

# **An Empirical Comparison of CBC and AHP for Measuring Consumer Preferences**

Martin Meißner and Reinhold Decker  
Department of Business Administration and Economics  
Bielefeld University  
33615 Bielefeld, Germany  
E-mail: {mmeissner, rdecker}@wiwi.uni-bielefeld.de

## **ABSTRACT**

Conjoint Analysis (CA) is a very popular class of methods for measuring consumer preferences, both in research and practice. However, since a couple of years, the Analytic Hierarchy Process (AHP) is being discussed in this field as well. Several empirical studies have shown the general potential of AHP, particularly in complex product evaluation tasks consisting of many attributes. But, in spite of its promising results, marketing practice seems to ignore AHP so far. This is somewhat astonishing if one considers the closeness of both approaches. Already almost thirty years ago Wind and Saaty (1980, p. 657) stated: “In some cases both the AHP and Conjoint Analysis can be used, and it is desirable to compare the results of the two approaches ...”. This paper contributes to fill this gap.

In order to increase the practical relevance of such a comparison for marketing practitioners, we run AHP against the commercially successful Choice-Based Conjoint Analysis (CBC) in an online survey. We outline how both the hierarchy and the scale used in AHP can be adapted to preference measurement and how Harker’s (1987) technique for incomplete pairwise comparison matrices can be applied in a beneficial way. Our results indicate that, indeed, both methods are equivalent with regard to convergent validity and the individual prediction of holdout tasks, but AHP significantly outperforms CBC in market share predictions.

Keywords: AHP, CBC, preference measurement, incomplete pairwise comparison matrices, market share prediction

## **1. Introduction**

Over the past three decades Conjoint Analysis (CA) has been marketers’ favorite methodology for eliciting consumer preferences. Ranking, rating and choice based types of CA have been developed as well as different techniques to estimate part-worth utilities. Particularly, Choice-based Conjoint Analysis (CBC) is very popular because of its realism in mimicking individual choice processes (Haaijer and Wedel, 2007). Choice-based data are most often used in market simulations to predict market shares and develop pricing strategies. In 2005, 77 % of market research institutions have used CBC (Sawtooth Software, 2005a).

The Analytic Hierarchy Process (AHP) (Saaty, 1980) has been of substantial impact in business research and particularly in managerial decision making for a long time (Forman and Gass, 2001; Vargas, 1990). In recent years, comparative studies have shown that AHP-based preference measurement shows high predictive accuracy (Mulye, 1998; Helm, Scholl, Manthey, and Steiner, 2004a; Helm, Steiner, Scholl, and Manthey, 2004b). Mulye (1998) states that a special strength of AHP lies in preference measurement for complex products, i.e., in cases where several attributes have to be included in the product evaluation problem. But in spite of the promising results, AHP did not attract much interest in the marketing research community until now. Several reasons might be responsible for this: First, AHP has only been run against variants of traditional full-profile CA (TCA). So, since even marketing practitioners are used to applying more sophisticated CA methods, such as the commercially successful ACA and CBC, the practical impact of these results might have been rated as limited. Second, computer-aided interviewing has become a standard in consumer preference measurement. The existing empirical studies, however, were conducted using paper-and-pencil technique, which diminishes the comparability with the standard online approaches. Last but

not least, the above-mentioned comparative studies are based on student samples. Therefore, AHP's potential for market share predictions in consumer research settings is still an open research issue. The empirical study being presented in this paper considers CBC and AHP in a real online consumer research setting in order to investigate the predictive and convergent validity of the two approaches.

The remainder of the paper is structured as follows: We first outline the methodological basis of the two approaches in Section 2. In case of AHP we delineate some theoretical and practical concerns necessary to adequately apply AHP to consumer preference measurement. In Section 3, a short overview of prior empirical comparisons of CA-based and AHP-based preference measurement is given. Section 4 depicts the design of our empirical study. Selected results are discussed in Section 5 before Section 6 concludes with some final remarks.

## 2. Methodological background

In 2000, CBC became the most widely used conjoint-related method in marketing research (Sawtooth Software, 2008). A major advantage of CBC, compared to ranking and rating based CA approaches, is that it imitates real shopping behavior (Chrzan and Orme, 2000). This means that in CBC respondents simply choose one among several product alternatives instead of rating or ranking them in terms of their preference. A typical choice task is depicted in Figure 1.

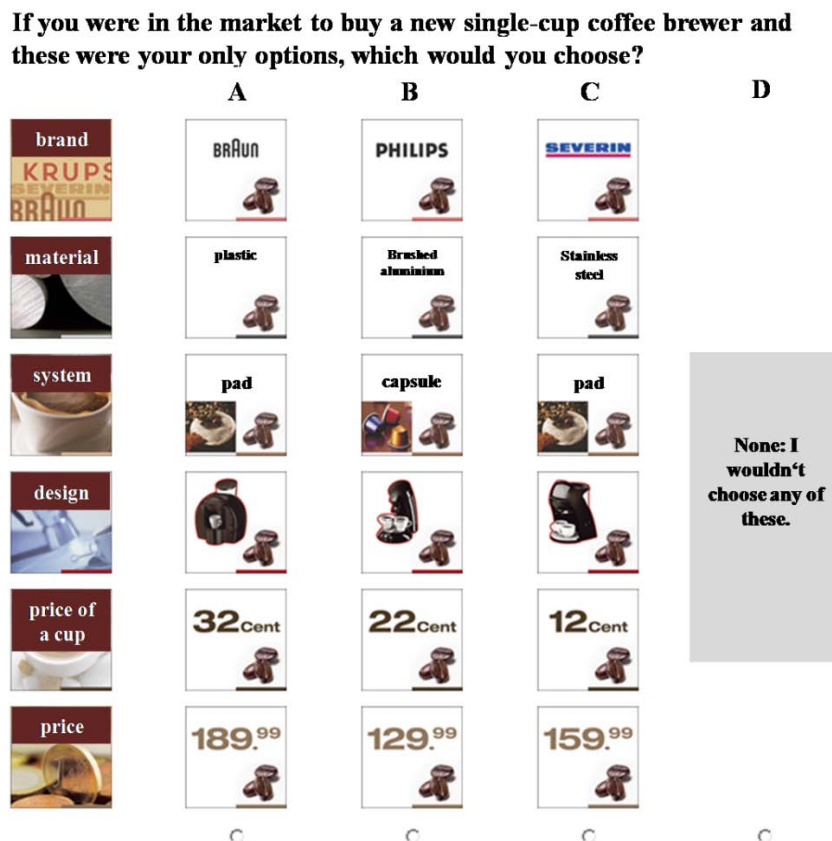


Figure 1: Design of a choice set in CBC

There are different approaches for designing and generating CBC choice sets. The most popular ones among them are random-choice set designs (Chrzan and Orme, 2000). Typically, CBC models are fitted at the aggregate level resulting in a single set of coefficients (or part-worth utilities) being valid for the whole sample of respondents. In order to estimate part-worth utilities at the individual level with relatively few questions, hierarchical Bayesian (HB) methods are employed (Rao, 2007). Individual parameters are estimated using sophisticated Monte Carlo simulation techniques such as Gibbs sampling and Metropolis-Hastings algorithms. For further discussions of these standard techniques we recommend Sawtooth Software (2008) and Sawtooth Software (2005b), as well as

Andrews, Absari, and Currim (2002) who particularly focus on the robustness of the HB estimation technique.

Conjoint-analytic approaches are based on the principle of decomposition, which means that overall evaluations of alternatives are used to derive part-worth utilities of attribute levels. In contrast, AHP, as a compositional approach, follows the principle of “divide and conquer”. Accordingly, complex decision problems are decomposed into smaller, manageable parts. The individual parts are then aggregated in order to derive an overall value for each alternative (Morera and Budescu, 2001).<sup>1</sup> Compositional preference measurement by means of self-explicated approaches has been proven to be quite “robust” (Srinivasan and Park, 1997) and therefore still frequently used as an alternative to CA approaches. There has been an ongoing controversy about the ability of compositional and decompositional approaches to elicit meaningful importance weights from the respondents (Von Nitzsch and Weber, 1993).

In the following, we outline three main adaptations of the original AHP (Saaty, 1980) to better meet the requirements of preference measurement. These changes concern the hierarchy, the scale and the reduction of pairwise comparisons by using incomplete matrices:

**Hierarchy:** AHP aims at structuring complex decision problems. The complexity of the decision problem is reduced by decomposing the decision problem into its smaller constituent sub-problems. When applying AHP to preference measurement, some authors, e.g., Helm et al. (2004a), suggest the use of simple three-layer-hierarchies. The top level of the hierarchy represents the individually best product, which is decomposed into a predefined number of characteristics (called attributes in the preference measurement context) on the second level and corresponding attribute levels on the third level.

An important question is whether to use complete or incomplete hierarchies. In the former case several holistic alternatives are presented at the bottom level of the hierarchy, whereas in the latter attribute levels are shown. Since individual preferences for product features are often heterogeneous, it is difficult to determine a-priori which alternatives should be compared at the bottom level. Particularly in new product development, probably the most important field of preference measurement in marketing, the number of possible product alternatives can be tremendous. In these cases incomplete hierarchies seem to be more feasible. Moreover, incomplete hierarchies also have advantages in terms of interview length and fatigue (Mulye, 1998). From a practical point of view we recommend a bottom-up evaluation of the hierarchy, because then the respondents first become aware of the attribute levels and ranges (difference between the most and the least preferred level of an attribute) before evaluating the importance of the attributes. This means that respondents judge all pairs of attribute levels on the bottom level and then proceed with pairwise comparisons of attributes.

**Scale:** In Saaty’s AHP, two scales are used to measure the ratio of preferences. First, the respondent has to indicate which of two elements he or she prefers. Then, a 9-point rating scale is used to measure the strength of this preference by means of verbal judgments. The numerous applications of AHP to expert decision making (Vargas, 1990) show that this scale leads to satisfying results if the respondents are familiar with the decision problem at hand and experienced in using the AHP scale. However, there is an ongoing debate on using alternative scales for different purposes (Ji and Jiang, 2003; Ishizaka, Balkenborg, and Kaplan, 2005). Gass and Standard (2002) argue that in some cases the compromise positions of the original AHP scale, which are not described by verbal statements, are used rather infrequently. When applying Saaty’s scale to preference measurement, we observed that respondents often did not understand how to use the scale. We therefore developed a simple bipolar scale to measure both the direction and strength of preferences simultaneously. Moreover, we limit the number of response categories to a 9-point-scale format which is common in consumer research.

---

<sup>1</sup> Please note that some authors, e.g., Morera and Budescu (2001) denote AHP as a decompositional technique. In preference measurement, however, compositional preference measurement means that the overall utility of an alternative is composed of the part-worth utilities of attribute levels. Therefore, from the perspective of preference measurement AHP is a compositional approach.

A second modification to the original AHP scale concerns the numerical values assigned to the verbal judgments of the respondents. In consumer research, verbal scales are typically interpreted as providing geometric increases in the preferences of neighboring scale levels. In contrast to that, when using the original AHP scale, the ratios of neighboring scale levels are unevenly dispersed (Salo and Hämmäläinen, 1997). We therefore transformed the verbal statements into the numeric values depicted in Table 1.

Table 1: New geometric AHP scale for use in preference measurement

Scale level	1	2	3	4	5	6	7	8	9
Numeric value	9.00	5.20	3.00	1.73	1.00	1/1.73	1/3.00	1/5.20	1/9.00

Figure 2 additionally illustrates the question and verbal judgments which are used to collect the priority ratios  $a_{ij}$  ( $i, j = 1, \dots, n$ ). In a given sub-problem, the respondent has to answer  $n \cdot (n - 1) / 2$  pairwise comparisons to come up with a complete pairwise comparison matrix  $\mathbf{A}$ . We apply the well known eigenvalue approach to calculate the priority weights of each sub-problem (see, e.g., Saaty 2003). An appealing strength of AHP in preference measurement is that the degree of consistency of a respondent's answers can already be quantified if more than the minimum of  $n - 1$  pairwise comparisons are evaluated. Particularly in online market research, serious concerns about the consistency of responses arise (Schwartz and Oren, 1988). In this case the consistency ratio can be used to track response quality.

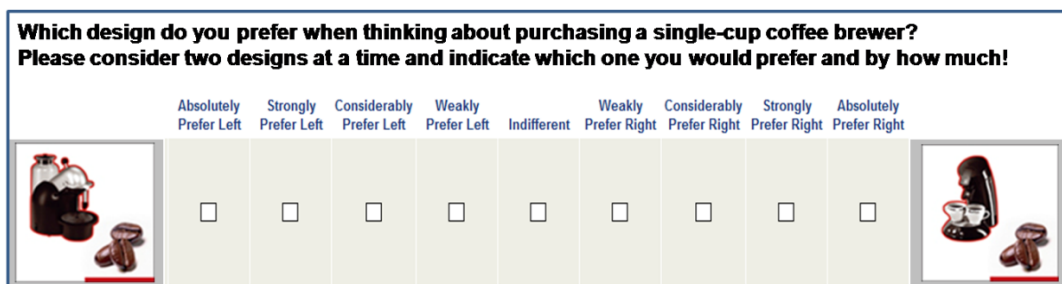


Figure 2: Pairwise comparison of attribute levels for the attribute design

**Reduction of pairwise comparisons:** As Mulye (1998) outlines, the “redundant” comparisons pose practical difficulties in large sub-problems where the number of pairwise comparisons increases rapidly with increasing numbers of elements. In fact, there is a need to reduce the number of pairwise comparisons if the respondents can be assumed not to have the time to complete all  $n \cdot (n - 1) / 2$  comparisons, to be unwilling to make a direct comparison between two elements or to be unsure about some of the comparisons (Harker, 1987). In marketing applications, particularly, interview time and mental fatigue are major concerns. Therefore, reducing the number of pairwise comparisons is desirable to increase the practicability of the approach.

Harker’s (1987) technique for incomplete pairwise comparison matrices can be applied to reduce the number of pairwise comparisons needed to derive the relative preference weights. Instead of approximating the missing entries  $c_{ij}$  in the initial, i.e., incomplete pairwise comparison matrix  $\mathbf{C}$ , the

respective priority ratio is set to be equal to  $\frac{w_i}{w_j}$ . If, for example, three elements have to be compared

in a sub-problem but only two comparisons  $c_{12} = 3$  and  $c_{13} = 9$  are carried out by the respondent, then the pairwise comparison matrix may read:

$$\mathbf{C} = \begin{bmatrix} 1 & 3 & 9 \\ 1/3 & 1 & w_2/w_3 \\ 1/9 & w_3/w_2 & 1 \end{bmatrix} \quad (1)$$

The following matrix  $\mathbf{B}$  can then be derived from the incomplete matrix  $\mathbf{C}$  as follows:

$$\begin{aligned} b_{ij} &= c_{ij} \quad \text{if } c_{ij} \text{ is a real number } > 0 \\ &= 0 \quad \text{otherwise} \\ b_{ii} &= m_i, \text{ the number of unanswered questions in row } i = 1, 2, \dots, n \end{aligned}$$

In our simple example matrix  $\mathbf{B}$  is given by:

$$\mathbf{B} = \begin{bmatrix} 1 & 3 & 9 \\ 1/3 & 2 & 0 \\ 1/9 & 0 & 2 \end{bmatrix} \quad (2)$$

Applying the eigenvalue approach to matrix  $\mathbf{A} = \mathbf{B} + \mathbf{I}$  (with  $\mathbf{I}$  representing the identity matrix) provides the respective preference weights. This approach can be applied as long as a pair of elements is directly or indirectly compared (which means that matrix  $\mathbf{A}$  is irreducible) (Harker, 1987). However, it should be noted that the random consistency indices have to be adjusted when calculating consistency ratios. Forman (1990) provides the respective indices for incomplete pairwise comparison matrices.

To choose an initial set of pairwise comparisons, the two-cyclic design developed by Miyake, Satomi, Keikichi, and Shinohara (2003) can be used. Then only  $2 \cdot n$  pairwise comparisons have to be collected. In the following study such a reduction of the number of pairwise comparisons has been applied if more than five elements had to be considered in a sub-problem.

All in all, AHP has two major advantages to conjoint-analytic approaches in large product evaluation problems: First, the sequential analysis of attributes and levels reduces the risk of cognitively overburdening the respondents. Second, with the help of two cyclic designs and Harker's technique for incomplete matrices, the number of pairwise comparisons can be kept at a managerial level.

### 3. Literature on comparisons of CA-based and AHP-based preference measurement

Helm et al. (2004b) give an overview of existing empirical comparisons between AHP and variants of TCA. The authors conclude that both the complexity of the evaluation task as well as the knowledge of the respondent/decision maker regarding preference measurement influences the quality of the results. According to other papers in this field we categorize a product evaluation problem as being complex if more than six attributes are included in a study (Rao, 2007; Haaijer and Wedel, 2007). If so, more sophisticated CA approaches, e.g., Adaptive Conjoint Analysis (ACA), are used in marketing practice. Table 2 provides an update of the overview given by Helm et al. (2004b) and shows that AHP obviously produces better results than CA if the complexity of the evaluation task is high. The study by Meißner, Scholz, and Decker (2008) in particular indicates that the AHP approach can also be used in online research settings with real consumers who are not familiar with the respective preference measurement approach.

Table 2. Empirical studies on the comparison of CA-based and AHP-based preference measurement

		Complexity of the evaluation task	
		Yes	No
Knowledge in preference measurement	Yes	<u>Mulye (1998) second study:</u> AHP better than TCA	<u>Mulye (1998) first study:</u> Similar results of AHP and TCA  <u>Helm (2004a):</u> AHP considerably better than TCA  <u>Helm (2004b):</u> TCA slightly better than AHP
	No	<u>Meißner et al. (2008)</u> AHP remarkably better than ACA	<u>Tscheulin (1991):</u> TCA better than AHP  <u>Helm (2004b):</u> TCA remarkably better than AHP  <b>This study:</b> <b>Comparison of AHP and CBC</b>

In the present study the description of the product is restricted to six attributes only. Thus the complexity of the evaluation problem can be considered as low. Both CBC and AHP are used in a real consumer research setting, where the respondents are not familiar with the respective approaches. From the existing empirical comparison of Helm et al. (2004b) one might assume that CBC as a decompositional approach might outperform AHP with respect to predictive validity. However, the above-mentioned modifications of AHP might prove beneficial regarding the performance of AHP.

#### 4. Design of the empirical study

To meet the requirements of CBC the most important attributes of a single-cup coffee brewer were determined in a pretest. The number of attributes was restricted to six because otherwise information overload could impair the predictive validity of the decompositional approach (Green and Srinivasan, 1990). The dual questioning technique (Myers and Alpert, 1968) was used to determine the most relevant attributes and levels. Figure 3 shows the respective hierarchy used to describe the single-cup coffee brewers. In order to avoid learning effects and keep the survey length manageable, we used a between subject design, i.e., each respondent had to pass either the AHP or CBC questionnaire. The product description design was the same for both surveys.

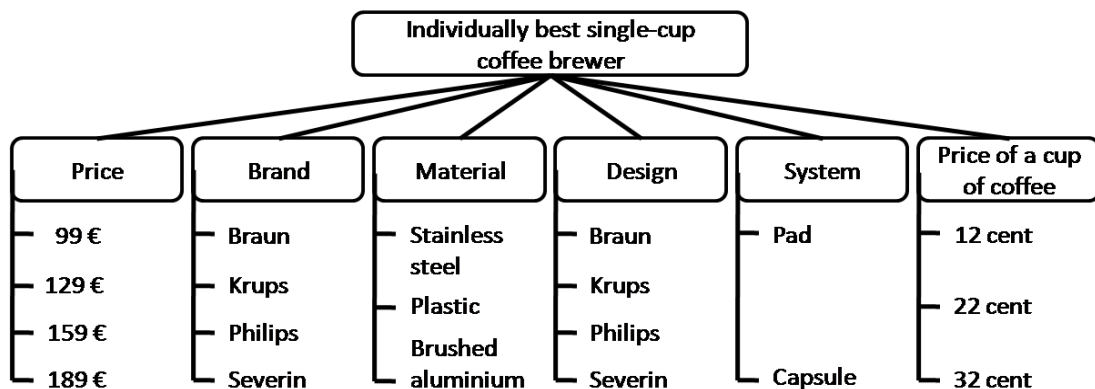


Figure 3. Hierarchy of attributes and attribute levels for single-cup coffee brewers

The online questionnaire form was divided into three parts: In the first part, the respondents were interviewed about their consumption of coffee. Respondents who did not drink coffee within the last 12 months were excluded from the survey because it could be assumed that in this case the purchase intention regarding a single-cup coffee brewer and product experience were quite low. Further

questions concerning the intensity of consumption (cups per week) and prior product experience were added. In the second part of the survey, the respondents were first informed about the attributes and attribute levels describing the different single-cup coffee brewers by means of brief textual and pictorial descriptions. Then, the respondents were randomly assigned to either the AHP or the CBC questionnaire.

The CBC approach was conducted via the popular CBC software of Sawtooth Inc., the leading provider of CA tools. The survey was set up according to the recommendations given in the Sawtooth Software manual. The respondents had to answer 13 choice tasks including three alternatives and a no-choice option. Hierarchical Bayes estimation (a common standard when applying CBC in preference measurement) was used to estimate part-worth utilities on the individual level. The AHP survey was set up as outlined in Section 2. Since the product evaluation problem included six attributes, we only conducted 12 out of the 15 ( $=6 \cdot (6-1)/2$ ) pairwise comparisons and used Harker's (1987) approach to calculate the importance weights. All in all, the AHP questionnaire included 37 ( $=12+6+6+3+6+1+3$ ) pairwise comparisons.

As is common practice in marketing research, two holdout choice tasks were included in the last part of the survey to measure the predictive accuracy of the two approaches. Each choice task consisted of three alternatives which were described on all six attributes included in the study. Standard instructions for selecting the best alternative were used (Sawtooth Software, 2008). The second choice task was presented twice in order to gain a measure of test-retest reliability at the beginning of the preference measurement part of the interview.

Furthermore, three rating scales were added to measure the perceived realism, the difficulty and the enjoyment of the survey. The respective questions read as follows:

- Q1: How realistic is this form of questioning?  
Scale: very unrealistic = 1, ..., 9 = very realistic
- Q2: How difficult was it to respond to the questions asked?  
Scale: very difficult = 1, ..., 9 = very simple
- Q3: How much did you enjoy the survey as a whole?  
Scale: very little = 1, ..., 9 = very much

At the end of the survey some additional questions on socio-demographics (age, gender, profession, size of the household) were asked. The respondents were invited to participate in the survey via a large public email directory. In total 61 and 58 respondents completed the AHP and CBC questionnaires respectively. Chi-square homogeneity tests indicate that the two samples stem from the same population. Therefore, the comparison of AHP and CBC should lead to meaningful results with respect to convergent and predictive validity.

## **5. Results**

### **I. Face Validity**

Product attributes can be nominal, ordinal or even quantitative (Orme, 2002). In the first case, it is unknown a-priori whether a respondent prefers one level over another. In the other two cases, most respondents would usually prefer the given levels in a certain order. For example, it can be assumed that a low price will be preferred to a higher one, but it can hardly be anticipated which of several designs will be favored. In this study the attributes "price", "material" and "price of a cup of coffee" can be assumed to be ordinal. The respective part-worth utilities should therefore follow the a-priori ranking of the attribute levels. The frequency of fulfilling this assumption is used as a measure of face validity.

Table 3 shows that for the majority of respondents the a-priori assumption is correct in both studies. However, the face validity is significantly higher for AHP compared to CBC in case of the attribute

“price”. A possible explanation for this difference is that the holistic product presentation in CBC significantly affects price expectations. In some cases respondents might prefer a higher price to a lower price, if they inferred the quality of the product from the price.

Table 3. Frequency of fulfilling the a-priori assumptions (in %)

Attribute	AHP	CBC
Price	77.05*	48.28
Material	72.13	72.41
Price of a cup of coffee	81.97	82.76

\* p<.05

## II. Convergent Validity

High convergent validity can be presumed if two different measurement methods come to similar results (Helm et al., 2004a). In the following, Spearman’s coefficient is used to quantify the extent of rank correlation between part-worth utilities on the aggregate level. The part-worth utilities of AHP are calculated by multiplying the local priority weights of the attribute levels with the corresponding importance measures. In order to make the AHP part-worth utilities comparable to the ones of CBC, they were transformed by normalizing the local priority weights within each sub-problem such that they sum up to zero (Mulye, 1998).

Figure 4 depicts the part-worth utilities as well as the attribute importances (in %). As can be seen, the part-worth utilities of both approaches show high structural similarity. Both approaches identify “price”, “price of a cup of coffee” and “material” as the most important attributes on the aggregate level. The rank correlation between the part-worth utilities of both approaches equals .93 and the corresponding correlation of the attribute importances equals .86. The ranking of the part-worth utilities is the same for four of the six attributes. Differences occur for the attribute “brand” and “design”.

The results are quite similar on the aggregate level. However, with respect to the differences regarding attributes brand and design, it cannot be said which method provides the better results. We therefore investigate the predictive validity of the two approaches on the individual level.

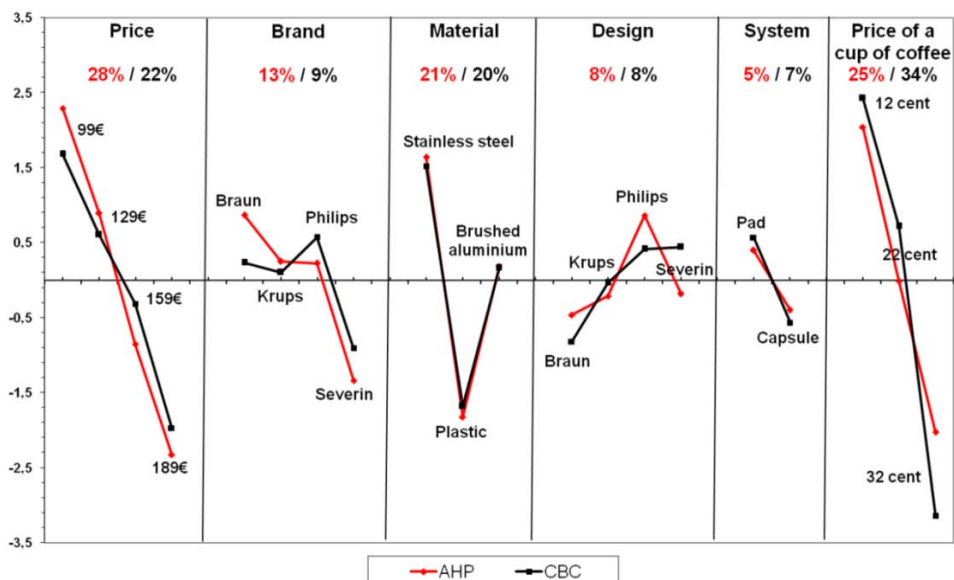


Figure 4. Aggregate part-worth utilities and attribute importances for single-cup coffee brewers

## III. Predictive Validity



In empirical studies the predictive validity is investigated by means of holdout choice tasks for two main reasons (Johnson, 1997): First, they provide an indication of validity, measured by the part-worth utilities' ability to predict choice decisions. Second, they permit the identification and removal of inconsistent responses. In preference measurement it is advantageous to repeat at least one of the holdout concepts to assess the test-retest reliability. In a between-subject design it is critical to adjust hit rates by the test-retest reliability for each group in order to determine whether one method performs better than the other (Orme, Alpert, and Christensen, 1997). The repeated holdout task can be used to calculate a theoretical upper limit for holdout predictability. In our study one of the holdout tasks was shown twice. The corresponding test-retest reliabilities are  $p(\text{AHP}) = 77.04\%$  and  $p(\text{CBC})=77.59\%$  respectively. According to Wittink and Johnson (1992) the maximum expected hit rate can be calculated as follows:

$$\pi = \frac{1 + \sqrt{(2 \cdot p - 1)}}{2} \quad (3)$$

where  $p$  is the agreement between independent replications of the criterion measure. Given the above test-retest reliabilities, the maximum possible hit rates for predicting the holdout choices are  $\pi(\text{AHP}) = 86.77\%$  and  $\pi(\text{CBC}) = 87.14\%$  in the present case.

Two holdout choice tasks with three alternatives each were used to measure the predictive validity. We compared the overall utilities of the alternatives in the holdout tasks with the actual choices of the respondents in the questionnaire. The first choice hit rate equals the frequency with which a method correctly predicts the single-cup coffee brewer chosen by the respondent. On the aggregate level we estimated the choice shares in the holdout tasks. The mean difference between the predicted choice shares of an alternative and the actual choice shares is captured by the average mean absolute error. As in many other comparative studies we used the first choice rule for market share predictions (Green and Srinivasan, 1990).

Table 4. Individual holdout validations

First choice hit rate	AHP	CBC
Choice task 1	68.85	70.69
Choice task 2	67.21	68.97
Average	68.03	69.83

Table 4 shows the first choice hit rates for the two holdout tasks and both preference measurement approaches. Compared to random prediction, which would lead to a first choice hit rate of 33% both methods perform reasonably well. CBC achieves slightly higher hit rates than AHP, but the difference is not significant at the .05 level. Bearing in mind that the reliability of the given answers was not that high in both samples, the predictive accuracy of both approaches seems to be quite good:  $68.03\%/86.77\% = 78.40\%$  (AHP) and  $69.83\%/87.14\% = 80.14\%$  (CBC) of the reliable responses have been correctly predicted.

Table 5. Aggregate choice share validations

Mean absolute error	AHP	CBC
Choice task 1	.0219	.0919
Choice task 2	.0109	.0345
Average	.0164*	.0632

\*  $p < .05$

Table 5 shows that AHP significantly (.05 level) outperforms CBC with respect to the mean absolute error of choice shares in the two holdout tasks. On average the choice shares predicted by AHP deviate from the real choice shares by 1.64%. This error is about more than three times higher for the CBC predictions.

To sum up, AHP seems to be at least on par with CBC with respect to predictive accuracy. This is a very surprising result if one considers that the majority of today's market research institutions use CBC to perform market share predictions.

#### **IV. Practical applicability**

To compare the practical applicability of AHP and CBC, we used the respondents' subjective evaluations (Helm et al., 2004b; Meißner et al., 2008) as described in Section 4. In marketing practice, respondents' ability and willingness to participate in a study is a major concern as the costs of a survey rise with interview length and undesired cancellations of the survey. Moreover, the reliability of the results might decrease, if respondents become fatigue during the interview.

In the CBC sample the respondents had to choose between three holistic alternatives. Thus the cognitive effort should be higher for CBC than for AHP. In the latter only two attributes or attribute levels have to be evaluated at a time. Therefore questions might be much easier to answer and it can be assumed that respondents evaluate CBC to be more difficult. Due to the holistic presentation of product profiles in CBC, this approach should be evaluated as being more realistic compared to AHP.

According to Table 6 all three measures considered are significantly higher/better for AHP. Surprisingly, CBC is not rated as being more realistic. One explanation for this might be that the choice tasks are too complex which results in some kind of information overload. This, in turn, might also impair the enjoyment of the survey. Interview length is another indicator for the practical applicability of an approach. In the present case 7.3 minutes were needed on average to complete the CBC survey and 8.0 minutes for the AHP questionnaire.

Table 6. Average values of measures for practical applicability

<b>Criterion</b>	<b>AHP</b>	<b>CBC</b>
Q1: Realism	5.53*	4.79
Q2: Difficulty	6.74*	6.14
Q3: Enjoyment	5.74*	4.83

\* p<.05

In sum, the results indicate that respondents answering the AHP questionnaire would be more motivated than those answering the CBC questionnaire, albeit the CBC interviews were on average shorter than the AHP interviews. This difference was not significant at the .05 level.

#### **6. Conclusion**

Today, conjoint analytic approaches are marketer's favorite instrument in research and practice. Several empirical studies have shown that AHP at least can keep up with traditional CA. To increase the practical impact of such empirical comparisons, we run AHP (with slight modifications with regard to the hierarchy, the scale and the number of pairwise comparisons conducted) against the popular Choice-Based Conjoint Analysis in an online survey. The empirical study focuses on the comparison in a real online consumer research setting with non-academic respondents to investigate the applicability as well as predictive and convergent validity of the two approaches.

The measures used in this study show that AHP is a good alternative for estimating market shares. AHP outperformed CBC in market share prediction and was almost as accurate as CBC with respect to first choice hit rates. This is an important finding which suggests AHP as a promising tool for preference measurement. Moreover, the results show that AHP is slightly better with respect to face validity as well as the subjective evaluation of the approaches by the respondents. The AHP survey was rated as being more realistic, less difficult and more enjoyable. No significant differences were found concerning the average interview length.

As outlined in the theoretical part of the paper, Harker's (1987) technique for incomplete pairwise comparison matrices can be used to reduce respondents' burden and with that the total interview length. In this context future research should investigate how many pairwise comparisons are needed to ensure a high predictive validity of the results.

Hierarchical structuring of decision problems has been found to be one common theme in the way humans deal with complexity (Forman and Gass, 2001). In this study, we used a simple 3-level hierarchy for modeling the product evaluation problem. However, further empirical investigations are needed to determine the influence different hierarchical structures have on preference measurement. By using hierarchies with more levels, the number of pairwise comparisons required could be further reduced. Albeit, in consumer online surveys respondents might have problems with the comparison of product categories consisting of multiple attributes. We leave this aspect for future research.

## REFERENCES

- Andrews, R.L., Ansari, A., & Currim, I.S. (2002). Hierarchical bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction and partworth recovery, *Journal of Marketing Research*, 39, 87-98.
- Chrzan, K., & Orme, B. (2000). An overview and comparison of design strategies for choice-based conjoint analysis, *Sawtooth Sequim*.
- Forman, E.H. (1990). Random indices for incomplete pairwise comparison matrices. *European Journal of Operational Research*, 48, 153-155.
- Forman, E.H., & Gass, S.I. (2001). The analytic hierarchy process: An exposition, *Operations Research*, 49, 4, 469-486.
- Gass, S.I., & Standard, S.M. (2002). Characteristics of positive reciprocal matrices in the analytic hierarchy process. *Journal of the Operational Research Society*, 53, 1385-1389.
- Green, P.E., & Srinivasan, V. (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54, 3, 3-19.
- Green, P.E., Krieger, A.M., & Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31, 3, 56-73.
- Haaijer, R., & Wedel, M. (2007). Conjoint choice experiments: General characteristics and alternative model specifications. In A. Gustafsson et al. (Eds.), *Conjoint Measurement (pp.199-229)*, Berlin: Springer.
- Harker, P.T. (1987). Incomplete pairwise comparisons in the analytic hierarchy process. *Mathematical Modelling*, 9, 11, 837-848.
- Helm, R., Scholl, A., Manthey, L., & Steiner, M. (2004a). Measuring customer preferences in new product development: Comparing compositional and decompositional methods. *International Journal of Product Development*, 1, 1, 12-29.
- Helm, R., Steiner, M., Scholl, A., & Manthey, L. (2004b). A comparative empirical study on common methods for measuring preferences. *Jenaer Schriften zur Wirtschaftswissenschaft*, from <http://www.wiwi.uni-jena.de/Papers/wp-sw2004.pdf>.

- Ishizaka, A., Balkenborg, D., & Kaplan, T. (2005). AHP does not like compromises: The role of measurement scales. *Joint-Workshop on Decision Support Systems, Experimental Economics & e-Participation 2005, Graz*, 46-57.
- Ji, P., & Jiang, R. (2003). Scale transitivity in the AHP. *Journal of the Operational Research Society*, 54, 896-905.
- Johnson, R. (1997). Including holdout choice tasks in conjoint studies. *Sawtooth software research paper series*, Sequim.
- Meißner, M., Scholz, S.W., & Decker, R. (2008). AHP versus ACA – An empirical comparison. In C. Preisach et al. (Eds.), *Data analysis, machine learning, and applications* (pp. 447-454), Berlin: Springer.
- Miyake, C., Satomi, H., Keikichi, O., & Shinohara, M. (2003). 2-cyclic design in AHP. *Journal of the Operations Research Society of Japan*, 46, 4, 429-447.
- Morera, O.F., & Budescu, D.V. (2001). Reduction of random error in analytic hierarchies: A comparison of holistic and decompositional decision strategies. *Journal of Behavioral Decision Making*, 14, 223-242.
- Mulye, R. (1998). An empirical comparison of three variants of the AHP and two variants of Conjoint Analysis. *Journal of Behavioral Decision Making*, 11, 263-280.
- Myers, J.H., & Alpert, M.I. (1968). Determinant buying attitudes: Meaning and measurement. *Journal of Marketing*, 32, 4, 13-20.
- Orme, B.K., Alpert, M.I., & Christensen, E. (1997). Assessing the validity of conjoint analysis – Continued. *Sawtooth Software Research Paper Series*, Sequim.
- Orme, B.K. (2002). Formulating attributes and levels in conjoint analysis, *Sawtooth Software Research Paper Series*, Sequim.
- Pöyhönen, M., Hämmäläinen, R.P., & Salo, A.A. (1997). An experiment on the numerical modelling of verbal ratio statements. *Journal of Multi-Criteria Decision Analysis*, 6, 1-10.
- Rao, V.R. (2007). Developments in conjoint analysis. In B. Wierenga (Ed.), *Handbook of marketing decision models* (pp. 23-53), New York: Springer.
- Saaty, T.L. (1980). *The Analytic Hierarchy Process*, New York: McGraw Hill.
- Saaty, T.L. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary? *European Journal of Operational Research*, 145, 85-91.
- Salo, A.A., & Hämmäläinen, R.P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-criteria Decision Analysis*, 6, 309-319.
- Sawtooth Software (2005a). Report on conjoint analysis usage among saw tooth software customers, from <http://www.sawtoothsoftware.com>.
- Sawtooth Software (2005b). The CBC/HB system for hierarchical bayes estimation. *Technical paper series – Sawtooth Software Inc.*
- Sawtooth Software (2008). The CBC system for choice-based conjoint analysis. *Technical paper series – Sawtooth Software Inc.*

- Schwartz, R.G., & Oren, S.S. (1988). Using analytic hierarchies for consumer research and market modeling. *Mathematical and Computer Modeling*, 11, 266-271.
- Tscheulin, D.K. (1991). Ein empirischer Vergleich der Eignung von Conjoint-Analyse und Analytic Hierarchy Process (AHP) zur Neuproduktplanung. *Zeitschrift für Betriebswirtschaft*, 61, 1267-1279.
- Vargas, L.G. (1990). An overview of the analytic hierarchy process and its applications. *European Journal of Operational Research*, 48, 2-8.
- Von Nitzsch, R., & Weber, M. (1993). *The effect of attribute ranges on weights in multiattribute utility measurements. Management Science*, 39, 8, 937-943.
- Wind, Y., & Saaty, T.L. (1980). Marketing applications of the Analytic Hierarchy Process. *Management Science*, 26, 641-658.