

HILCA: A NEW CONJOINT PROCEDURE FOR AN IMPROVED PORTRAYAL OF PURCHASE DECISIONS ON COMPLEX PRODUCTS

*by Raimund Wildner, Holger Dietrich and Ansgar Hölscher**

ABSTRACT

This paper introduces HILCA, a new conjoint analysis procedure with which purchasing decisions on complex products can be better modelled. The survey and analysis procedures are described in detail. The procedure is compared in a methodology test with a standard conjoint analysis procedure used for such research issues. It produces with roughly the same length of interview much more valid and easier to interpret results.

* *Dr. Raimund Wildner is Vice-President and Managing Director of GfK-Nürnberg e.V. and also Manager of Methodology and Product Development of the GfK Group, Nordwestring 101, 90319 Nuremberg, Germany, tel. +49 (0) 911 3 95-25 73, email: Raimund.Wildner@gfk.com.*

Holger Dietrich, graduate statistician, is Manager of Basic Research at GfK-Nürnberg e.V., Nordwestring 101, 90319 Nuremberg, Germany, tel. +49 (0) 911 3 95-20 20, email: Holger.Dietrich@gfk.com.

Ansgar Hölscher is Associate Principal at McKinsey & Company in Hamburg. He manages the Customer Insights Function within the European Marketing Practice, McKinsey & Company Inc., Am Sandtorkai 77, D-20457 Hamburg, Germany, tel. +49(0)40 36 12 13 14; fax: +49(0)40 36 12-13 15; email: Ansgar_Hoelscher@mckinsey.com

Originally published in GfK Jahrbuch der Absatz- und Verbrauchsforschung, Vol. 52 (2006), No 4, pp. 332-348.

1. Research Issue

The configuration of products and services is one of the most important parameters within the marketing mix. How many revolutions per minute should a washing-machine have, what hard disc capacity should a notebook have? Should a giro account be provided with a debit card facility and with a credit card facility as well? Should a bank account have a standard monthly fee, or should there be a fee for each transaction? Will a life assurance policy be more profitable if it is provided with employment protection or not? The answers to these and similar questions affect not only the costs of the product-range, but also – either directly or indirectly due to changes in the willingness to pay – its profitability. Correct answers can therefore be critical to the profitability of a product (1).

Direct questioning about preferences for attribute levels and their importance which used to be in common use for this purpose has not proved to be successful. It has regularly led to inflation of requirements, in other words claiming that all positive features are important and also in particular, an attractive price. This inevitably leads to the recommendation that high-value products at an attractive price will meet with a high level of customer acceptance. This is neither surprising nor particularly helpful.

On the other hand, **conjoint analysis** (Hartmann/Sattler 2002), which was developed in psychology as early as 1964 by Luce and Tukey, but which only became well-established in market research from the 1980s, has the advantage that when making a purchase decision, the interviewee must as in the real world weigh up the various attributes against each other in order to be able to determine the **total utility** of a product. He therefore decides whether he prefers a better specified but more expensive product or a less well specified and cheaper one. The result for each attribute level is a measured value of the preference, named the **part-utility value**.

With conjoint analysis it is assumed that the part-utility values of a product add up to its total utility value. It also assumes that with the total utility value, the probability rises that the interviewee will buy the product. It therefore unites the preferences for individual attribute levels with prospective buying behavior.

The early forms of conjoint analysis include profile and trade-off methodology. With the **profile methodology**, complete features of a product are shown for assessment. With the **trade-off methodology** combinations of features are shown. In both cases, the products are to be ranked by the respondents in terms of their order of preferences. This makes it possible in a first stage to estimate total utility values and from them part-utility values. Neither of these procedures was able to become established in practice. Two newer conjoint analysis procedures, i. e. Choice Based Conjoint (CBC) and Rating Based Conjoint analysis (RBC) have primarily established themselves in practice.

■ Choice Based Conjoint

With Choice Based Conjoint (CBC), the interviewees are repeatedly shown (e.g. 20 times) a set comprising a number of product concepts on which full details are provided. They then select the concept that they would buy. They also have the option of selecting none of the concepts.

CBC offers the **advantage** that the research approach to a large extent reflects the type of task that has to be completed when products are being bought in real life. Even if the choices made are of a hypothetical nature, the results can therefore be regarded as particularly valid (2). A **disadvantage** is that the questioning is not efficient, because in each exercise only very little information is collected. Since the number of choices to be made rises very sharply with the number of attributes, a traditional CBC only can handle five or six product characteristics or attributes. Additionally, the information collected from each interviewee is not sufficient to generate part-utility values per interviewee; they can only be generated for groups of interviewees in total. But individually calculated part-utility values are nevertheless required if interviewees are to be allocated to segments with different preference structures, so that the manufacturer can offer different products that are targeted at different market segments.

More recent developments use the **Hierarchical Bayes (HB)** methodology to generate individual part-utility values (*Sawtooth* 2005). Additionally, using HB the number of product attributes covered can be increased to up to ten. But for complex products this is not enough. Moreover, in addition to individual data, HB also uses information about the interviewee group as a whole. The results are therefore not fully available at the individual level.

■ Rating Based Conjoint

With Rating Based Conjoint various product attributes and/or complete product concepts are compared together. **ACA** (Adaptive Conjoint Analysis from *Sawtooth Software Inc.* (2005)) has emerged as the market leader, because software is available for it, with which both the data collection and the analysis can be easily undertaken.

With ACA, the **survey** goes through the following stages:

- (1) Identification of the **Unacceptables**: They are understood as attribute levels, which would not be acceptable under any circumstances. If for instance, the issue is the only car in the household, then for “the number of seats“ attribute, a “two“ seat level would not be acceptable for a family with children.
- (2) **Ranking the attribute levels**: Are four or five seats preferred?
- (3) Details on the **importance of the attributes**, although that depends on the levels that are offered. It could be for instance that the number of seats is not important

if only the levels “four” or “five” are offered, but that it is very important if the levels offered are “two”, “four” and “five”.

- (4) **Comparison pairs** with more complex concepts: There is an increase in the number of attributes per concept. The selection of attributes and their levels depends on the answers previously provided (for this reason the procedure is described as „adaptive“).

The **main advantage** of Rating Based Conjoint is that the required data can be collected very efficiently. As a result, in theory ACA can handle up to 30 attributes, and in practice up to 20. A further advantage is that the part-utility values can also be calculated on an individual basis. This makes possible a segmentation of interviewees into groups with similar preference structures.

A **disadvantage** is that the ACA methodology has little similarity to a genuine purchase decision. Additionally direct questioning on Unacceptables can easily lead to individual levels being too readily categorized as such. Often, the interviewee wants to say that he regards a level as very unsatisfactory, rather than it makes a purchase totally impossible for him. Both these issues lead to the methodology producing results of little validity (3).

To summarize: firstly, with **CBC** a methodology is available which combines a high level of validity but on a **limited number of attributes**. This methodology is therefore not suitable for complex products with a very large number of different attributes. It also permits only very limited market segmentation capabilities. With **ACA** on the other hand, a methodology is available which in fact can handle up to twenty attributes at the individual level, but which produces **results with very little validity**. What is required is a procedure which can handle a large number of attributes at the individual level (where twenty attributes for many research projects such as cars are insufficient) *and* at the same time generates valid results.

This paper introduces in the following Section 2 a **new type of procedure** which **combines** together these **desirable attributes**. The basic principles are described in 2.1, the data collection in 2.2, and the analysis in 2.3. Section 3 describes a test of the methodology together with test design (3.1) and results (3.2), while the short Section 4 summarizes the results.

2. Describing the HILCA methodology

2.1 Basic principles

The objective for the development of the HILCA conjoint analysis methodology was to generate valid results from a large number of attributes. The abbreviation stands for „*Hierarchical Individualised Limit Conjoint Analysis*“.

The starting point for the development was the fact that a category is often characterized by a very large number of attributes, but that for every single buyer of a category only a few of the attributes strongly influence the purchase decision on a product, and that these important attributes differ from person to person.

This can be made clear by taking the **purchase of a notebook** as an example: people who want to use the notebook for writing long memos on business trips, will primarily be concerned with battery life, perhaps also with weight, but will be less concerned with processing capacity or with graphic card quality. On the other hand the screen-size and the graphic card quality will be important for somebody who mainly uses his notebook for computer games. Finally a consumer who uses his notebook primarily as a mobile DVD-player, will above all be concerned with the brightness of the screen and also with battery life.

Previous conjoint analyses have handled all attributes equally. But if attributes can be important or unimportant in the interviewee's buying decision, then it makes sense to proceed **hierarchically**, in other words to create an initial broad-brush overview of the attributes, and then to go into detail on the more important ones. As the important attributes differ from person to person, this has to be done on an **individualised** basis. This also makes the interview more interesting, since the interviewees are not questioned in detail about attributes which they regard as unimportant and which therefore do not interest them.

Finally it is assumed that a product must have a certain appeal, if it is to be bought. This is achieved by including a **limit card** in the interview. With this methodology, which was developed by *Voeth/Hahn* in 1998, the interviewees are shown a number of products arranged in order of attractiveness; they are asked to insert a limit card in the series, at the point where they would in principle be willing to consider buying all products above the limit card, but none of those below it.

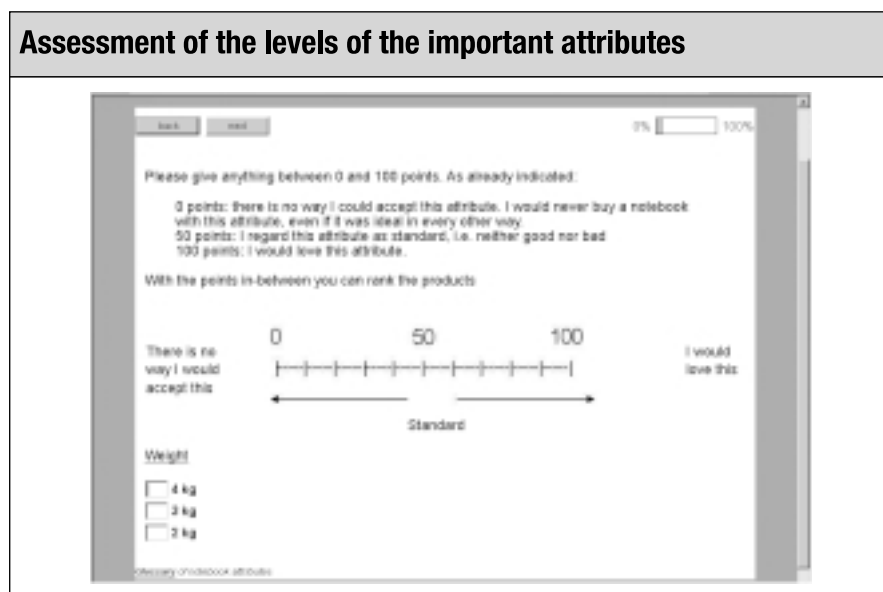
The basic characteristics of HILCA were initially developed by *Voeth* (2000) who, working closely with *McKinsey* (*Ansgar Hölscher*) and *GfK* (the first two authors of this paper) together brought the product to maturity. An empirical test with a pilot client proved the validity of the procedure (cf. Section 3 of this paper).

2.2 Data collection

The survey is conducted in the following stages:

- (1) **Selection of the attributes**, which have a role in the purchase decision: at this stage, all the attributes with their various levels are shown on a screen. The interviewees state for each attribute whether they are important to them or not, when making their buying decision.
- (2) **Assessment of the attribute levels** in terms of preference using a scale from 0 („no way I could accept this“) to 100 points („I would love this attribute“). For this purpose, each attribute is shown on a screen with all its levels and with the scale (cf. *Figure 1*). The proportion of the interview completed so far can always be seen (*Figure 1* top right) and a glossary of the terms employed is also available (left below). If zero points are given to an attribute level, then it is interpreted as Un-acceptable.
- (3) **Selection of the five most important attributes** from the attributes set out in Stage 1: For this purpose, all the attributes with their different levels and the scores given are shown. The interviewees select their five most important attributes.

Figure 1



Five attributes are selected, because various research projects have shown that generally people are only able to compare about five attributes simultaneously (cf. for instance Miller 1956). It can therefore be assumed for the purpose of assessing alternative products, that initially about five especially important attributes are used.

- (4) **Assessment of concepts:** here a maximum of 25 product concepts with varying levels of the five most important attributes are shown. The selection of the concepts is done with the „Orthoplan“ procedure from SPSS, that generates what are called „orthogonal designs“ (SPSS 1991), with the levels of the attributes appearing independently of each other so that the evaluation of the attribute levels can follow from the evaluation of the concepts. In practice the possible orthoplan designs were generated in advance and saved on the hard disk of the notebook being used for interviewing purposes. The relevant design (depending on the number of levels within the five most important attributes) is selected by the questionnaire software. The concepts shown are to be evaluated by the respondent on a scale from 0 („The

Figure 2

Assessment of concepts

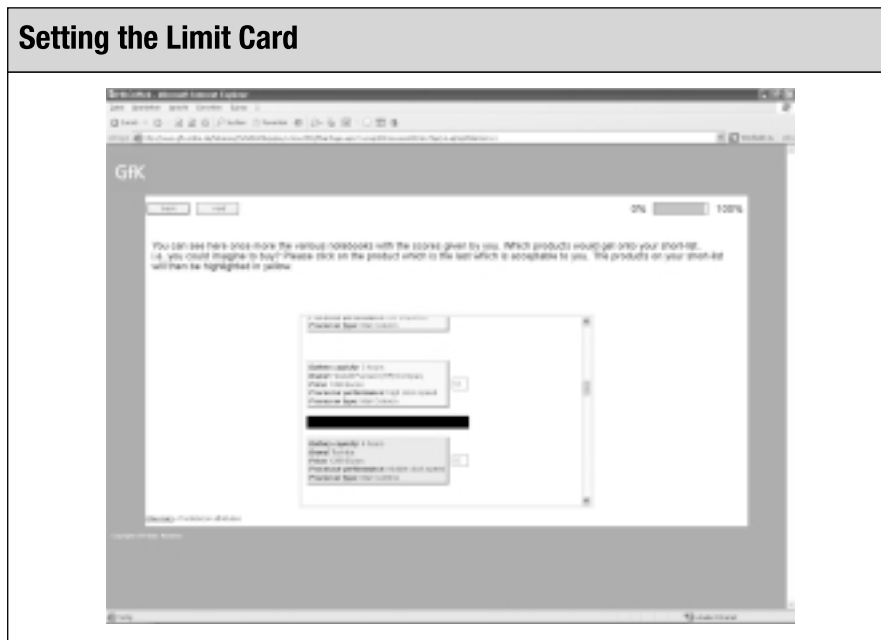


product is totally at odds with my wishes“) to 100 points („The product fits in perfectly with my wishes“).

For this purpose, initially three concepts are shown that are constructed so as to cover a large proportion of the total utility. For this, the evaluation of the attribute levels is used (see step 2). A concept that has been evaluated by the respondent is moved to the right side of the screen and shown together with all the other concepts already evaluated in ranked order. Then another concept is shown on the left side of the screen, which has to be evaluated in the same way and is moved to its place in the ranking of concepts on the right side of the screen (cf. Figure 2). This is repeated up to 25 times. In order to avoid biases from ordering effects, concepts are shown in a randomized order. At any time the respondent can modify evaluations of former concepts.

(5) In the final stage, the complete list of sorted concepts is shown and the respondents are asked to click on the concept which is the worst they could imagine considering for purchase. This step is called “setting the limit card”. A black bar appears

Figure 3



below the final concept. The concepts above the black bar are highlighted in a different color. (cf. *Figure 3*).

It is worth noting that the questionnaire is not developed with one of the standard software packages designed for producing computer-assisted questionnaires. Stage 4 in particular cannot be handled by such software packages. So a special program had to be written which is controlled by the standard software; it performs the operation and returns control to the standard software when the task is completed.

It seems more important that this research procedure is **completely transparent at every stage to the interviewees**. It will always be clear why one attribute is used and another not. This leads to a better acceptance of the questionnaire and in the consequence to a higher quality of answers.

Additionally, in the whole questionnaire **only one scale**, i.e. the 100-point scale, is used. This has proved to be particularly useful in the identification of the **Unacceptables** (Stage 2), because with a scale that has such a large range, the extreme score of 0 points is less used for non-genuine Unacceptables. This leads to a further benefit of HILCA in comparison with ACA: While with ACA the number of Unacceptables per attribute is limited, so that at least two levels per attribute are needed, HILCA accepts what are called „**Must Haves**“, i. e. attribute levels a product has to have if the product is to be acceptable.

2.3 Analysis

HILCA distinguishes between three different **groups of attributes** the utilities of which are estimated differently:

- **Unimportant attributes** that were explicitly identified as such in Stage 1. In the test survey (cf. Section 3) on average 12.7 out of 19 attributes were rated unimportant.
- **Important attributes** that were identified as such in Stage 1. For their attribute levels, scores are available from Stage 2. In the test survey (cf. Section 3) on average 1.5 of the 19 attributes were rated important.
- **Very important attributes** that were identified as such in Stage 3: for their attribute levels scores are available from Stage 2. Additionally utilities per attribute level can be computed from the scores given to the total concepts in Stage 3. 4.8 out of the 19 attributes on average were very important (4).

Not surprisingly the allocation of an attribute to one of these groups differs from person to person. The different levels of importance correspond with the **different estimating procedures** for the utility values that are done at an individual level:

- Utilities of the very important attributes are computed on the basis of the scores for all the concepts (Stage 4 of the interview) since the simultaneous evaluation of these attributes comes closest to the actual buying process.
- Utilities of the important attributes are estimated on the basis of the scores given in Stage 2. For the estimation, the evaluation of the very important attribute levels is used, with their scores from Stage 2 as an independent variable and the utilities estimated from the evaluation of the total concepts in Stage 4 as the dependent variable. The co-efficient of determination in the methodology test was 57 %. This regression is then applied to the scores of the important attributes as independent variables resulting in the estimated utilities.
- No utilities are estimated for unimportant attributes. These are assumed to be zero.

The calculations are done by GfK. After the estimation of the utilities has been completed, they are transferred to a **Simulation Program**. This makes available to the client a large variety of analyses that can be easily undertaken. *Figures 4 and 5* provide

Figure 4

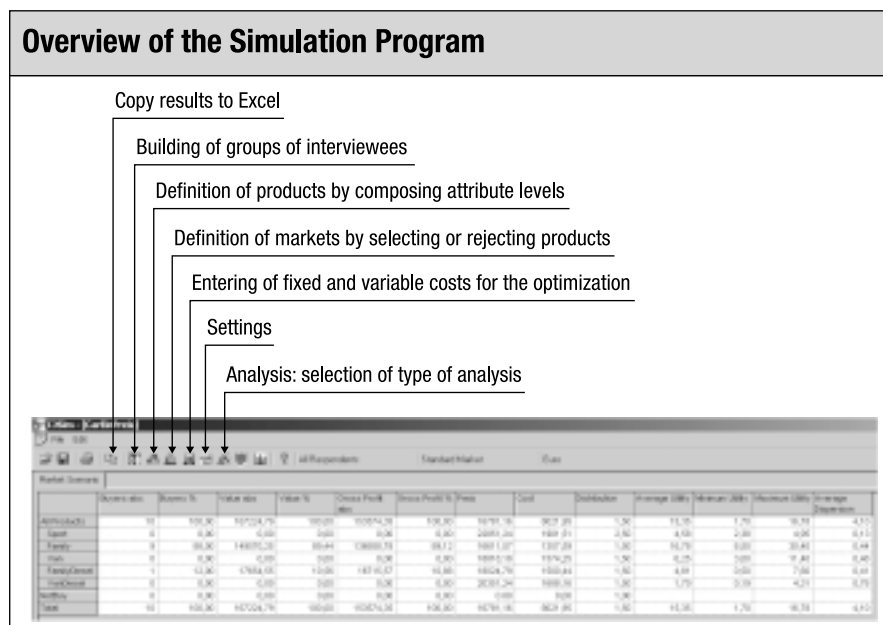
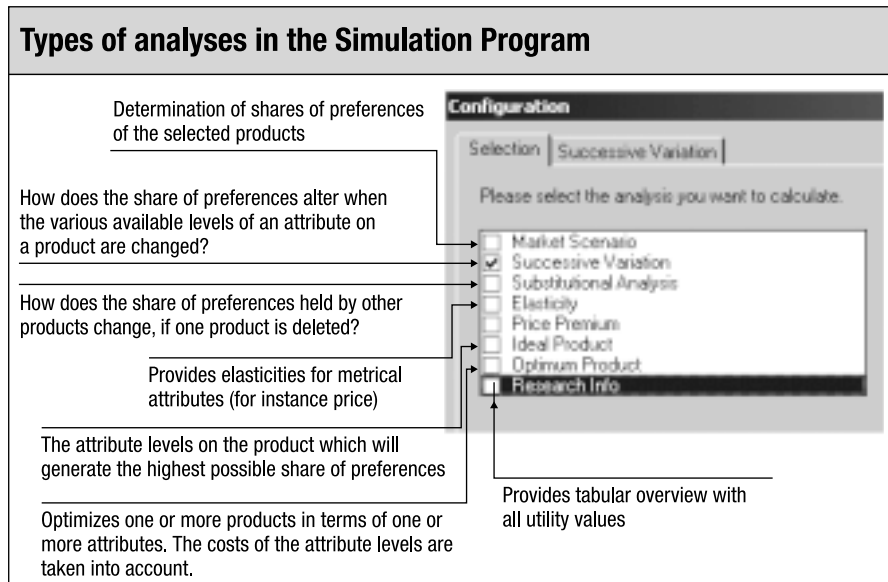


Figure 5



an overview. The program is usually shown to the client by the market research company consultant in a workshop after the final presentation and it can then be used by the client.

The Simulation Program is very important for two reasons:

- It **speeds up** the analysis. If additional analyses are required, these generally lead to a telephone call being made to the market research company so that they can make the calculations and report back to the client. Just the communications required for this will take more time than the computer simulation.
- Additionally **analyses become feasible**, which up to now were impossible. For instance, clients can themselves enter the costs for producing attribute levels, which because of security considerations cannot be passed on to other people. This cost data can be used to optimize products in terms of their profitability. For this purpose, a non-linear numerical optimization procedure is included.

3. Test Survey

3.1 Test Design

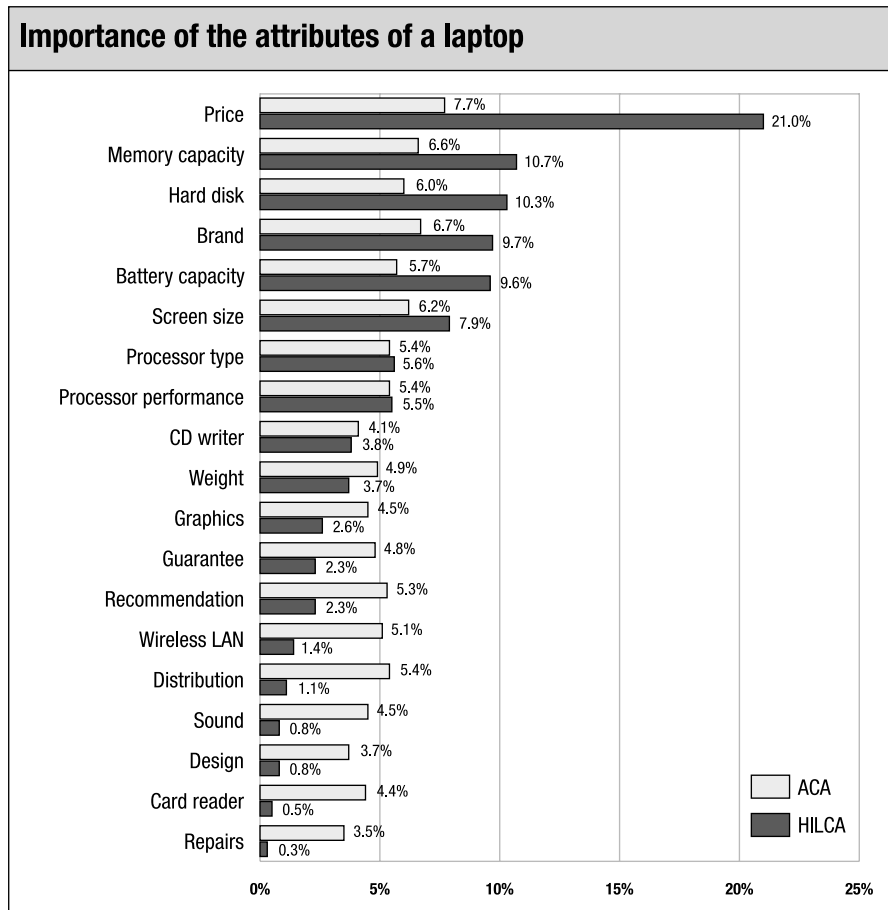
Theoretical considerations are important and provide the foundation for a good research tool. Its practical usefulness can though only in the final analysis be proved by an empirical test in which results from the HILCA procedure are tested against the methodology in general use up to now, i.e. ACA.

For this purpose **two groups with identical profiles** were formed. One group comprising 367 people was interviewed using the **HILCA methodology**. A second group comprising 387 people was interviewed using the **ACA methodology**. ACA was chosen as the basis for comparison because it is widely used and because with the large number of attributes to be covered in the research, it was not possible to use CBC. The survey took place between November 2005 and January 2006. It was designed with all those involved in the project working closely together with Fujitsu Siemens and it was carried out by GfK.

The research issue for the methodology test was the **purchase of a notebook**. In the first part of the CAPI questionnaire it was established whether the interviewee belonged to the target group. The **target group was defined as those persons** who in the last two years had bought a notebook or planned to do so in the next six months. This ensured that the interviewees were to some extent personally involved. General questions on buying a notebook followed. Then the **conjoint questionnaire** formed the main part, alternating between ACA and HILCA. It covered 19 attributes with a total of 64 levels. The attributes comprised the technical design and availability of the notebook (cf. *Figure 6*). Finally both groups were given three „**Hold-Out-Tasks**“ as they are called. In each of them three realistic notebooks, described in full detail, were presented to the interviewees and they were asked whether they would buy one or none of them, and if so, which.

The purpose of these Hold-Out-Tasks was to check the validity of the results. The choice made (one of the three notebooks or none of them) in each of the Hold-Out-Tasks was interpreted as a „*decision to purchase*“. On the basis of the results of the conjoint analysis it could also be predicted for each Hold-Out-Task, how the interviewee would decide by taking into account his estimated utilities: if for instance the utility of none of the shown products exceeded the threshold from the limit card question, it was predicted that the interviewee would select none of the products; otherwise it was predicted that the product with the highest total utility would be selected. A comparison of the predicted purchase decision with the actual decision generated what is called the **hit-rate quota**, i.e. the proportion of decisions that were correctly predicted.

Figure 6



It is important to know how both methods predict purchase or non-purchase decisions. HILCA explicitly tries to predict non-purchase decisions; therefore in the prediction non-purchase was allowed for. On the other hand ACA assumes that in principle a purchase decision will be made. In order to be able to predict a non-purchase, it was assumed that a product with non-acceptable attribute levels would not be bought. Accordingly, if all three products had one or more unacceptable attribute levels, a non-purchase was predicted.

In evaluating the hit-rate quota, it is necessary to keep in mind that with a total of four choices (purchase of one of three notebooks or a non-purchase), a prediction based on chance would generate a hit-rate quota of 25 %.

3.2 Test Results

Firstly in terms of **interview length**, times were comparable but showed a slight advantage for HILCA. The HILCA interview lasted on average 35 minutes, the ACA interview 37 minutes (cf. *Table 1*). There were differences though in the structure of the interviewing time: while with ACA all the attributes were handled in the same way, with HILCA more time was given to the important attributes.

There were significant differences in the **quality of the predictions**. While with HILCA 48.0 % of the purchase decisions in the Hold-Out-Tasks were predicted correctly, with ACA it was only 36.8 %. That means that ACA was less than 12 percentage points better than the chance result of 25 %. But even the percentages achieved by HILCA are really more a reason for undertaking further research than for satisfaction. Nevertheless, the significant better performance by HILCA compared with that by ACA can be regarded as a success: in an earlier version, which has since been optimized, the performance of ACA and HILCA were still similar in this respect (*Kraus 2003, p. 214*).

HILCA also shows significantly **fewer problems with the Unacceptables**. With HILCA, only 30.5 % of the products selected in the Hold-Out-Tasks contained one or more levels of attributes, which were declared as a result of the award of zero points in the interview to be not acceptable. With ACA, 88.2 % of the selected products contained an unacceptable attribute (cf. *Table 1*).

Table 1

Key Numbers for Methodology Comparison		
	HILCA	ACA
Average interview length	35 min.	37 min.
Prediction quality	48.0%	36.8%
Conflicts with unacceptables	30.5%	88.2%
Number of interviews	367	387

The results from HILCA are not only in formal respects superior to those of ACA; they are also regarded by the pilot client as significantly **more plausible** and **more informative** than the results from ACA. For instance, in terms of their relative importance the HILCA results were perceived as much more differentiated than those from ACA and were regarded as more plausible from a practical point of view. In particular, the underestimation of the importance of price which has been criticized with ACA, has been corrected (*Sawtooth* 1996 and *Figure 6*).

HILCA also provides **more information** than other procedures. Must-have levels are only possible with HILCA. For instance for some of the interviewees it was essential that a new notebook provided Wireless LAN or was of a particular brand. These attribute levels form a sort of threshold which must be surmounted before a purchase can even be considered in this product group, and that is not adequately handled with ACA.

4. Conclusion

With HILCA a **new conjoint analysis procedure** is available, which makes it possible to take account of a **large number of attributes** and which generates significantly **more valid results** than previous procedures. It is particularly suited to more complex products, where the purchase decision is made in an extended process, but not so suitable for products which are bought spontaneously or which have few attributes. In such cases, CBC is an appropriate tool. It is also important that with HILCA all estimates are made on an **individual basis**, which is precisely what makes a segmentation of the interviewees possible. These **advantages** do not entail the disadvantage of a longer interview.

In addition to information about the design of the physical attributes of a product, HILCA also generates inputs to **operational marketing**. Price must primarily be mentioned here; its importance is not, as with ACA, underestimated. Additionally, elasticities and the willingness to pay for individual product characteristics can be generated. Finally all the information outlined here can be very easily accessed via a **simulation program**.

Nevertheless, there is a **need for further improvement**. Even if the hit-rate quota of 48 % at the validation stage **represents** a significant improvement on what had been previously achieved, it still cannot be regarded as satisfactory. Work is currently proceeding to achieve additional improvements through further modifications to the data collection and by using different calculations of *utilities*.

Notes

- (1) In the following text, the concept of the “product” covers both physical products and also services.
- (2) For discussion, see *Brzoska* (2003), p. 140 ff.
- (3) The distributor of the software himself wrote as follows: “When price is just one of many attributes, ACA may assign too little importance to it”; *Sawtooth* (1996).
- (4) Attributes with a Must-Have level (that means attributes where all levels except for one are Unacceptables), are excluded here. Although each interviewee has selected five very important attributes, the average is therefore slightly lower.

References

- Backhaus, K./Erichson, B./Plinko, W./Weiber, R.* (2003): *Multivariate Analysemethoden [Multivariate Analysis Methodologies]*, Berlin/Heidelberg.
- Backhaus, K./Voeth, M./Hahn, C.* (1998): *Limit Conjoint Analyse*, Arbeitspapier [Working Paper] Betriebswirtschaftliches Institut für Anlagen und Systemtechnologie der Universität Münster, Münster.
- Brzoska, L.* (2003): *Die Conjoint Analyse als Instrument zur Prognose von Preisreaktionen [Conjoint Analysis as a Tool for Predicting Reactions to Prices]*, Hamburg.
- Hartmann, A./Sattler, H.* (2002): *Commercial Use of Conjoint Analysis in Germany, Austria and Switzerland*, Research Papers of Marketing and Retailing, University of Hamburg, Hamburg.
- Kraus, H.* (2003): *Preissetzung im Aktienfondsgeschäft auf Basis von Erkenntnissen über das Kauf- und Preisverhalten von Fondsinvestoren – eine empirische Analyse der Investitionsentscheidung von Privatkunden mit Hilfe der Conjoint-Analyse*, Dissertation Universität Bayreuth [Pricing in the Share Market on the Basis of Experience with Behavior in Terms of Purchasing and Attitudes to Pricing by Investors in Bonds – an Empirical Analysis on Investment Decisions by Private Customers Using a Conjoint Analysis; Dissertation from the University of Bayreuth], Bayreuth.
- Luce, R.D./Tukey, J.W.* (1964): *Simultaneous Conjoint Measurement*, in: *Journal of Mathematical Psychology*, No.1, pp. 1–27.
- Miller, G.A.* (1956): *The Magical Number Seven, Plus or Minus Two: Some Limits On Our Capacity for Processing Information*, in: *The Psychological Review*, No.63, pp. 81–97.
- Sawtooth (Ed.)* (1996): *Staying out of Trouble with ACA*, sequim (available from www.sawtoothsoftware.com).
- Sawtooth (Ed.)* (2005): *The CBC/HB System for Hierarchical Bayes Estimation, Level 4.0 Technical Paper*, sequim (available from www.sawtoothsoftware.com).
- SPSS (Hrsg.)* (1991): *SPSS® Statistical Algorithms*, 2nd Edition, Chicago.
- Voeth, M./Hahn, C.* (1998): *Limit-Conjoint-Analysis*, in: *Marketing, Zeitschrift für Forschung und Praxis*, Nr. 2, pp. 119–132.
- Voeth, M.* (2000): *Nutzen-Messung in der Kaufverhaltensforschung [Utility Measurement in Research into Buying Behavior]*, Wiesbaden.