LCA Methodology

Measuring Preferences on Environmental Damages in LCIA

Part 1: Cognitive Limits in Panel Surveys

Part 1: Cognitive Limits in Panel Surveys · Part 2 (under review): The Question Format in Panel Surveys

This series of two papers discusses the elicitation of weights for damage categories in LCIA with the aid of panel surveys. The papers focus especially on methodological aspects in panel surveys. **Part 1** discusses potential cognitive limits of the panel members to understand the reference that is weighted. **Part 2** focuses on the influence of the question format and compares results from two different weighting tasks: discrete choice (between alternatives) and score allocation.

Thomas Mettier[1]*, Roland W. Scholz[1] and Olaf Tietje[2]

[1] Swiss Federal Institute of Technology Zurich, Department of Environmental Sciences, Chair of Natural and Social Science Interface (NSSI), ETH Zentrum, 8092 Zurich, Switzerland
[2] Syst[aim], Pfingstweidstr. 31a, 8005 Zürich

* **Corresponding author** (thomas.mettier@env.ethz.ch)

DOI: http://dx.doi.org/10.1065/lca2005.08.218

Abstract

Goal, Scope and Background The weighting of environmental impacts and damages on the safeguard subjects Human Health, Ecosystems, and Resources is a significant step of full aggregated LCIA. Panel surveys have become a common approach in LCIA research to investigate the preferences of stakeholders on environmental impacts and damages. Despite the numerous studies, the knowledge on how to elicit reliable weights is still poor and inconsistent. We present a questionnaire study with 58 environmental science students to investigate so-called framing effects in panel surveys.

Main Features. The study investigates the significance of different framings, which were provided by three references. In addition, the significance of quantitative information provided in the questionnaire is tested. The references are (a1) safeguard subjects without specified additional information, (a2) damages in Europe as they are perceived by the panelist, and (a3) quantified scenarios derived from Eco-indicator99. All participants ranked and rated the importance of the safeguard subjects three times, once within each reference system. According to a test-of-scope study, quantitative information given to the panelist was varied. One level (b1) included data from the Ecoindicator99 methodology, whereas the other group (b2) received data with significantly higher Human Health damages and lower Ecosystem damages, ceteris paribus. This design allows testing the influence of quantitative data on the rating.

Results. The weighting of the safeguard subjects (a1) reveals that Human Health is considered a slightly more important safeguard subject than Ecosystems. However, both are judged to be significantly more important than Resources. This picture changes for the references (a2) and (a3) where damages were weighted. For both references, the respondents rated damages to Ecosystems as most important followed by Resources and Human Health, showing by far the lowest weights. Therefore, the framing of the reference that was weighted played a significant role. The ratings of the subgroups (b1) and (b2) did not differ with respect to the importance of damages, though substantially different quantitative information was given.

Conclusions and Recommendations. The participants of the study were obviously insensitive with respect to quantitative information provided. This raises three questions, which are discussed. What is the mental model upon which respondents base their beliefs and values? Can we expect that 'more sophisticated' subjects would respond differently? Which prerequisites should an empirical weighting procedure fulfill in order to incorporate numerical data? We propose different approaches for future procedures in order to accurately analyze these questions.

Keywords: Environmental damages; framing; life cycle impact assessment (LCIA); panel surveys; test-of-scope; valuation tasks; weighting

Introduction

Life Cycle Impact Assessment (LCIA) – a specific method of goal oriented, functionalistic evaluation – follows a three-step protocol of analytical decomposition. As a first step, the functional unit under investigation (i.e. the products or technological processes) is decomposed into several aspects, perspectives, or criteria relevant for evaluation. Then, each of these aspects, perspectives, or criteria is assessed separately in terms of every option or action alternative. Finally, a concluding synthesis process integrates the decomposed scores of the criteria for each option. According to the ISO 14042 standard (ISO 2000), the decomposition step within LCIA is described as the 'selection of impact categories, category indicators and characterization models'. Separate assessment of each criterion is conducted through the two mandatory steps 'assignment of LCI results to the impact categories (classification)' and 'calculation of category indicator results (characterization)'. These calculations yield a collection of indicator scores (LCIA profile). For the composition step, ISO 14042 suggests two ways of integrating the indicator scores. One way is a direct, intuitive, holistic interpretation based on the LCIA profiles. The second way consists of further aggregation by calculating one or a small set of composite scores for each option by normalizing ('calculating the magnitude of category indicator results relative to reference information'), grouping, and weighting the indicator scores, for example by using a linear model. In the latter method, the interpretation includes the comparison of the composite scores attained for the different options.

Thus, LCIA thinking relies strongly on the decision theoretic framework. From a decision theoretic perspective, the LCIA procedure is a special case of so-called *bootstrapping* (Dawes 1971, Elstein et al. 1978). In this procedure, different experts work on different components of this analytic procedure at different stages in the process. LCIA provides the framework for this and serves as a knowledge integration tool (Scholz et al. 2002), and aids in decision making (Werner et al. 2002) by organizing decomposition, assessment and composition steps. There is still a dispute over which of the two above described methods to use with the composition step. Aggregation is criticized by some for introducing societal values that are neither scientific nor internationally harmonized (Owens 1999, Schmidt et al. 2002). We agree with the argument that values are ubiquitous at all stages of LCIA, (Hofstetter 1998, Hertwich et al. 2000, Werner et al. 2002) and that values are especially at play in the two ways of concluding evaluations mentioned above. From the perspective of psychological decision research, it is recommended that values are systematically taken into account in the composition step (somewhat contradicting the ISO recommendations for comparative assertions disclosed to the public, ISO 2000). This is because – as we know from decision theory – judges tend to intuitively attribute equal importance to every category unless the composition procedure is explicit, conscious, and controlled (Kleindorfer et al. 1993). The process of intuitively attributing equal weights to every category is even encouraged by the ISO standard (for comparative assertions disclosed to the public); it suggests that "the comparison shall be conducted category indicator by category indicator". In addition, graphical representations of LCIA profiles, where bars for different normalized category indicators are on the same plot, intuitively encourage the reader to add up the bars for an overall interpretation. Since there is no scientific rationale for such equality, we favor systematically evaluating the relative importance of the selected categories.

The composition rule for an LCIA profile can follow approaches other than the weighted sum approach, such as the verbal argumentative algorithm developed for the German Umweltbundesamt method (i.e. the German Environmental Protection Agency, (UBA 1995)), the mixing triangle method (Hofstetter et al. 1999), or any other of the various procedures derived from the framework of decision analysis (Seppälä et al. 2002). All of these approaches allow values to be systematically included.

Within LCIA, panel methods have already become a common and seemingly accepted approach to investigating stakeholders' preferences and the weights they assign to environmental categories[1]. Several surveys and panel workshops have been conducted in order to get information about reasonable weights for environmental category indicators (Nagata et al. 1996, Puolamaa et al. 1996, Huppes et al. 1997, Lindeijer 1997, Sangle et al. 1999, Seppälä 1999, Virtanen et al. 1999, Harada et al. 2000, Mettier et al. 2000, Itsubo et al. 2003). We make a distinction between midpoint weighting and endpoint weighting. Most of these studies have been conducted using midpoint indicators; in these approaches, the impact categories are subject to weighting. The studies of Mettier et al. (2000), Harada et al (2000) and Itsubo et al (2003) apply endpoint weighting, as the category indicators presented are expressed as damages to the category endpoints.

[1] For a more detailed discussion of weighting approaches and their advantages and disadvantages, see (Hofstetter 1998, Finnveden 1999, Finnveden et al. 2002).

Despite these numerous studies, knowledge is still limited on how to elicit weights reliably and about which biases, disturbing factors, and cognitive limits to take into account when measuring weights. The lack of knowledge is still particularly strong in terms of how best to represent those impacts or damages being evaluated and how the information provided to panelists, both quantitative and qualitative, affects the weighting.

The main goal of the survey presented is to investigate salient methodological problems associated with panel studies in LCA and to invoke discussion about them. This can improve the weighting techniques and the interpretation of LCIA data, especially for damage-oriented methodologies, such as Eco-indicator 99 (Goedkoop et al. 1999). The Life Cycle Impact Assessment Programme of the UNEP/SETAC Life Cycle Initiative aims at providing guidance to users on how to derive consistent weighting procedures and sets of weighting factors for LCIA results (Jolliet et al. 2003). The methodological questions focused on in that survey may contribute to that aim.

1 The Methodological Problems Focused on in the Survey

One of our concerns is that many respondents participating in weighting panels do not assign appropriately quantified weights to the environmental categories to be judged. We suspect that their answers reflect instead general belief strengths or values associated with these categories. Panelists are supposed to ignore the variation or updating of quantitative information included in the weighting task. We refer to the principles of bounded rationality (Gigerenzer et al. 1996, Todd et al. 2003), the arguments of the psychometric paradigm (Kahneman et al. 1982), and the constructive perspective on the elicitation of monetary values (Gregory et al. 1993) to support our hypothesis that respondents are directed by qualitative issues they perceive and extract from the 'task-story' (which they then link to their personal experience and worldview), rather than by the abstracted, numerical information provided. Therefore, when valuing environmental damage categories in a survey, the framing of the valuation task is a major issue. There are three aspects to the framing of valuation tasks (Kahneman et al. 1981, Scholz 1987, Payne et al. 1992). Respondents can be influenced by the context in which a task is presented (e.g. within a political agency or for a scientific study), the emotional and cognitive associations elicited by the content, or the reference point chosen (e.g. a relative improvement to the current situation will be valued very differently from a downturn in the situation). The main focus of the survey was to investigate what effect the information presented, particularly the qualitative and quantitative information, had on the preferences and weighting judgments of the participants. To investigate the qualitative aspects of the information presented, we tested for differences when different cognitive references are used (see Section 2.1). To study the quantitative aspect of presentation of the information, we conducted similar tests for different data models (see Section 2.2).

1.1 Valuation of different cognitive references: Safeguard subjects, perceived damages, and data models

When constructing a weighting survey, one of the issues critical for framing the valuation task is the method of presenting the environmental categories to the panelists. For the EI '99 project, several possible framings or reference systems were discussed. We compare the weighting of three different reference systems in this survey. Logically, the most simple reference system for eliciting weights is based only on the definitions of the safeguard subjects. This is the first reference system, referred to as 'safeguard subjects'. With aid of this reference, we intend to elicit the intrinsic values that the respondents attribute to the concepts of the safeguard subjects.

From a methodological point of view, weights are used to aggregate the damage (or impact) category results for a product system. Besides the intrinsic value attributed to the concept, the weighting factors should also account for the scarcity of a safeguard subject, i.e. the effective damage situation. Therefore, a reference system representing damages (or impacts) would better match the models used in LCIA than one that only defines the safeguard subjects. In general, reference systems consisting of damages can be introduced in two ways. One is to not introduce any data in the survey, but rather rely solelyon the panelist’s perception of and previous knowledge about the damage categories. A second reference is referred to as 'perceived damages'. Another way is to introduce additional information and data about the environmental categories. In our study, we use damage information derived from the characterization models of EI99. We refer to this third reference as 'data models'.[2]

In the questionnaire used for the EI '99 survey, data on damages in Europe were provided for Human Health Ecosystems and Resources (Mettier et al. 2004); that is, a data model framing was applied. For the present survey, therefore, it is of special interest to investigate how panelists cope with other reference systems and what bias these other reference systems may introduce. In a later section, we compare the influence that the three reference systems – safeguard subjects, perceived damages, and data models – have on the individuals' judgments.

[2] This reference could also be labeled 'expected' or 'predicted damages' (Hofstetter 1999). In order to avoid a confusion between perceived and expected damages we chose the term 'data model'.

1.2 Valuation of different data models: A test of scope

The topic how to specify a reference and present it to the panelists is also discussed in applications of Multi-Attribute Utility Theory (MAUT) (Humphreys et al. 1989). In MAUT, the tradeoffs between the various criteria may not be based on comparison of the of the criteria themselves, but rather on reasonable changes in the values of the criteria (von Winterfeldt et al. 1986). The question to answer is not how important one criterion is relative to another, but how much change in one criterion a respondent is willing to trade off for how much change in another criterion at a certain point. This fact has been neglected in previous panel studies thus far, as the elicitation of weights has generally been based on the total load of one impact or damage category, rather than tradeoff rates which rely on criteria value changes. The valuation of the data models was therefore framed change oriented; the values reflect changes in the three damage categories rather than the total damages. As the characterization models of EI99 are based on marginal modeling, one would expect marginal changes (e.g. the normalization values of EI99) as a reference for the weighting task. But respondents tend to have problems to weight small changes (Mettier & Hofstetter 2004). Therefore, we decided to use data on a larger scale (see Section 2.1). In contingent valuation (CV), there has been a heated debate over whether or not respondents are aware of the magnitude or scale of a weighted good. Therefore, so-called scope insensitivity is often discussed as a major source of bias in CV surveys, especially when the good is complex and tasks are unfamiliar (Fischhoff et al. 1993). Several studies have revealed that the respondents were not aware of the amount of the valued good (Kahneman et al. 1992, Hanemann 1996, Frederick et al. 1998). A way to prove that respondents understand the amount of the good that is valued are tests of scope. In a test of scope, the sample is split and a different version of a damage (or good) is presented to each half of the sample. To express this in terms of experimental social sciences, we investigate the effect of manipulating the basic information as an independent variable on the individuals’ weighting as a dependent variable.

The damages presented differ only in the amount of the good that is damaged; the rest of the information remains the same. According to utility theory, a greater amount of damage should also be given a higher weight. Therefore, the results of the two versions of the survey can be compared and it can be tested whether higher amounts of damages result in higher weights. In CV, such tests of scope are taken as proofs of validity (Arrow et al. 1993).

We chose to conduct such a test of scope for this study, so we made two versions of the questionnaire. In one of the versions, the damages reflect the European damages of the EI’99 report (Goedkoop et al. 1999), which are used as normalization values. In the other version, the damage to human health is higher (five times as high) and the damage to the ecosystem is lower (half as high). Consequently, we expected that the two versions of the questionnaire would yield different weighting factors. If this were to turn out to be true, we could assume that respondents are able to comprehend the magnitude of the figures we provide.

2 Survey Design

The sample consisted of 58 undergraduate students of Environmental Sciences at the ETH-Zurich. Although not experts, the students in this sample had had some experience with environmental valuations, as all had participated in exercises on weighting attributes for a multi-criteria analysis prior to this survey. The questionnaire was distributed during a course in Environmental System Analysis that included an introduction to LCA. The students filled out the questionnaire at home. Participation was voluntary, took about 90 minutes, and was not required to pass the course. Two versions of the questionnaire were distributed. These versions only differed in the figures presented to the participants in the data section (see below), which were varied to allow for a test of scope analysis.

The questionnaire was structured into the following five sections:

- Introduction and Personal Data: The survey was introduced, and questions pertaining to participants' age, gender, and knowledge of LCA were posed.
- Definition and rating of the safeguard subjects: The concept of safeguard subjects in LCA was introduced and definitions of the safeguard subjects Human Health, Ecosystems, and Resources were given. After that, the respondents had to rank the importance of the safeguard subjects and rate them on a graphical scale ranging from 0 (not important for LCA) to 100 (the most important safeguard subject for LCA). This data was used to evaluate the rating of the safeguard subjects, as the first cognitive reference described above.
- Perception of damages in Europe: An introduction was given, outlining how impacts and damages to safeguard subjects within LCA can be used for environmental management. Respondents were asked to state which safeguard subject they thought was damaged most in Europe and which was damaged least. Then, for every safeguard subject, they had to rate the seriousness of the damages in Europe on a five-point scale (from 'no serious damages' to 'very serious damages').
- Valuation of damage indicators*:* The definitions of the damage indicators from Eco-indicator99 (Goedkoop et al. 1999) and the concepts behind them were introduced – DALYs[3], PDF[4], and surplus energy[5] in particular. A reference scenario (see Section 3.3) was introduced and the students were asked to refer to this when assigning weights.
- Questions concerning cultural theory and attitudes towards the environment: 18 questions were posed to measure the level of agreement with the various cultural perspectives (Marris et al. 1996) and attitudes towards the environment (Thompson et al. 1994, Siegrist 1996).[6]

[3] DALYs (Disability Adjusted Life Years) are a concept about measuring damages to human health. These are used by the WHO and the World Bank, among others.) (Murray et al. 1996).

[4] In the Eco-indicaror99 methodology, PDF (the potentially disappeared fraction of species) is used as an indicator for damages to *Ecosystem Quality*.

[5] The concept of surplus energy is used within the Eco-indicator99 to express damages to resource stocks of minerals and fossil fuels.

[6] Data from this section will not be discussed further, as the questions did not allow classification of the respondents into different groups. This was not possible as the students' cultural perspectives and attitudes were too homogeneous, unlike the EI99 survey (Mettier & Hofstetter 2004), where the questions possessed good discriminative power.

In order to avoid having the previous weightings bias the respondents, different scales were provided for each of the three weighting tasks (safeguard subjects, perceived damage in Europe, and the data model). This made it so that the respondents had to think freshly about every weighting and could not just state the same answer. To make it so, the ratings were still comparable, all of the weighting tasks first contained a ranking.

2.1 The reference scenarios

After rating the safeguard subjects and the damages in Europe, which are two references that are not based on data, a damage scenario for a small region with 100,000 inhabitants was introduced to the panelists. Two versions were distributed in order to conduct the test of scope. Version A is based on the normalization values of EI 99[7]. Version B describes a region that has 5 times higher damages in Human Health, but half the damages in Ecosystems. This represents a region where certain densely populated areas are impacted strongly while wide expanses remain almost natural, as land is used extensively. Damage to the safeguard subject Resources is the same for A and B. The following data was presented:

- Human Health: Damage is 800-1600 DALYs/a for Version A and 4000-6400 DALYs/a for version B.
- Ecosystems: For Version A, an average of 45–55% of all species that could occur in a certain area are not found[8]. For Version B, this amount is 25–30%.
- Resources: For versions A and B the surplus energy is 600-800 Mio MJ/a (this is equal to the per capita [brutto-] energy use of 3800–5400 Swiss inhabitants).

3 Results of the Survey

A total of 109 questionnaires were distributed; 56 of Version A and 53 of Version B. Of these, 58 questionnaires were returned. That is a 53% rate of return, which is high for a survey. Of the returned questionnaires, 34 had Version A (normalization values of EI 99) and 24 had Version B (modified damage data).

First, we present a comparison between the valuation of the safeguard subjects and perceived damages in Europe. For the valuation of the safeguard subjects and the perceived damages in Europe, the two versions of the questionnaire were identical. Therefore, the results of all of the respondents are presented together. We address the question of whether the valuations for these references differ significantly (see Section 3.1). For the valuation of the scenarios, the questionnaires contained different data models. Thus, the answers from the two versions must be compared in order to conduct a test of scope (see Section 3.2).

[7] The normalization values from the EI99 report have been multiplied by 100,000 for the reasons described in section 2.2. The highest and the lowest normalization value from the three cultural perspectives were presented in order to cover parts of the uncertainty.

[8] The damage to Ecosystems is described as the average potentially damaged fraction of species (PDF) for that region. The normalization values from the EI 99 report (PDF*km^2*yr/yr) were divided by the size of the region to get the PDF. In the questionnaire, a region of 100,000 inhabitants was chosen, containing 1/3800 of the reference population in EI 99. Therefore, the size of the region chosen is 1/3800 of the reference area in EI 99 (≈ 1000 km^2).

3.1 Valuation of safeguard subjects and perceived damages

As a first valuation task, respondents ranked and rated the safeguard subjects according to importance. This task implies a trade-off between the intrinsic values respondents attribute to the safeguard subjects. The most important safeguard subject was assigned a value of 100%. The safeguard subjects that ranked 2^{nd} and 3^{rd} had to be placed on a graphical scale of 0% to 100%. The respondents were told that 100% means 'equal importance', compared to the safeguard subject ranked '1^{st}' and that 0% means the 'safeguard subject should not be included in LCA'. As is commonly done in LCA weighting, the figures were transformed into weights that add up to 1 ($W_{HH} + W_{EQ} + W_R = 100\%$). Results from ranking and rating the safeguard subjects are shown in **Table 1** and **Fig. 1**. A t-test reveals that the importance of the safeguard subjects Human Health and Ecosystems is rated significantly higher ($p < .0001$) than the importance of Resources. Human Health and Ecosystems are rated as equally important.

In Fig. 1a, a box plot depicts the statistics for each safeguard subject: the mean values (circle), the 95% confidence interval for the mean value (small T-line), the range that contains 50% of the values (box), the median (line across the box) and the extreme values (dotted T-line)

Fig. 1b shows a mixing triangle, which contains the assigned weights for all 56 participants that filled in the questionnaire. Every cross marks the weights assigned by a participant. The dotted lines mark the center of the mixing triangle where every safeguard subject is weighted equally. Drawing similar lines from every cross to the three axes allows reading every participant's attached weights possible.

For the next valuation task, we explained that some LCA methods result in damage indicators for the assessed products and that weighting damages to the safeguard subject is relevant for LCA when comparing products. Then, the respondents had to rank and rate the damages in Europe that arise from anthropogenic influences. No further data was given; the participants had to rely on their own knowledge and experience (perceived damages). To do the rating, respondents made a selection by filling in a cross on a five-step scale ranging from 'damage low' to 'damage high'. For the statistical analysis, the answer categories were given a code from 1 (damage low) to 5 (damage high)[9]. These codes were again converted into weights that add up to 100%. The results of the ranking are shown in **Table 2**. The ratings are presented in **Fig. 2**.

In the question about ranking, a majority responded that Ecosystems is the safeguard subject damaged the most in Europe and Human Health, the one damaged the least. Statistical analysis (t-test) of the rating data revealed significant distinctions in the rating of the perceived damages. Damages to Ecosystems are rated higher than those to Resources

Table 1: Ranking of the safeguard subjects (n=56)

	Most important safeguard subject (in %)	2nd most important safeguard subject (in %)	Least important safeguard subject (in %)
Human Health	51.8%	32.1%	16.1%
Ecosystems [1]	44.6%	42.9%	12.5%
Resources	3.6%	25.0%	71.4%

For the questionnaire, the term Ecosystems was chosen instead of Ecosystem Quality (EQ). For clarity's sake, we use the terminology from the questionnaire in this chapter.

[9] The results are not sensitive to the chosen range for the codes. Whether codes are from 1 to 3 or from 1 to 9, the results of the statistical tests are the same.

Table 2: Ranking of the perceived damages to the safeguard subjects in Europe

	Safeguard subject that is damaged the most (in %)	Safeguard subject that is damaged the least (in %)
Human Health	5.4	63.6
Ecosystems	71.4	5.5
Resources	23.2	30.9

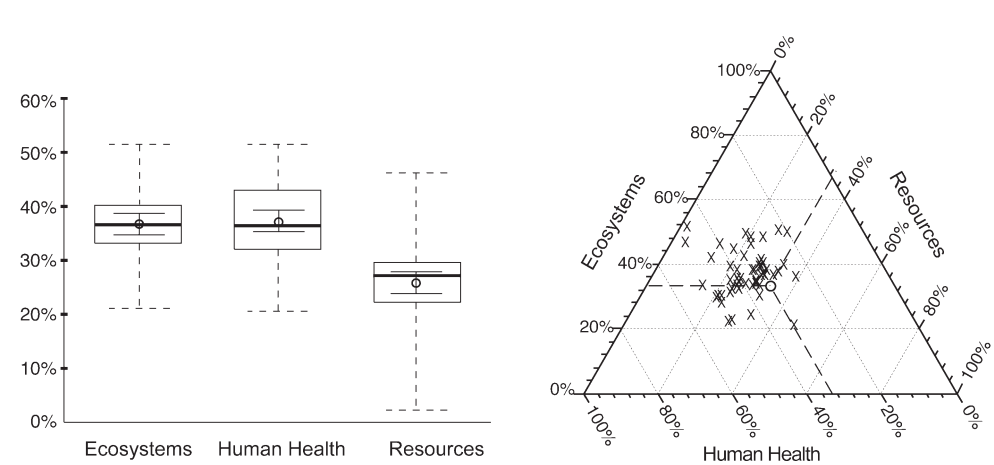


Fig. 1a&b: Weights assigned to the safeguard subjects Ecosystems, Human Health, and Resources

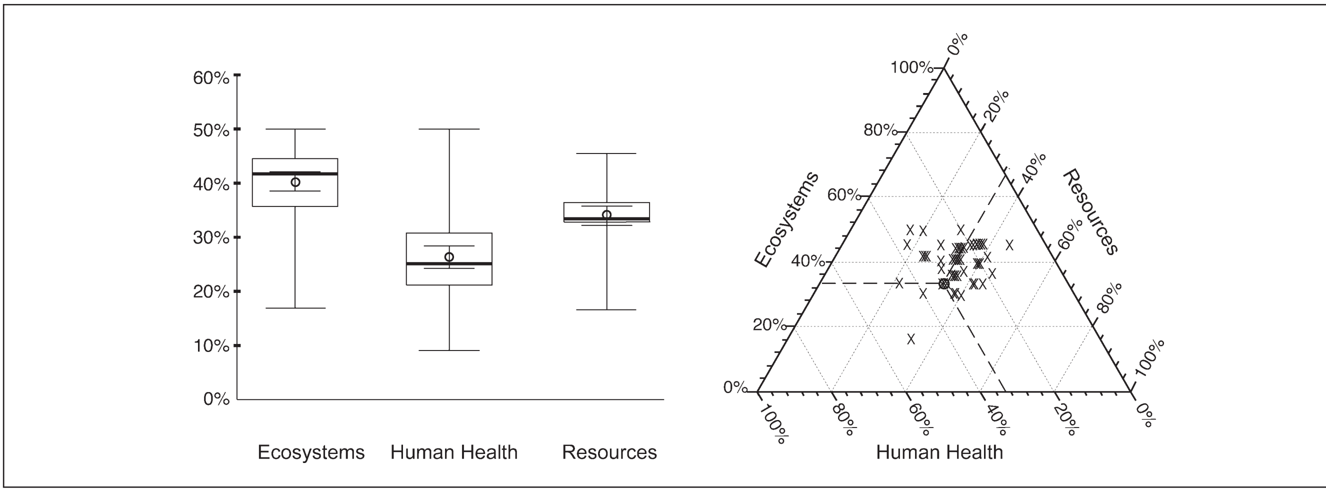


Fig. 2: Weights associated with the perceived damages in Europe (n=52). For explanations of the box plot or the mixing triangle, see Fig. 1

(p < .001) or Human Health (p < .0001). Moreover, Resources are perceived as being more damaged than Human Health (p < .0001).

3.2 Valuation of data models: Is there a difference between the two scenarios in a test of scope?

In order to prepare for the valuation of the data models, we introduced the damage indicators from EI99. After that, the respondents had to decide whether or not they acceptedthese indicators. This was done in order to avoid the situation of respondents rating a damage category low, because they do not agree that the indicator represents a damage. The acceptance was measured on a scale from 1 (poor indicator) to 5 (very good indicator). The acceptance of all three indicators was good, with an average of 3.3 for Human Health and 3.7 for Resources and Ecosystems. This meant that lack of acceptance of the indicators should not bias the valuation of the damages presented in the scenarios.

After introducing the indicators, a scenario was presented, as described in Section 2.1. In order to conduct a test of scope, we already mentioned that the two versions of the survey that had been given to the respondents contained significant differences in their data on the damage to the region. As a first valuation task, respondents had to choose between different reduction targets. The reduction targets were formulated such that a trade-off between the different damage categories had to be made. Such choice questions and assessment of tradeoff rates offer an interesting method for eliciting values on damage indicators. Moreover, the results from these choice questions cannot be easily compared to the weights directly allocated. Therefore, we will present the method and the results in Part B of this paper (in an upcoming issue of Int J LCA).

Following that, respondents had to allocate 50 reduction points among the three damages in order to maximize the resulting total reduction in environmental damage. The figures given by the respondents were multiplied by 2 in order to get weights that add up to 100%. The data was analyzed using a t-test to compare results from the two versions. Based on the data presented, we expected that the respondents who filled in Version B would assign higher weights to Human Health and lower weights to Ecosystems, but no such difference could be found. The p–values are far from being significant (p > .4 for all three damages). This finding can easily be understood when looking at **Table 3** where the weights are shown. The distribution of the weights is almost the same for both groups. This finding indicates that the data given had no influence on the weighting of the damages. The test of scope failed. We discuss the implications of this important finding further in the Conclusions section.

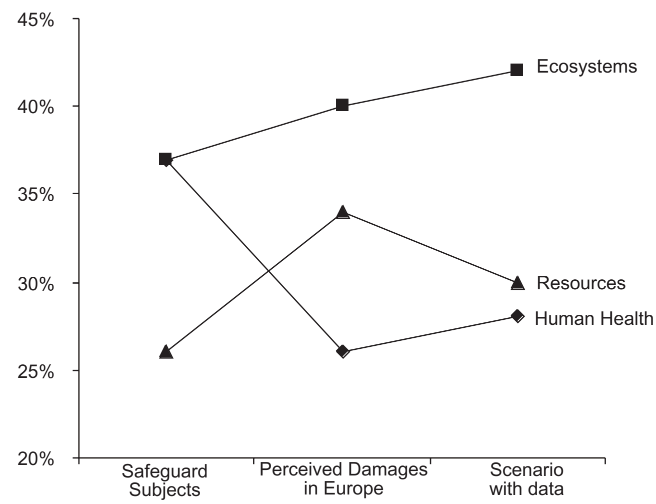
Table 3: Average weights assigned to the two versions of the questionnaire (in %)

	Weights Version A	Std. Dev.	Weights Version B	Std: Dev.	t-test (p-value)
Human Health	29.3	15.2	30.3	15.6	.42
Ecosystems	41.7	12.6	42.6	12.6	.72
Resources	29.0	13.2	27.1	13.5	.57

3.3 Are there differences in the ratings of the three references?

In order to compare the allocation of weights for the three references, we analyzed the data using t-tests. Because the weights attributed to the damage scenarios by the two groups did not differ, the data for this reference is analyzed for all of the respondents together. The average weights and the results of the tests are shown in **Fig. 3**.

The analysis revealed that the safeguard subjects (intrinsic values) are rated significantly differently than the perceived damages in Europe or the damage scenarios. When rating the importance of the safeguard subjects, Human Health was the most important; but it is the least important when rating the damages in Europe or the data given in the scenario. In contrast, the importance of the safeguard subject Resources is rated the least important, but Resources is rated higher in terms of damages. This shift is also recognizable when testing the differences between the ratings of the three references statistically. Significant differences are found be-



	Intrinsic values vs. perceived damages (p-value)	Intrinsic values vs. data model (p-value)	Perceived damages vs. data model (p-value)
Human Health	<.0001**	<.0001**	.44
Ecosystems	.019*	.003*	.11
Resources	.019*	<.0001**	.021*

Fig. 3: Ratings of the environmental categories for three different cognitive references (diagram) and statistics from a t-test (Table). The p-value denotes the probability that the mean rating of an environmental category is the same for the references compared. Significant differences in the rating (p< .05) are marked with *; highly significant (p< .001) with **

tween the rating of safeguard subjects and the rating of perceived damages; for Human Health, the differences are even greater (see the first column in Table 1). The same holds true for a comparative rating of safeguard subjects and the data model (see the second column in Table 1). As one can see, this shift in the ranking does not arise from a bias of the different scales used to elicit the weights, because the ranking task was the same for each reference. We can, therefore, conclude that panelists distinguish between the importance of a safeguard subject (intrinsic values) and the importance of perceived damages in Europe. This is important since, to weight the damage indicators, we need to elicit the panelists' assessment of the actual damages, rather than their assessment of the general importance of the safeguard subjects. The step-by-step presentation of the information did not allow the order of the questions to be random[10]. Therefore, the order may be a source of bias (sequence effect), as the previous weighting of perceived damages in Europe may influence the final weighting of the data model. The sequence effect is not relevant for most panel surveys where only one reference is rated.

When comparing the perceived damages in Europe to the data model, only Resources is rated significantly differently, whereas Human Health and Ecosystems are rated more or less the same. This is surprising, as the method of eliciting the weights – especially the scale used – was quite different. We conclude that the importance of the safeguard subjects is perceived to be different from the importance of the perceived damages or of the data model, while the ratings of the perceived damages in Europe and the data presented are similar.

[10] In a randomized questionnaire, all of the information must be provided before the valuation tasks are performed, because the two references referring to damages cannot be valued before the safeguard subjects are introduced. To avoid starting the questionnaire with a huge introduction, information was given step by step.

4 Conclusions and Recommendations

The study was based on damage categories. Therefore, the conclusions we draw may not apply to surveys based on indicators located earlier in the cause-effect chain. The most important finding of this study is that the respondents were not sensitive to numerical data that was presented as reference information for weighting damage categories. The analysis indicates that they obviously were not able to take into account the magnitude of the data presented. It is evident that the test of scope failed. This raises three questions. First, if the respondents are not influenced by the data that is presented in the questionnaire, what is the reference or mental model upon which respondents base their beliefs and values? Second, can we expect that 'more sophisticated'[11] subjects would respond differently? Third, which prerequisites should an empirical weighting procedure for student samples or sophisticated decision makers fulfill in order to incorporate numerical data.

With respect to the first question, it seems relevant that the weights from the valuation of the data model were similar to those that resulted from the valuation of personal beliefs about damages in Europe, although the method of measurement and context of the questions were very different. The general values and beliefs about the damage categories which a respondent has prior to the survey seem to play a more important role than the data presented in the survey. Moreover, the qualitative aspects of the category that is valued seems to be more decisive than the figures specified in the questionnaire. This is supported by the finding that respondents rate the importance of safeguard subjects significantly differently than they do the

[11] Scholz (1987, p. 134) distinguishes between "experienced and sophisticated decision makers. By experienced decision makers we mean individuals who either have performed a certain decision repeatedly or have experienced the information, e.g. have observed the concrete history of the environmental damage indicators. By sophisticated decision makers, we mean individuals who possess the ability or knowledge to cope with a situation or find an appropriate solution for a problem."

damages to these safeguard subjects. Obviously, endeavors at explaining indicators for the damage categories and introducing numerical data about damages are not successful, especially if the sample does not consist of experts. Similar outcomes could be found in our former survey (Mettier & Hofstetter 2004), where two thirds of the respondents stated that they were not influenced by any figures presented.

This result is in line with findings from risk communication (Scholz et al. 1993) and diagnostic decision making (Gigerenzer 1999), which show that subjects and decision makers are very insensitive to numerical data, when this information cannot be linked to an individual's perceivable, experiential knowledge. Just changing abstract numbers is, thus, not sufficient for conveying changes to environmental reality.

The second question concerns the participants' knowledge and skills. One could question to what extent the results depend on the sample. Is it likely that experienced environmental experts would incorporate the numerical data in the questionnaire in a 'better', (i.e. more sensitive) way? This question is difficult to address and cannot be answered without empirical research. Nevertheless, it is likely that there are only a few people who have extensive experience or the degree of sophistication to prioritize and weight environmental problems and categories appropriately. Cognitively representing, integrating, and differentiating among statistical data relevant for LCA damage weighting seems to be – even for educated scientists – a difficult task, and few possess sufficient aptitude for this. Our data suggest that people who are not used to such valuation tasks tend to report a general belief about the categories rather than assign weights that refer to the reference specified. Therefore, it is important to provide more careful, adequate preparation for the valuation task, so that the statement about the reference is more valid; otherwise, one may accept that the reported figures are general statements. Possible methods for introducing the valuation task more carefully include using focus groups or multi-round surveys in which respondents get feedback and have the opportunity to change their opinion (Delphi methods).

The latter issue is the core of the third question. How should questions about weighting be posed in further studies? Our study supports some conclusions for future weighting surveys about three aspects of framing: context, emotional and cognitive associations, and the reference point. Our point-of-view on the reference point is that the main goal of LCA studies – to reduce environmental consequences associated with product life cycles – already defines a reference point. The reference valued in weighting surveys should therefore be expressed as a defined reduction in an environmental category. Although we could not find major differences in weighting between (the total) perceived damages in Europe and the reduction of the damages in the local scenarios, we think that from a methodological point of view it is important to specify the reference as a relative change to the category, not the total of the category.

For weighting surveys in LCIA, the values are elicited in order to interpret the importance of category indicator results. This application already introduces major aspects of the context. Regarding the consistency of the LCIA models, the context should therefore be characterized by data describing the consequences of an environmental impact or damage category in relative terms. Our study showed, in terms of results, that (for non-experts) the qualitative aspects of this data might have more influence than the quantitative.

Because qualitative information is better understood, it becomes easier to focus increasingly on information about the model structure, i.e. the environmental problems integrated into a damage category. For example, every environmental problem that contributes to a damage category could be presented to the participants, thereby potentially reducing so-called prominence or availability effects (Nisbett et al. 1980, van der Pligt et al. 1998). Availability effects are often reported in CV studies where a category gets higher values when it is valued on its own than when it is part of (or embedded into) another category. It is therefore probable that environmental problems integrated into a large category containing numerous problems are systematically underestimated. In the EI99, the damage category Human Health contains six environmental problems (respiratory effects, climate change, radioactive emissions…) while Resources only contains two (energetic and non-energetic resources). This suggests that the weighting of an environmental problem associated with Resources (e.g. energetic) may end up being too high as compared to the weighting of an environmental problem associated with Human Health (e.g. climate change). Therefore, the environmental problems integrated into a damage category may provide decisive information for enhancing the quality of a weighting procedure for damage categories. This conclusion is analogous to the 'out of sight, out of mind' bias in decision research (Fischhoff 1982, Kleindorfer et al. 1993), which implies that a category is considered to be less likely or less important, the fewer subcategories are explicitly listed.

A gap may exist between two requirements the polled panel should meet: internal validity and external validity. The panel should be externally valid; that is, representative for a broad stakeholder group of LCA (Mettier & Hofstetter (2004)). At the same time, the results should be internally valid, meaning that we really measure what we intend to measure. As only some experts can manage the data presented, there may be a conflict between internal and external validity. We therefore see two ways of proceeding with further weighting surveys. One is to focus on internal validity and conduct comprehensive multi-round expert procedures. These procedures are based on a lot of environmental information, like the data model in this study. The experts become experienced and, thus, experts as well, by repeated measurements. The experience and time necessary for comprehending this data restricts the selection of participants. Nevertheless, we do not expect that the limited number of experts in the field of LCIA and weighting restricts similar studies among experts. Another interesting approach would be to focus on external validity and poll a broader population's general beliefs about the environmental categories. For this procedure, one may choose a reference similar to the perceived damages in Europe; that means not relying on data. The outcome wouldn't match the specific category indicators used in LCIA as well, but would be more representative.

As mentioned, we also examined the influence of different question formats on the weighting results. These findings will be presented in Part B of this paper.

Acknowledgements. We would like to thank Patrick Hofstetter, Thomas Baumgartner, Thomas Köllner, Ruedi Müller Wenk and Arthur Braunschweig for the valuable discussions and the Swiss National Science Foundation for the founding of Thomas Mettier.

References

Arrow K, Solow R, Leamer E, Portney P, Radner R, Schuman H (1993): Report to the NOAA Panel on Contingent Valuation. Federal Register 58 (10)
Dawes R M (1971): A case study of graduate admissions. Application of three principles of human decision making. Amer Psychol 26, 180–188
Elstein AS, Schulman LS, Sprafka SA (1978): Medical problem solving. Cambridge, MA, Harvard University Press
Finnveden G (1999). A Critical Review of Operational Valuation/Weighting Methods for Life Cycle Assessment. Stockholm
Finnveden G, Hofstetter P, Bare J, Basson L, Ciroth A, Johansson J, Mettier T, Norris G, Seppälä J, Volkwein S (2002): Normalization, Grouping and Weighting in Life Cycle Impact Assessment. In: Udo de Haes H, Finnveden G, Goedkoop M, Hauschild M, Hertwich E, Hofstetter P, Jolliet O, Klöpffer W, Krewitt W, Lindeijeret E et al.: Life-Cycle Impact Assessment: Striving towards best practice. Brussels, SETAC Press
Fischhoff B (1982): Debiasing. In: Kahneman D, Slovic P, Tversky A: Judgement under Uncertainty: Heuristics and Biases. Cambridge, Cambridge University Press: 422–444
Fischhoff B, Jacobs Quadrel M, Kamlet M, Loewenstein G, Dawes R, Fischbeck P, Klepper S, Leland J, Stroh P (1993): Embedding effects: stimulus representation and response mode. Journal of Risk and Uncertainty 6, 211–234
Frederick S, Fischhoff B (1998): Scope insensitivity in elicited values. Risk Decision and Policy 3, 109–124
Gigerenzer G, Goldstein DG (1996): Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev 103 (4) 650–669
Gigerenzer G, Todd, PM, ABC-Research Group (1999): Simple Heuristics that make us smart. Oxford, Oxford University Press
Goedkoop M, Spriensma R (1999): The Eco-indicator 99 – A damage oriented method for Life Cycle Assessment: Mehodology report. Amersfoort, PRé Consultants B.V.
Gregory R, Lichtenstein S, Slovic P (1993): Valuing environmental resources: A constructive approach. Journal of Risk and Uncertainty 7, 177–197
Hanemann MW (1996): Theory Versus Data in the Contingent Valuation Debate. In: Bjornstad DJ, Kahn JR: The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs. Cheltenham, Edward Elgar
Harada T, Fujii Y, Nagata K, Inaba A, Mettier T (2000): Panel Test for Japanese LCA Experts Aiming to Weight Safeguard Subjects. Proceedings of the 4th Int. Conference on EcoBalance, Tsukuba, Japan, Epochal Tsukuba
Hertwich E, Hammitt J, Pease W (2000): A theoretical foundation for life-cycle assessment. Recognizing the role of values in environmental decision making. JIE 4 (1) 13–28
Hofstetter P (1998): Perspectives in Life Cycle Impact Assessment: A structured approach to combine models of the technosphere, ecosphere and valuesphere. Boston, Kluwer Academic Press
Hofstetter P, Braunschweig A, Mettier T, Müller-Wenk R, Tietje O (1999): The Mixing Triangle: Graphical Decision Support for Comparisons in LCA – Why correlation matters. JIE 3 (4) 97–115
Hofstetter P (1999): Debate: Best Available Practice Regarding Impact Categories and Category Indicators in LCIA – Comments by Patrick Hofstetter, September 1999: Starting Point Weighting Methods: Hints and Hits – 2 Global LCA Village (03) 2002 <http://www.scientificjournals.com/sj/lca_village/Pdf/aId/4857>
Humphreys PC, Wisudha AD (1989): Handling decisions problems: A structuring language and interactive modules. Final report 1 Sep 87–3 Oct 89. London, London School of Economics and Political Science
Huppes G, Sas H, de Haan E, Kuyper J (1997): Efficient Environmental Investments. Paper presented at the SENSE International Workshop, Amsterdam
ISO (2000): ISO 14042:2000(E) Environmental management – Life cycle assessment – Life cycle impact assessment. Geneva, International Organisation for Standardisation
Itsubo N, Sakagami M, Washida T, Kokubu K, Inaba A (2004): Weighting Across Safeguard Subjects for LCIA through the Application of Conjoint Analysis. Int J LCA 9 (3) 196–205
Jolliet O, Brent A, Goedkoop M, Itsubo N, Müller-Wenk R, Peña C, Schenk R, Stewart M, Weidema B (2003): Final Report of the LCIA Definition Study, Life Cycle Impact Assessment Programme of the Life Cycle Initiative
Kahneman D, Knetsch J (1992): Valuing public goods: The purchase of moral satisfaction. Journal of Environmental Economics and Management 22, 57–70
Kahneman D, Slovic P, Tversky A (eds) (1982): Judgment under Uncertainty: Heuristics and Biases. New York, Camebridge University Press
Kahneman D, Tversky A (1981): The framing of decisions and the psychology of choice. Science 211, 453–458
Kleindorfer P, Kunreuther H, Schoemaker P (1993). Decision Sciences: An Integrative Perspective. Cambridge, Cambridge University Press
Lindeijer E (1997): Results Try-out Japanese/Dutch LCA Valuation Questionnaire 1996. Amsterdam, University of Amsterdam, IVAM Environmental Research
Marris C, Langford I, O'Riordan T (1996): Integrating Sociological and Psychological Approaches to Public Perceptions of Environmental Risks: Detailed Results from a Questionnaire Survey. Norwich, Center for Social and Economic Research on the Global Environment (CSERGE)
Mettier T, Baumgartner T (2000): A non-monetary approach to weight environmental damages in life cycle assessment: Panel methods compared to contingent valuation and other monetary approaches. <http://www.wu-wien.ac.at/esee2000>
Mettier T, Hofstetter P (2004): Survey insights into weighting environmental damages: influence of context and group. JIE 8 (4) 189–209
Murray CJL, Lopez AD (1996): The Global Burden of Disease – Volume I of Global Burden of Disease and Injuries Series. Boston, WHO / Harvard School of Public Health / World Bank, Harvard University Press
Nagata K, Fujii Y, Ishikawa M, Yokota M, Ureshino M (1996): Developing an Impact Assessment Methodology Using Panel Data. Proceedings of RITE International Workshop on Valuation. Tokyo
Nisbett RE, Ross L (1980): Human inference: strategies and shortcomings of social judgment. Englewood Cliffs, NJ, Prentice-Hall
Owens W (1999): Why Life cycle assessment is now described as an indicator system. Int J LCA 4 (2) 81–86
Payne JW, Bettmann JR (1992): Behavioral decision research: A constructive processing perspective. Annual Review of Psychology 43, 87–131
Puolamaa M, Kaplas M, Reinikainen T (1996): Index of Environmental Friendliness: A Methodological Study. Helsinki, Statistics Finland
Sangle S, Ram Babu P, Khanna P (1999): Evaluation of Life cycle impacts: Identification of societal weights of environmental issues. Int J LCA 4 (4) 221–228
Schmidt W-P, Sullivan J (2002): Weighting in life cycle assessment in a global context. Int J LCA 7 (1) 5–10
Scholz RW (1987): Cognitive Strategies in Stochastic Thinking. Dordrecht, Reidel
Scholz RW, Mayer J (1993): Repräsentationsmodi in computergestützten Lernprogrammen zu bedingten Wahrscheinlichkeiten. [Modes of representation of conditional probability in computer-supported learning programs]. Paper presented at the 27. Bundestagung für Didaktik und Mathematik [27. Annual Meeting for Didactics and Mathematics], Freiburg, Schweiz. (pp. 323–327) Bad Salzdetfurth, Franzbecker
Scholz RW, Tietje O (2002): Integrating Knowledge with Case Studies – Formative Methods for Better Decisions. Thousand Oaks, CA, Sage Publications
Seppälä J (1999): Decision Analysis as a Tool for Life Cycle Impact Assessment. LCA Documents Vol. 4, ecomed publishers, Landsberg, Germany
Seppälä J, Basson L, Norris G (2002): Decision analysis frameworks for life cycle impact assessment. JIE 5 (4) 45–68
Siegrist M (1996): Fragebogen zur Erfassung der ökozentrischen und anthropozentrischen Umwelteinstellung. Zeitschrift für Sozialpsychologie 27, 290–294
Thompson S, Barton M (1994): Ecocentric and anthropocentric attitudes towards the environment. Journal of Environmental Psychology 14, 149–157
Todd PM, Gigerenzer G (2003): Bounding rationality to the world. Journal of Economic Psychology 24, 143–165
UBA (1995): Ökobilanz für Getränkeverpackungen. Text 52/95. Berlin, Umweltbundesamt
van der Pligt J, van Schie JCM, Hoevenagel R (1998): Understanding and valuing environmental issues: The effects of availability and anchoring on judgment. Z Exper Psychol 45 (4) 286–302
Virtanen Y, Torkkeli S, Wilson B (1999): Evaluation of a Delphi Technique Based Expert Judgment Method for LCA Valuation – DELPHI II. Espoo, Technical Research Centre of Finland
von Winterfeldt D, Edwards W (1986): Decision Analysis and Behavioral Research. Cambridge, Cambridge University Press
Werner F, Scholz RW (2002): Ambiguities in Decision-Oriented Life Cycle Inventories: The Role of Mental Models. Int J LCA 7 (6) 330–338

Received: June 24th, 2004
Accepted: August 2nd, 2005
OnlineFirst: August 3rd, 2005