions between the interests of individuals and communities in the case of environmental problems

At the moment there is no appropriate method for assessing the importance of various interventions to the environment. Partly as a consequence of this, but also for traditional ideological reasons, valuing the ecosystem services and natural capital is hardly ever included in economic thinking. (Constanza et al., 1998)

1.2 Environmental issues in different decision making situations

1.2.1 Administrative, political view

In the Finnish environmental administration new needs for valuing environmental issues have recently increased, for example, by the IPCC directive. The environment should be handled as a whole. The authorities have to able to assess the importance of the separate environmental impacts, to take into consideration the local conditions, etc. Until now, decisions have usually been concerned with decreasing one emission at a time. Now it has been recognised that the problems are more complex.

Results gained with an LCA tool may be complicated, but there are also examples, as the previous LCA study of Finnish Beverage Packaging Systems (Mälkki et al., 1995), that the assessment between the alternatives under comparison could be done with help of the inventory results. Developing one method for valuation is not seen possible, since "the values cannot be commensurated", but additional means are needed. In valuation, transparency is very important.

LCA studies on total sectors will help producing overall views for the environmental political decision making. Monetary valuation methods may be helpful in producing information on the magnitude of the environmental impacts.

1.2.2 Industry

Driving forces for the enterprise (Peattle and Charter, 1997) comprise the attitudes of the olders, the competitors, the impact of the people inside the enterprise, the knowledge on the problems existing, the examples of the environmental accidents etc… A relevant scope for consideration would be the management of the whole value chain, system studies.

Situations where environmental information is needed for the decision making in the companies may range from taking environmental issues into consideration as solely external factor till really trying to understand and learn about the problems. For the latter, it has been recognised that a systematic information collection with an LCA tool is useful.

Impact assessment and valuation are considered undeveloped for the time being. There are too many methods and different methods end up with very different conclusions. Simple and clear methods would be needed. When starting the impact assessment, firstly the local real impacts should be considered.

1.2.3 Ecological view

From an ecological point of view it seems that the product centred way of living leads to eco-catastrophe. Solving the environmental problems would require a great change in the way of living. Anyhow, people are not perceiving the real conditions of the state of the earth. It is thought possible to treat nature "separately"… but it cannot be avoided that what is done to nature is done also to humans themselves.

Considering environmental impacts is dominated by physical and chemical concerns. The real problem is not the chemical destruction of the environment, but the mechanical destruction – extinction of species and land use. How could issues like changes in biodiversity or in landscape, treatment of animals, noise or the impact of human activities to floods and even to earthquakes be included in LCAs?

Nature's values should be taken into account in the decisions, at least, because the conditions of human life is based on the vitality of nature. Furthermore, nature may also be considered valuable as such.

1.2.4 Consumers, civic organisations' view

A normal consumer is learning to be environmentally conscious and responsible with the products offered, but has difficulties with the large diverse information flow on difficult issues. The lay people may easily be misled with too optimistic presentations of information about "comprehensive assessments", for example, in connection with the environmental labels etc. Environmental statements are perhaps the most easy to take as "truth" – trying to find the justifications for them would be too difficult. However, measuring everything in money may be questionable. For individuals, the value of money and willingness to pay depends on many things, like ability to pay etc. Relying on expert judgements may also lead to one-sighted solutions. A problem with valuation methods may be that the end result is not transparent.

The role of the environmental organisations is to ask questions and strive for paying attention to particular environmental problems. Otherwise, the complexity of the problems would paralyse…

Factual grounds are required for environmental constraints and norms, for example, for free trade organisations.

1.2.5 Needs of the users for a valuation method

The significance of the environmental issues in decision making situations has various levels, depending on the case. One may only want to take environmental issues into account in some less important meaning, in which case for example a group of experts may be used to assess the differences in the environmental issues between the alternatives. Or one may want to participate oneself in valuing of various impacts or understand and learn more deeply about the problems.

The demands for the transparency, intelligibility and acceptability of the method depend on whether the assessment results are meant for the internal or external use. For instance, assessing the environmental friendliness of a product from a company view, when the information is used for backing up the product development, and for seeking continuous improvement poses different requirements for a valuation method from assessing impacts of the product for an environmental label, in which case the criteria must be simple and measurable.

Decision makers need simple tools which help to structure the decision problem, clarify the alternatives etc… With help of such tools it should be possible to give transparent and good quality information. But, would it be possible to develop a generalised model, which could be applied in various situations?

1.3 Acceptance of the valuing methods and tools in environmental decision making

1.3.1 Generally

Considering environmental impacts, issues

In the studies of environmental impacts, especially also in LCAs the physical-chemical view is presently highly dominating. For instance, landscape and biodiversity issues are normally neglected. Measurability means “important”. It may be asked, whether engineering knowledge is the only way to consider what is important?

Also political issues, and not necessarily only the immediate environmental optimum, may be important. For ex. recycling may be seen worth promoting because of it may teach people benign behaviour. In valuation, the best available techniques have also to be considered.

Generally, the consideration of total environmental problems is being separated into sectors, for example, car accident belongs to traffic sector. Instead, all the different environmental issues should be handled within the same conceptual system. In the suggested model the changes caused by humans in the ecological systems may be studied at four levels: energy flow, materials cycle, natural processes and systems and ecosystems as whole. (Willamo and Stabmej 1988).

Models and tools

The methods are deficient but the evolution of the methods can still be seen. Since the first Finnish LCA in 80’es assessing environmental costs has clearly improved.

Assumption with the weighting method is that the environmental impacts may be described with one index. But in the Delphi II study it was found that choosing the basic point for the index may significantly affect the end results.

It must be kept in mind that models are tools for the problem solving. There is no one right model a nd no one truth can be produced. The models must be submitted to the problem. A problem may be approached with many ways and methods. It is important to ask what is being solved? For what purpose? Context? In the case of monetary models it should always be explained, for what kind of problem, in what kind of context the model is relevant. The process and understanding in decision making situations is more important in stead of numbers.

The consideration area should always be defined. The damage functions are usually relevant only within a certain area and certain damages are different in different areas. For example, saving clean water is very important in some locations. On the other hand, each kg of CO_2 should be equally damaging, everywhere.

Too rigid truth seeking in valuation methods may be harmful. It could suffice to say, how valuation was done, what was considered. The analysis should be transparently anchored to certain perspective: "with these basis the weight is this…, more data is being gathered, the process will be continued".

There is a danger that the transparency of the valuation process is diminished, when only the end result is being looked at. It must be remembered that the meaning of the method is to help in decision making situation. Someone should bear the responsibility for the decisions.

Although the knowledge would increase how much, the values remain subjective and even irrational. For instance, when considering risks in working conditions the workers may see also positive values in confronting dangerous situations. It has been detected that when splitting the attributes in valuation tasks, i.e. making the criteria more accurate, the weight given for the split attributes increases.

The values included in a tool should be questioned: Is it worthwhile of LCA to take into consideration other than external values related, for example, to products and economy? Internal values, nature's intrinsic values, should also be taken into account in addition to the normal values of people… How to resolve the demands to include various values in decision making situations? At least the values in the tool or in the decision making process could be nominated. An ideal situation in business decisions would be that the external values and internal values would be considered equally.

Different stakeholders bring up different views, and the decision maker decides who he is listening to. Justification and acceptance of the decisions in the society should be taken into account. Representation of the future generations should be secured in some way. Question of who decides should also considered. Globally there is a discrepancy in the environmental protection between the industrialised and development countries…

When basing the decisions on values it must be taken into account that they change over time. For example, the case of water power plants; when most of the plants were built the valuations were different from what they are today.

1.3.2 LCA

In which level the outcomes of LCA studies are to be aggregated? Inventory of the interventions may already be sufficient, at least when the activities compared are not totally different. In the LCA world, the present way is comparing the alternatives on the impact assessment level. Methodological debate on the valuation is missing though.

Valuation and impact assessment are not the only subjective phases in LCA, choices are made during the whole process, for example, in allocation, or when placing the system boundaries.

In the LCA, normally only the disadvantages are considered. Benefits should also be assessed, as well as controlling of residual damages, surprises, accidents… The wholeness should be considered ! LCA is not the only overall view. The "overall" is also subjective, variable…

LCA is in the most cases "engineering". It should, however, be extended, but on the other hand it is already very complex. development trend could also be focusing on the input side.

Taking into account the deficiencies, the consumers should not be given too optimistic picture of comprehensive assessment in eco-labels etc.

1.4 Different approaches to valuation

1.4.1 Monetary valuation

Environmental problems and benefits should be valued in money – it is better to try to develop methods for that than do it only arbitrarily… However, environmental aspects have been considered in economics quite arbitrarily and irrationally… A holistic approach would be needed. For example in the case of wood processing industry the chlorine was a very critical subject, other impacts were forgotten.

A monetary approach is needed for the administrative activities as for the permission practices and for the environmental taxes (the cost of the emission should equal the marginal harm). LCAs cannot be extended far enough, they cannot be "right". But, if everything could be internalised in prices…

When giving monetary value the cause effect relationships are easily forgotten as well as the data deficiencies and initial assumptions - a parallel use with other valuation methods would be recommendable. Use of several methods would recommendable also in the impact assessment.

A problem with monetary methods is that the values of nature may easily be left out, or they are not considered to sufficiently at all. There are costs related to the changes in the environment, which can not be priced enough. Willingness to pay data needed in valuation methods has it's restrictions.

1.4.2 Panel judgements

When panels are used, it is an important issue, how they are chosen. It is a political decision, the citizens, politicians etc. should be included as well. The possibilities are a comprehensive choice of the panellists versus choosing particular panellists and describe their qualifications. In any case, the values and knowledge of the panellists should be assessed…

For example, in the permission proceedings, opinions of citizens have also been included. The experts may be one-sighted ! For example, in the case of nuclear power.

The results of valuation panel are always connected to the time and place – therefore they should be continuously renewed.

Judgements should be provided with arguments. Idea of a databank of arguments was suggested: For example, for nitrogen oxides a description of BAT technology for managing the problem or possibilities for compensating…

It could be assumed that the views in the expert panels would become closer if there would be more basic knowledge available. On the other hand in Delphi I two broad groups could be distinguished: one considering the global climate change the most important (not well understood effect of the future), the other setting the first priority on immediate, visible and locally perceivable effects, such as eutrophication. In the analysis made in Delphi II it could be seen that there is clearly a subjective component in valuation. The experts are of different views and iteration does not eliminate that.

According to the views of the experts gained in Delphi II: "The result is representing a view of a certain expert group", there is "a danger to misuse the results in marketing" and "everyone who contributes in producing environmental impacts has to participate in valuing".

1.4.3 Delphi-technique

According to the Delphi II study, the opinions of the experts differ much from each others. How much this depends on different values and how much on differences in knowledge etc. is impossible to assess. Also, how much the technique used and the statistical handling of the expert answers may have impacted the eventual scores of different interventions is difficult to assess.

The future success of the Delphi-technique depends on abilities to accomplish the following issues acceptably (Kuusi 1998):

- Choosing the experts

The expertise and values of the experts have to be reported. Possibilities are a comprehensive choosing of the panellists with relevant attributes or choosing certain panellists and describe their qualifications.

- Argumentative methods

Complement the answers or numbers given by experts with arguments. Problems with commitment to anonymous argumenting?

- Content of the questions and tasks

There is a danger that the task will be understood wrongly. For example, if it would be asked if concentration of a certain substance is a problem – the answer depends easily on where the question is asked, although, it should not – the concentration as such is a problem.

- Structured discussion, where the competence of topics and arguments are continuously assessed.

The importance, desirability and conditions of the scenarios formed would also be needed.

- An ability to collect a bank of the relevant arguments – arguments, from several and different experts and arguments representing different weights

- Relevance of data produced for the decision making

1.4.4 Conclusions of the workshop

- **Responsibility questions**

Models are necessary to develop and use for analysing and structuring decision problems. Anyway, models do not eliminate the responsibility of the final decisions, it must be born by someone, and it should be clear by whom.

- **Choosing the experts, world views**

There is two ways for choosing experts (or panellists) for valuation task: comprehensively, choosing the panellists with relevant attributes, concerning knowledge and worldview or choosing the certain panellists and describe their qualifications.

- **Context**

The context of a model should always be explained, for what kind of problem, in what kind of context it is relevant. Target area and time should be defined. Valuation weights free from the context are not possible, at least yet.

- **Methods**

Models can sturcture the process and help understanding in decision making situations. It is recommendable to use several models side by side.

- **Learning process**

Understanding the decision problem, the causalities involved is more important than the resulting numbers. Expert knowledge is also a slice of reality.

Literature cited:

Constanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V, Paruelo, J., Raskin, R.G, Sutton, P. and van den Belt, M. (1998) The value of the world's ecosystem services and natural capital. Nature, Macmillan Publishers Ltd.

Kuusi, O. (1998) Delfoi-menetelmä.

Mälkki, H., Hakala, S., Virtanen, Y., Leppänen, A. (1995) Life Cycle Assessment of Environmental Impacts of Finnish Beverage Packaging Systems. The Association of Packaging Technology and Research.

Miettinen, P. and Hämäläinen, R. (1998) Indexes for Fixed and Flexible Environmental Target Setting – A Decision Analytical Perspective. Submitted manuscript 17.2.1998. 20 p.

Miettinen P. and Hämäläinen, R. (1997) How to benefit form decision analysis in environmental life cycle assessment (LCA). European Journal of Operational Research 102 (1997), pp. 279–294.

Vilkka, L. (1997) The Intrinsic Value of Nature. Amsterdam/Atlanta, GA. 168 p.

Vilkka, L. (1998) Ympäristöfilosofia etsii kestävää luontosuhdetta. In: Ekokaari 1/1998.

Vilkka, L. (1996) Ympäristöeettinen päätöksenteko. In: YKL-posti 2/96.

Willamo, R. and Stabmej, V. (1998) Rajat rikki! Ympäristömuutosten arviointi. (YMPS2.2A). Kurssin opintomateriaali. 4. Painos. Ympäristönsuojelun opetusmoniste n:o 25. Helsingin yliopisto. Limnologian ja ympäristönsuojelun laitos. ISSN 1235-2144.

List of environmental categories

CATEGORY	DESCRIPTION
Emissions to air	
Main global warming gases	CO_2, CH_4, N_2O *excluding* CFCs, HCFCs, HFCs, HCs and halons
Ozone depleting gases	CFCs, HCFCs, HFCs, HCs, Halons, methyl bromide
Heavy metals (air)	As, Ba, Cd, Cr^{3+}, Co, Cu, Fe, Pb, Mn, Hg, Mo, Ni, Sn, V, Zn
Hydrocarbons	Unhalogenated/halogenated, aromatic/aliphatic, PAHs *excluding* CH_4
Particulates	Including PM_{10}, soot
Nitrogen compounds (air)	NO_2, NO, NH_3 (*excluding* N_2O)
Other acidification gases	SO_2, HCl, HF
Discharges to water	
Heavy metals (water)	As, Ba, Cd, Cr^{3+}, Co, Cu, Fe, Pb, Mn, Hg, Mo, Ni, Sn, V, Zn
Nutrient discharges to water	N and P in wastewater treatment discharges, and in run-off from agriculture and non-sewered households
Pesticide applications	Insecticides, herbicides, fungicides, disinfectants
Solid wastes	
Hazardous wastes placed in landfills	Regulated chemicals
Non-hazardous wastes placed in landfills	Inert materials, sewage sludges, domestic wastes, industrial non-hazardous wastes, incinerator ashes
Resources	
Use of fossil fuels	Crude oil, coal, lignite, natural gas (Ignore any downstream effects, e.g. combustion products)
Loss of agricultural soil	Soil erosion, salination, desertification
Loss of forest/woodland	Tropical and temperate, original and planted
Other	
Radiation	Routine emissions from nuclear fuel processing, reprocessing and electricity generation
Odours	Odours from industrial, domestic and agricultural processes
Noise	Noise arising from industrial, domestic and agricultural processes (including transport)