Int J Life Cycle Assess (2008) 13:468–476
DOI 10.1007/s11367-008-0022-z

METHOLOGOY • RESEARCH ARTICLE

Measuring preferences on environmental damages in LCIA. Part 2: choice and allocation questions in panel methods

Thomas Mettier · Roland W. Scholz

Received: 9 April 2008 / Accepted: 11 June 2008 / Published online: 15 August 2008

Abstract
Background, aim, and scope Within life cycle impact assessment (LCIA), 'panel methods' has become a common term to denominate methods that elicit and measure stakeholders' stated preferences on environmental impact categories. Such panel procedures use different question formats to elicit information on weighting across impact categories from the stakeholders. The two most frequently used question formats are score allocation and choice between alternatives. The differences between these two question formats were analyzed in order to give advice on how to frame future panel procedures.
Materials and methods A choice-based weighting procedure (choice experiment) for the three damage categories of human health, ecosystems quality, and resources was developed and executed. A logistic regression model was applied in order to estimate the weighting factors for the polled sample. Results from this choice-based procedure were compared to the results from an allocation-based procedure described in part 1 of this paper.
Results When weighting factors are elicited by score allocation questions, panelists tend to distribute the scores more equally. A factor of 1.5 between the least and the most weighted damage category was found. Weighting factors from a choice experiment were more spread, i.e., the most important category was weighted considerably higher, whereas the other two categories were weighted less. Thus, for the choice experiment, the range between the most and the least weighted categories was considerably bigger—by about a factor of 4.
Discussion A comparison of the two procedures revealed that the weighting of environmental damage categories is considerably influenced by the question of format. The reason for these variations may be different cognitive routines that are applied. In addition, several advantages and shortcomings of choice experiments are discussed.
Conclusions The developed, choice-based procedure provided meaningful results. Thus, choice experiments, often used for the monetary valuation of environmental goods, can also be applied in LCIA to elicit nonmonetary weighting factors.
Recommendations and perspectives Choice experiments form a new interesting approach for weighting procedures in the future as they have some advantages over the often used score allocation methods. They are simple and more realistic than other procedures, as panelists have practiced in choice tasks from everyday life. We, therefore, recommend such choice-based procedures for future panel studies.

Keywords Choice experiments · Framing · Panel surveys · Stated preference · Weighting of damage categories in LCIA

1 Background, aim, and scope

In general, life cycle impact assessment (LCIA) finishes up with a set of three to 12 impact category indicator results that describe the impact of a product system on the environment. Weighting across impact categories is often needed in order to interpret these category indicator results and to draw conclusions. Within LCIA, 'panel methods'

Responsible editor: Atsushi Inaba

T. Mettier (✉) · R. W. Scholz
ETH Zurich, Institute for Environmental Sciences,
Natural and Social Sciences Interface (NSSI),
CH-8092 Zurich, Switzerland
e-mail: Thomas.Mettier@alumni.ethz.ch

Table 1 Overview of stated and revealed preference methods

Stated preference (panel methods)		Revealed preference
Choice questions	Allocation	
Contingent valuation (referendum design with bidding cards)[a]	Direct allocation of weights Swing weights[b]	Distance to target
Conjoint analysis (choice design)	Conjoint analysis (rating design)[c]	Reduction costs
Choice experiments	AHP[d]	Travel costs
	Contingent valuation (open-ended design)[a]	Hedonic pricing

[a] In contingent valuation, respondents can state an amount of money they are willing to pay for an environmental good (open-ended design) or accept/reject an amount noted on a bidding card (bidding cards design)
[b] Swing weights are weighting factors that do not add up to 1 (or 100%). In general, the most important category is set to 1 (or 100%) and the other categories are rated accordingly. Regular weighting factors are derived by normalization of the swing weights
[c] Conjoint analysis is a term used for a variety of methods originally developed by market researchers to elicit consumer's stated preferences for new products (Green and Srinivasan 1990). In conjoint analysis, there also exist designs where a series of options is ranked and rated (rating design) in order to value the attributes. But nowadays, a simple choice design is applied in most conjoint studies. We use the term 'choice experiment' in this paper for this kind of design, as it is often used in studies that value environmental goods
[d] In the AHP, a pairwise comparison of attributes leads to relative scores that are used to calculate the final weighting factors

has become a common term to denominate methods that elicit and measure stated stakeholders' preferences on environmental impact categories. This information attained with these methods can be used for the grouping or weighting of impact categories within a life cycle assessment (LCA) study. Contrary to revealed preference methods, which are based on observations or reported behavior, stated preference methods elicit values that are expressed in response to hypothetical scenarios or experiments. A relevant question concerning panel procedure is how to elicit weighting factors in LCIA. Stated preference methods use ranking, score allocation, or choice tasks to obtain information on respondents' preferences.

The most straightforward method to elicit weighting information on impact categories surveys the direct allocation of scores that add up to 100% (see, e.g., Lindeijer 1996; Nagata et al. 1996). In this case, the allocated scores are equal to the weighting factors used. Other score-allocating methods are based on pairwise comparisons and the scores represent the relative importance between two categories. The weighting factors are finally calculated based on these relative scores, e.g., in the analytical hierarchy process (AHP) (Saaty 1980). Such methods have been used by Puolamaa et al. (1996), Sangle et al. (1999), Seppälä (1999), Harada et al. (2000), and Mettier and Hofstetter (2004). Methods for the monetary valuation of environmental goods,[1] such as contingent valuation or conjoint choice experiments (see Section 2.1) are often based on choice questions. A choice task is much easier, less time-consuming, and often more realistic than the rating or ranking tasks used in the other elicitation techniques. It is believed, though difficult to prove, that the more closely a research task mimics real behavior, the more valid and reliable the results (Sell et al. 2007).[2] In contingent valuation, for example, respondents have to state whether they would be willing to pay an amount of money marked on a bidding card for a defined environmental good. However, in conjoint analysis studies and choice experiments, respondents have to choose between different alternatives. In LCIA, such a conjoint analysis method has been used by Itsubo et al. (2004), but most panel studies in the LCIA context make use of allocation techniques.

Table 1 shows an overview of the most common allocation- and choice-based stated preference methods as well as revealed preference methods used to elicit preferences in LCIA or in environmental economics. Allocation- and choice-based methods are commonly used techniques in stated preference studies, but knowledge is still poor regarding the differences between these two approaches, as well as their strong points and drawbacks. Irwin et al. (1993) provided some interesting fundamentals, as they demonstrated that results from choice and (multicriteria) allocation questions often contain inconsistencies. Thus, respondents often prefer an alternative in choice experiments that scores weaker than other alternatives when using allocation techniques. This inconsistency between allocation and choice is an example of preference reversal. The preference reversal between choice and multicriteria choice experiments can be explained by the different cognitive processes and decision processes that are applied. In choice experiments, respondents often prefer the alternative that scores best in the most important category. This procedure

[1] For an introduction to monetization and LCA, see Finnveden et al. (2002).

[2] See also the discussion on representative design (Dhami et al. 2004).

has been called elimination by aspects (Tversky 1972; Hutchinson and Gigerenzer 2005). In this lexicographic choice, heuristic respondents emphasize the most important category. When allocating scores to categories, respondents often utilize a so-called anchoring and adjustment heuristic (Kahneman et al. 1982). This means respondents utilize an anchor value which they adjust to the allocation task. The anchor value is often the average score (e.g., 33% in the case of three impact categories that have to be weighted) and the adjustment is often small. Thus, more or less equal weights are stated for each category. These findings support the assumption that the question format significantly influences the outcome of a weighting procedure. But how large is this influence?

In this paper, we investigate the differences between an allocation- and a choice-based panel procedure in order to appraise the significance of the question format. For this analysis, we employ the damage categories specified in the Eco-indicator 99 (EI'99) (Goedkoop and Spriensma 1999), namely, human health (HH), ecological quality (EQ), and resources (R) and a sample of students of environmental sciences (see part 1 of this paper, Mettier et al. 2006). A choice-based panel procedure has been worked out (see Section 2.1) in order to elicit weighting factors for these damage categories that can be compared to the results from an allocation-based procedure already published in part 1 of this paper. In a first step, we evaluate whether respondents answer in a consistent manner when confronted with an allocation task and a choice experiment. In a second step, we test the hypothesis that a weighting procedure based on a choice experiment leads to a larger spread between the most and the least important impact categories than a direct allocation of weights. We put forward this hypothesis based on the assumed underlying cognitive processes, although the conjoint analysis conducted by Itsubo et al. (2004) did not reveal any evidence for it—the spread between the most and the least weighted damage category was below a factor of 1.5 in this study. Based on these interesting insights, we are able to give guidance on how to frame a future panel procedure.

2 Materials and methods—a choice-based weighting procedure

2.1 Choice experiments

The allocation-based procedure has already been described in part 1 of this paper. Therefore, we focus here on the choice-based procedure that has been developed.

Choice experiments are frequently used to elicit the value of environmental goods (Boxall et al. 1996; Alpízar et al. 2001; Hearne and Salinas 2002; Carlsson et al. 2003; Lehtonen et al. 2003; Christie et al. 2006; Colombo et al. 2006). Choice experiments are also applied in various other fields in order to measure preferences of people. In medical science, for example, choice experiments are used in patient studies to evaluate various forms of cancer treatment (Sculpher et al. 2004).

One assumption underlying choice experiments is that people generally have preferences among features of an alternative and are willing to accept various trade-offs. For example, a respondent may accept a higher use of natural resources in return for less impact on human health. In general, a choice experiment asks individuals to choose one alternative from a choice set where each alternative is described by a bundle of attributes. In our experiment, these attributes are represented by the damage categories. Several choice sets are presented to each individual in an experiment. These choices reflect the importance individuals assign to each damage category. Contrary to allocation methods, these choices do not, in the majority of cases, allow one to calculate a distinct personal score for every respondent. But the pattern of these choices can be statistically analyzed with a logit analysis for the polled sample (see Section 2.3 for details) to produce an overall relative importance or weighting factors for each damage category. If cost is included as an attribute, money-equivalent values can additionally be calculated for each damage category. But in our study, costs have been excluded as an attribute, as the study focuses on weighting factors. For a choice experiment, a minimum of three attributes is required.

Designing an experiment that attempts to incorporate intangible attributes requires much care (Shaw et al. 1989). Intangibles (like many impact categories) do not generally have a readily comprehensible measurement scale. To use such attributes in a choice experiment, it is, therefore, necessary to produce some unambiguous form of measurement that is understood by the respondent, while still meaning something in the LCIA context. If such a scale can be successfully produced, the respondents will understand the exercise and the results can be used to group and weight impact categories.

In order to produce such a scale for the three damage categories, the definitions and normalization data from EI'99 have been presented in the questionnaire. Thus, a reference scenario has been described that defined the damage level for every damage category according to the normalization data of EI'99 (see part 1, section 3.1). The choice questions presented to the respondents referred to two different reduction programs. Every reduction program reduces the three damage categories HH, EQ, and R by a certain percentage (see box for exemplarily chosen programs A and B in Fig. 1). The reduction targets presented were not marginal, although marginal changes would be preferable as

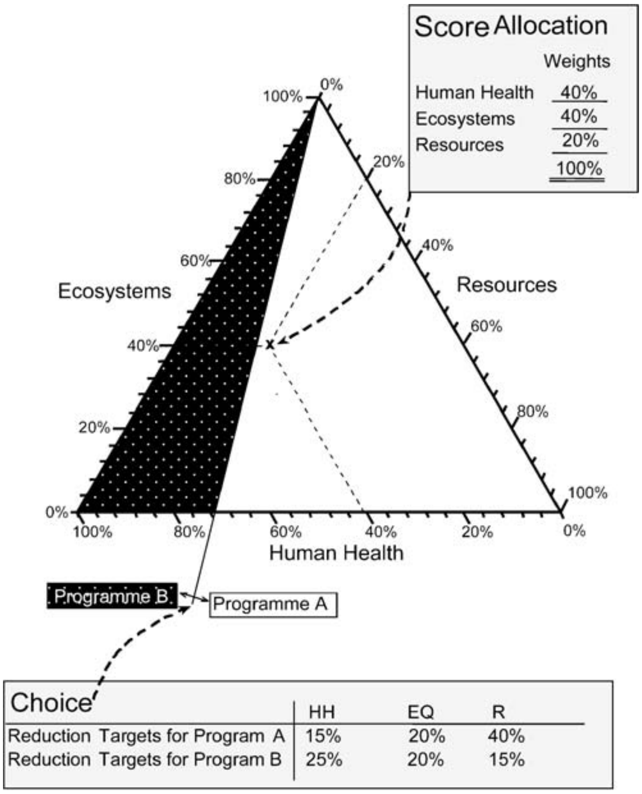

Fig. 1 Difference between the type of data from a choice and score allocation question. As an example, the 'x' marks the weighting set of a respondent that allocated weights of 40% for human health, 40% on ecosystems, and 20% on resources. The *shaded area* represents the preference area for a respondent who prefers reduction program B; the *white area* on the other side of the line of indifference would represent the preference area for program A

the weighting factors are used to weight marginally modeled damage categories. This approach was followed because we imply that marginal changes are difficult to value, as they do not produce a meaningful reference to the respondents.

The differences between the allocation- and choice-based weighting tasks are assessed in two steps. In a first step, it was analyzed whether figures given in the allocation question are consistent with the choices respondents made. Consistency between allocation and choice questions was assessed using preference areas within the mixing triangle, which are explained in Section 2.2. In a second step, the data from choice questions was statistically analyzed in order to generate weighting factors for the sample. For data analysis, we used logistic regression analysis. In Section 2.3, we introduce the basics of the logistic regression and of logit models in general, which are often used in choice experiments. These weighting factors derived from choice questions are compared to the weighting factors from direct allocations. Thus, we can test our hypothesis that choice questions lead to a larger spread between the weighing factors.

2.2 Preference areas in the mixing triangle

As described by Hofstetter et al. (1999), a set of weighting factors, w_{HH}, w_{EQ}, and w_{R}, for the three damage categories can be represented as a distinct point in the mixing triangle (see score allocation 'x' in Fig. 1). Such a set of weighting factors for every respondent results from the allocation task where weighting factors are directly allocated on the damage categories (see part 2 of this paper). On the contrary, the choice method worked out here does not produce such a distinct set of weighting factors for one respondent. The respondent has to choose between two reduction programs. The line between the white area and gray area in Fig. 1 represents all sets of weights for which the two programs are equal (line of indifference). For the example presented in Fig. 1, the line of indifference would be represented by Eqs. 1 and 2:

$$15\% \times w_{\mathrm{HH}} + 40\% \times w_{\mathrm{R}} = 25\% \times w_{\mathrm{HH}} + 15\% \times w_{\mathrm{R}} \quad (0 \leq w_{\mathrm{HH}}, w_{\mathrm{R}} \leq 1) \tag{1}$$

or simplified as:

$$10\% \times w_{\mathrm{HH}} = 25\% \times w_{\mathrm{R}} \Rightarrow \; w_{\mathrm{HH}} = 2.5 \times w_{\mathrm{R}} \quad (0 \leq w_{\mathrm{HH}}, w_{\mathrm{R}} \leq 1) \tag{2}$$

For all points on the line of indifference, a reduction of one unit of HH corresponds to a reduction of 2.5 units of R. If the respondent weights of HH are higher than the ratio of 2.5, one would choose program B and a set of weighting factors in the gray area results. Likewise, for program A, a set of weighting factors would result if the ratio was lower than 2.5. Thus, by choosing a program, one preference area can be located on either side of the line of indifference. So, each choice between two reduction programs delimits the area containing the most preferred sets of weighting factors. The survey included six choice questions. Each question included a trade-off between two damage categories. The third damage category was set equally for both programs in order to ease the task. Therefore, all six lines of indifference ran through the corners of the mixing triangle. Six such trade-off questions were posed, two for each pair of damage categories. The range of the reduction targets was chosen between 15% and 40% in order to distinguish between respondents for which the spread between the most and the least important damage categories exceeds a factor of 2.5 and respondents that assign weighs more equally (Fig. 2). A wider range of reduction targets could reveal more extreme weightings (according to Eq. 2). But, in order to limit the response time of the questionnaire, no additional questions were introduced. Implications of that selection are discussed at the end of Section 3.

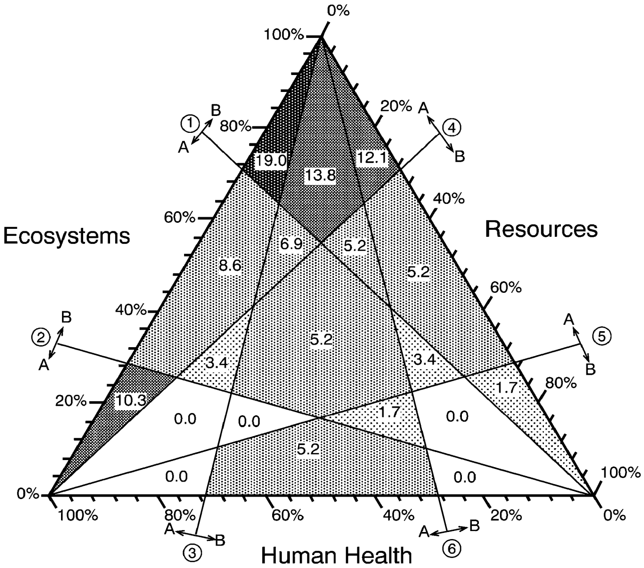

Fig. 2 Results from six choice questions (1 to 6). The figures represent the percentage of respondents whose preferences lie within the respective *shaded area*

2.3 Estimation model

The estimation model establishes a link between the difference of the outcomes of the programs and the observed frequencies of the program choices. The questionnaire included six different trade-offs. The differences in the reduction targets in every trade-off (see Fig. 1) are represented by a vector $x_i = \left(x_{i,\mathrm{HH}}, x_{i,\mathrm{EQ}}, x_{i,\mathrm{R}}\right)$, $i = 1, \ldots, 6$. For example, the vector $\boldsymbol{x}_i$ for the trade-off specified in Fig. 1 would be $\boldsymbol{x}_i$ = (25%, 0%, −10%).

The vectors of the differences between the reduction targets $\boldsymbol{x}_i$, can be conceived as the independent (or predictor) variables. The frequencies of the programs chosen are the dependent (or predicted) variables. In order to estimate the influence of the reduction targets HH, EQ, and R on the choice, we refer to regression analysis (i.e., the so-called linear model) and postulate that the frequency of the choices can be predicted as a linear combination of the differences in reduction targets x_{ik}, $k \in K = \{\mathrm{HH}, \mathrm{EQ}, \mathrm{R}\}$. However, instead of applying a linear regression (that can be used with a predicted variable that is continuous), we apply a logistic regression, i.e., a special case of a logit model (see, e.g., Hartung and Elpelt 1986, p. 132ff.).

In general, logit models are used to analyze the effect of categorical and continuous independent variables (the predictors, i.e., the differences in between the reduction targets of alternatives, the price, etc.) with respect to a categorical dependent variable (the response variable, i.e., the alternative chosen). Logit models are appropriate for choice experiments in order to monetize the value of environmental goods (see, e.g., Boxall et al. 1996) as the price can be linked with the other (environmental) attributes of the choice alternatives. As mentioned above, the predictors for our choice experiment are the differences between the attributes of the six programs ($x_{i,\mathrm{HH}}$, $x_{i,\mathrm{EQ}}$, and $x_{i,\mathrm{R}}$) and the dependent variable is the program chosen (A or B).

Logit models do not predict the choice frequency, which we will conceive as a probability p_i. Instead of the probabilities p, logit models utilize the log of the odds ratio (i.e., a logit transformation of p_i). The odds ratio is the probability of an event (favorable case) divided by the probability of a nonevent (nonfavorable case):

$$\mathrm{odds}_i = \frac{p_i}{1 - p_i}. \tag{3}$$

As an example, the odds ratio for a probability of 0.5 is 1:1, whereas it is 1:2 for a probability of 0.333. Following Eq. 3, odds can have values between 0 (for $p \rightarrow 0$) and ∞ (for $p \rightarrow 1$). If we take the log of the odds ratio, i.e., the *logit* (p_i), values between −∞ (for $p_i \rightarrow 0$) and ∞ (for $p_i \rightarrow 0$) result. Thus, the probability function (the choice frequency, which we want to estimate), which can have values between 0 and 1, is transformed into the logit function that has values between −∞ and ∞. For a binary response variable (only two alternatives can be chosen), the logit model is

Table 2 Results of the choice experiment and comparison with the allocation task

	Choice experiment				Allocation task[a]	Difference between choice experiment and allocation task and comparison of the mean weights in a t test (for D=0)			
	Results of the logistic regression								
Damage categories	β	SE	$\mathrm{Sig}_{\mathrm{reg}}$	Mean weighting factors	Mean weighting factors	Difference of mean (D)	95% confidence interval for D		$\mathrm{Sig}_{t\ \mathrm{test}}$
HH	0.031	0.012	0.013	0.22	0.28	−0.06	−0.026	−0.097	0.001
EQ	0.088	0.015	0.000	0.62	0.42	0.20	0.172	0.234	<0.000
R	0.022	0.013	0.088	0.16	0.30	−0.14	−0.107	−0.176	<0.000
	Percentage correct=70.0%								

[a] The weights derived from allocation questions are described in part 1 of this paper (Mettier et al. 2006) and are presented here again to illustrate the difference

equivalent to the logistic regression. Therefore, we use the logistic regression, which has the form:

$$\log\mathrm{it}(p_i) = \log\frac{p_i}{1-p_i} = \sum_{k\in K}\beta_k \times x_{ik} = \beta' \times x_i \qquad (4)$$

where K={HH, EQ, R} represents the damage categories and i=1,…,6 represents the pairs of programs presented for choice.

Thus, the logistic regression models the logit transformation of the ith (i=1,…,6) choice question's probability p_i (left part of Eq. 4) as a linear function of the regression coefficients β_k and the explanatory (dependent) variables in the vector $x_i = \left(x_{i,\mathrm{HH}}, x_{i,\mathrm{EQ}}, x_{i,\mathrm{R}}\right)$ (right part of Eq. 4). The regression coefficients β_k (in our case β_{HH}, β_{EQ}, and β_{R}) are also called β-weights. The β-weights represent the influence of the kth explanatory variable (in our case, the damage categories) on the choice. Thus, the regression coefficients β_{HH}, β_{EQ}, and β_{R} may be interpreted as reflecting the effects of the category indicators on the (logit) of a choice or on the underlying utilities of the alternatives. Therefore, these regression coefficients can be interpreted as weighting factors of the category indicators and will be transformed in the traditionally utilized weighting factors w_i that add up to 1 (or 100%) (see left side of Table 2).

If we solve Eq. 4 for p_i, we obtain:

$$p_i = \frac{\exp(\beta' x_i)}{1+\exp(\beta' x_i)}. \qquad (5)$$

With Eq. 5, we can calculate the probability for every alternative in all choice tasks (cases). All choices made by a single respondent are treated as independent observations. If the probability of program A is p_i>0.5, we assume that this program is chosen. If we assume that p_i<0.5, then A is not chosen but B instead. Comparing the predicted choice from the regression model with the real choice as made by the respondents leads to the percentage of correctly predicted cases by the regression model.

3 Results

As explained before, we asked six choice questions (marked with 1 to 6 in Fig. 2). Thus, there are six lines of indifference that separate the mixing triangles into 19 areas. Five respondents (9%) filled in the choice questions inconsistently, i.e., no preference area could be found.[3] For the other 52 respondents, preference areas are shown in Fig. 2. The allocation task yields precise sets of weights. Therefore, a comparison with the choice procedure cannot be performed directly. First, we analyzed whether the two methods produced consistent results. We, therefore, check that the preference field (the space of all possible weighting sets) derived from choice questions contained the allocated weighting set. For 27% of the respondents, the preference area contained the set of weights given in the allocation task. That means these respondents allocated weighting factors that matched with the preferences they showed in the six choice questions. Seventy-three percent showed a shift between the two tasks, as the figures they gave in the allocation task did not match with their choices. For those respondents, we analyzed whether the shift showed the trend we expected, i.e., the weights are closer to the middle than the preference areas. For 28 respondents (54%), weights are closer to the middle than the nearest point of the preference field. These respondents allocated weighting factors that are more equal than the preferences showed in the choice questions would predict. For 10 respondents (19%), the allocated weights are further from the center than the nearest point of the preference field. For these respondents, the allocation- and choice-based tasks show different results, but the direction of the shift is not determinable. In a statistical binomial test, the distribution between the respondents that show a shift toward more equal weights in the allocation task and those that do not is only significant for a significance level of α=0.1[4] (p=0.08). Nevertheless, the test gives us clear evidence as the criteria of the nearest point is the strongest criteria we can apply. If we choose, e.g., the center of gravity of the preference fields, only four respondents' allocated weights that are further from the center than the center of gravity can be considered. Thus, allocating weights too equally is the main reason for the preference shifts between allocation and choice task in this experiment.

In a second step, a logistic regression (see Section 2.3) of all choice questions is calculated (left side of Table 2). The regression coefficients β_k express the influence a damage category has on the choices of the whole sample and can, therefore, be interpreted as weighting factors for the damage categories. Thus, the regression coefficients (β values) have been used to calculate weights that add up to 100%. $\mathrm{Sig}_{\mathrm{reg}}$ denominates the probability that a category result has no influence on the alternative chosen (β=0), which means that the weighting factor is 0. These weighting factors calculated from the regression coefficients again reveal insights on the influence of the question format on the outcome of a valuation panel. Table 2

[3] Such an inconsistency could, for example, occur if a respondent chose program B in question 1 and program A in question 2.

[4] The α error denotes the probability of rejecting a null hypothesis (no difference between the choice and allocation task) when it is actually true.

includes a statistical figure that indicates the fit of the model with the data. Percentage correct is the percentage of correctly predicted cases in the model (see Section 3). For a binary outcome (program A or B is preferred), a random model would correctly predict 50%. As the correct percentage is 70%, the fit of the regression model with the answers obtained is satisfactory, though not exact.

Compared to the results obtained from direct weighting, there are big differences. It is obvious that the spread between the most and the least important damage category is much bigger (a factor of 4) for the choice task than for the direct allocation (a factor of 1.5).

The differences between the mean weights derived from choice task (C) and allocation task (A) are tested by a one-sample *t* test. The $\mathrm{Sig}_{t\ \mathrm{test}}$ values in Table 2 indicate highly significant differences. This analysis reveals the tendency of the respondents to value the damage categories more equally in allocation than in choice tasks. As highly significant differences are obtained from a medium-sized sample (n=58), the effect must be quite strong.

The discrepancy between the two procedures is considerable, especially for the most and the least valued damage category. We would expect an even bigger spread between the weights if the selected range of reduction targets was bigger, i.e., if the respondents were able to state wider trade-offs (see Section 2.2).

4 Discussion

The focus of this survey was to investigate the differences between the results from choice and allocation questions. For this purpose, a weighting procedure based on choice questions has been developed and has provided meaningful results.

Thus, choice experiments, often used for the monetary valuation of environmental goods, can also be applied in LCIA to elicit nonmonetary weighting factors and may become an interesting approach for the future.

Our hypothesis about the differing results of the two question formats could be verified: the spread between the weighing factors is highly sensitive to the question format used in a survey. In LCIA, most past panel studies have been based on allocation tasks and the spreading of the resulting weights has been quite low and, at most times, below a factor of 2.5 (Hofstetter and Mettier 2003). Similar results were found for the allocation procedure presented in part 1 of this paper. But the choice-based procedure applied in this study led to much bigger differences between the most and the least valued damage categories. Thus, the two procedures produce significantly different weighting factors for the same sample and the same data presented. The main reason for this finding may be the different cognitive processes underlying the different valuation procedures. For choice questions, we postulated that the lexicographic choice heuristic is of importance and was often applied by the respondents. For allocation tasks, the anchoring and adjustment heuristic has been supposed to be at work. The findings do not originate from the statistical treatment (logit analysis) used to calculate weights; a comparison of directly allocated weights with the preference areas in the mixing triangle reveals the same facts.

Regarding the differences between the two question formats, one may ask which question format is favorable, for example, more reliable, valid, or practicable. Reliability could be assessed comparing the results from several similarly framed studies. The validity of weighting factors, in contrast, cannot be proven in an experiment, as there is no external objective reference that we could compare the results to (since it often holds for characteristics measured in social sciences).[5] One can only argue about validating evidence. We think, though have not yet proven, that the choice task mimics real situations, is cognitively easier to perform, and provides more valid and reliable results. Choice situations involving goal conflicts and trade-offs are common in professional as well as everyday life, whereas rating and allocation situations are rare. This argument is in line with the findings of Sell et al. (2007) who argue that investigating the preferences by multiattribute analysis is appropriate to gain insight into the basics of the preference structure concerning how it is communicated by the decision maker, whereas choice experiments are closer to what people really do. Choice-based procedures have some more advantages. In general, they are not shaped in a way that they elicit the anchoring and adjustment biases. But they also have some important disadvantages. Choice-based methods pool data across all individuals and, as such, do not obtain estimates at the individual level. Therefore, studies involving different value positions, e.g., value characterization according to cultural theory (Mettier and Hofstetter 2004) or according to sustainability perspectives (Steen 2005), are harder to run as they need bigger samples. That means it is harder to handle value plurality in choice-based procedures. Moreover, it seems to be easier in rating-based procedures to work with large sets of categories, and their respective attributes, and to gain a differentiated insight into the preference structure. As mentioned in the introduction, score allocation procedures based on pairwise comparisons of categories (like the AHP; Saaty 1980) have been developed. These procedures account for the fact that humans have limitations on the number of criteria they can handle at the same time.[6] The allocation task is split into

[5] See the discussion about construct validity in Mettier (2006).

[6] See, for example, the work of Miller (1956) on cognitive limits of information processing.

different tasks that are easier to accomplish. The same can be done for choice-based procedures that include more than five or six categories. However, a complicated, aggregated design must be applied, which often becomes too complex for empirical studies. In aggregated designs, every choice task only contains a subset of all categories and final preferences are calculated based on relative preferences compared to a common category (in most cases, money). Thus, choice-based procedures with a larger set of categories can be conducted, but the setup is more complicated. Despite this drawback, we think that, in the context of LCA panel studies on eliciting weights, choice-based procedures should be favored for three reasons. Firstly, choice tasks are simple to conduct. Second, choice situations are a routinely practiced in everyday life, but allocation is not. Third, we think that choice tasks correspond more to the goal of many LCA studies to select among different products.

5 Conclusions and recommendations

As shown, the influence of the question format is considerable. Therefore we opt—if possible—to use more than one method as a kind of sensitivity analysis to check whether the results are robust. In many LCA studies, only little weighting information is needed in order to identify the best alternative(s), especially if endpoint indicators are applied. In these cases, the ranking of the alternatives only changes for extreme weighting sets. One only has to decide, for example, if one category indicator shall be weighted higher than 1% (see Hofstetter et al. 1999). In other cases, the ranking of alternatives is more sensitive to weighting information and interpretation should include a sensitivity analysis. For this sensitivity analysis, a reasonable variance of the weighting factors must be determined. We can conclude from our study that the ranking of the category indicators is not influenced by the question format and the hierarchy of the categories stays the same. But the spread between the most and the least weighted categories may depend on the question format and can be varied in a sensitivity analysis.

5.1 Some lessons learned

We will conclude this paper with some lessons learned from the weighting procedure described in part 1 (Mettier et al. 2006) and part 2 of this paper.

All weighted categories should be in the same order of magnitude and refer to the same reference (in space and time), for example, a percentage of normalization values. It is, for example, difficult to value a reduction of worldwide global warming against the reduction of species on a regional level. We are aware that defining a common normalization reference among all categories is especially challenging for midpoint indicators. Nevertheless, interpretation of midpoint category indicators depends on such a common normalization reference in order to comprehend the relevance of a category result.

Quantitative data provided may not have a big influence on the expressed preferences as many LCA stakeholders insufficiently process quantitative data. This is different for some experts who can link the data to prior knowledge.

The qualitative descriptions of the valued categories seem more determining. In this study, we emphasize the model structure. That means it is important to indicate which and how many environmental problems contribute to a damage category.

If only one type of valuation task can be included in a procedure—and no individual or subgroup assessments are required—we favor a choice-based procedure for its simplicity and practice. We, therefore, recommend choice experiments for future LCA panel studies because of the simplicity, the routine of the respondents, and the match with the goals of LCA studies to select among different products.

As shown in part 1 and II of this paper, the framing of the context and the valuation task can have a significant influence on the results. These findings reveal the constructive nature of stated preference procedures. But this should not form an obstacle to apply these procedures, but provides a challenge to search for an appropriate framing. We hope that this article can contribute to this aim. Finally, the framing and interpretation of such future choice experiments for LCA panel studies can benefit from the experience and vast literature of economical and medical research.

References

Alpízar F, Carlsson F, Martinsson P (2001) Using choice experiments for non-market valuation. Econ Issues 8(1):83–110

Boxall P, Adamowicz W, Swait J, Williams M, Louviere J (1996) A comparison of stated preference methods for environmental valuation. Ecol Econ 18(3):243–253

Carlsson F, Frykblom P, Liljenstolpe C (2003) Valuing wetland attributes: an application of choice experiments. Ecol Econ 47:95–103

Christie M, Hanley N, Warren J, Murphy K, Wright R, Hyde T (2006) Valuing the diversity of biodiversity. Ecol Econ 58(2):304–317

Colombo S, Calatrava-Requena J, Hanley N (2006) Analysing the social benefits of soil conservation measures using stated preference methods. Ecol Econ 58(4):850–861

Dhami MK, Hertwig R, Hoffrage U (2004) The role of representative design in an ecological approach to cognition. Psychol Bull 130 (6):959–988

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 HA, Finnveden G, Goedkoop M, Hauschild M, Hertwich E, Hofstetter P, Jolliet O, Klöpffer W, Krewitt W, Lindeijer E et al (eds) Life-cycle impact assessment: striving towards best practice. SETAC, Brussels

Goedkoop M, Spriensma R (1999) The Eco-indicator 99—a damage oriented method for life cycle assessment: methodology report. PRé Consultants B.V., Ammersfoort, The Netherlands

Green PE, Srinivasan V (1990) Conjoint analysis in marketing—new development with implications for research and practice. J Mark 54(4):3–19

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 International Conference on EcoBalance, Tsukuba, Epochal Tsukuba, Japan

Hartung J, Elpelt B (1986) Multivariate Statistik: Lehr- und Handbuch der angewandten Statistik, [Multivariate statistics: teaching book and handbook of applied statistics], 2nd edn. Oldenbourg, Munich

Hearne R, Salinas Z (2002) The use of choice experiments in the analysis of tourist preferences for ecotourism development in Costa Rica. J Environ Manag 65:153–163

Hofstetter P, Mettier T (2003) What users want and may need: insights from a survey among users of a life cycle tool. J Ind Ecol 7 (2):79–101

Hofstetter P, Braunschweig A, Mettier T, Müller-Wenk R, Tietje O (1999) The mixing triangle: graphical decision support for comparisons in LCA. J Ind Ecol 3(4):97–115

Hutchinson JMC, Gigerenzer G (2005) Simple heuristics and rules of thumb: where psychologists and behavioural biologists might meet. Behav Processes 69:97–124

Irwin JR, Slovic P, Lichtenstein S, McClelland GH (1993) Preference reversals and the measurement of environmental values. J Risk Uncertain 6:5–18

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 Life Cycle Assess 9 (3):196–205

Kahneman D, Slovic P, Tversky A (1982) Judgment under uncertainty: heuristics and biases. Cambridge University Press, New York

Lehtonen E, Kuuluvainen J, Pouta E, Rekola M, Li C-Z (2003) Non-market benefits of forest conservation in southern Finland. Environ Sci Policy 6:195–204

Lindeijer E (1996) Normalisation and valuation. In: Udo de Haes HA (ed) Towards a methodology for life cycle impact assessment. SETAC Europe, Brussels

Mettier T (2006) Panel methods in LCIA: procedures to support normative choices in life cycle impact assessment. Dissertation No. 16443, ETH, Zurich

Mettier T, Hofstetter P (2004) Survey insights into weighting environmental damages: influence of context and group. J Ind Ecol 8(4):189–209

Mettier T, Scholz RW, Tietje O (2006) Measuring preferences on environmental damages in LCIA. Part 1: cognitive limits in panel surveys. Int J Life Cycle Assess 11(6):394–402

Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81–97

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, Japan

Puolamaa M, Kaplas M, Reinikainen T (1996) Index of environmental friendliness: a methodological study. Statistics Finland, Helsinki

Saaty TL (1980) The analytical hierarchy process: planning, priority setting, resource allocation. McGraw-Hill, New York

Sangle S, Ram Babu P, Khanna P (1999) Evaluation of life cycle impacts: identification of societal weights of environmental issues. Int J Life Cycle Assess 4(4):221–228

Sculpher M, Bryan S, Fry P, Winter P, Payne H, Emberton M (2004) Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. Br Med J 328:382–383

Sell J, Koellner T, Weber O, Pedroni L, Proctor W, Scholz RW (2007) Ecosystem services from tropical forestry projects—the choice of international market actors. Forest Policy and Economics 9 (5):496–515

Seppälä J (1999) Decision analysis as a tool for life cycle impact assessment. Ecomed, Landsberg, Germany

Shaw J, Giglierano J, Kallis J (1989) Marketing complex technical products—the importance of intangible attributes. Ind Mark Manage 18(1):45–53

Steen B (2005) Describing values in relation to choices in LCA. Int J Life Cycle Assess 11(4):277–283

Tversky A (1972) Elimination by aspects: a theory of choice. Psychol Rev 79:281–299