

## Research Articles

# A User's Guide to the Galaxy of Conjoint Analysis and Compositional Preference Measurement

By Michael Steiner and Martin Meißner

Academics and practitioners aiming to measure preferences for the first time are faced with a 'galaxy' of conjoint analytic and alternative compositional preference measurement approaches. Potential users are confronted with questions, such as how to determine which approach to use, how to set up a study that generates reliable and valid results and how to interpret these results. The aim of this paper is to provide guidance to these key challenges. Thus, the first objective is to guide potential users in the selection of appropriate preference measurement approaches. Second, we discuss important craft factors, i. e. design elements that substantially impact the validity of preference measurement results, such as the definition of attributes and levels and their introduction to respondents. Third, we provide users with practical guidance how to interpret conjoint results, evaluate the quality of empirical conjoint data, set up market simulations and discuss factors that influence the studies' external validity. Our paper, thus, serves as a user's guide to the 'galaxy' of consumer preference measurement and helps sensitize researchers towards potential pitfalls when conducting their own preference measurement studies.



Michael Steiner is Professor of Marketing at University Witten/Herdecke, Alfred-Herrhausen-Straße 50, 58448 Witten, Germany, Phone: +49/2302 926 587, E-Mail: michael.steiner@uni-wh.de. \* Corresponding author



Martin Meißner is Associate Professor of Marketing and Innovation at the University of Southern Denmark, Niels Bohrs Vej 9, 6700 Esbjerg, Denmark, Phone: +45/6550 8653, E-Mail: meissner@sam.sdu.dk and Adjunct Senior Lecturer at Monash University.

## 1. Introduction

Humans make a multitude of choices every day that range from speedy decisions (e. g., which food to eat, task to do, film to watch or mode of transport to take) to more complex purchase decisions on cars, laptops or insurances. Managers, in particular, are required to make decisions regarding recruitment, machinery purchase and investment options. Across all contexts, decision makers must choose from a variety of alternatives. These choice alternatives are often described using product **attributes** (i. e., characteristics used to compare alternatives such as the colour of a car) and related **attribute levels** (specific product features, e. g., the red colour of a car).

The objective of **conjoint analysis** is to measure how respondents **trade-off** various alternatives and their respective attribute levels; for example, the trade-off a consumer faces is whether to buy a cheaper car with higher fuel consumption or a more expensive one with lower fuel consumption. Conjoint analysis is a survey technique that asks respondents to **evaluate products**. These product evaluations are then decomposed to derive estimates on the utilities of the products constituent attribute levels (the **part-worths**). Conjoint analysis is thus a **decompositional** approach.

Conjoint analysis is a class of methods and comprises a large variety of alternative measurement approaches. **Choice-based conjoint analysis** (CBC; also known as **discrete choice modelling**) is, by far, the most commonly used conjoint approach. A recent survey suggests that more than 80 % of all conjoint analysis studies apply CBC (Sawtooth Software 2016). However, other approaches can be used as well. Preferences for specific attribute levels can be surveyed directly. Such direct methods are commonly named **compositional** approaches since the total utility of a product can then be computed (composed) as the sum of its attribute levels' utilities. Furthermore, **hybrid** approaches aim at combining advantages of compositional and decompositional methods. Hybrid and compositional preference measurement approaches have advantages in certain application contexts (as discussed further below).

Numerous studies suggest that conjoint analysis methods (but also compositional and hybrid approaches) well predict real choices (e. g., Benbenisty 1983; Krishnamurthi 1988; Chapman et al. 2009; Louviere and Timmermans

Focus of this paper	Focus of many textbooks and research papers
1) Selection of a Preference Elicitation Technique	(often focus on a rather limited number of approaches and newly developed approaches might be missing)
2) Definition of an Attribute Set	
	3) Definition of an Experimental Design
4) Sample Selection, Explicating the Decision Context and Warm-up Tasks	
	5) Respondents' Evaluations
	6) Part-worth Estimation
7) Validity and Reliability (but also applicability measures and face validity)	(previous textbooks and research papers often focus on internal and predictive validity)
8) Interpretation of Part-worth Utilities	
9) Market Simulations and External Validity	

Tab. 1: Steps within conjoint-analysis studies

1992; Natter and Feurstein 2002; Parker and Srinivasan 1976; Robinson 1980; Srinivasan and Park 1997; Tscheulin 1991). Conjoint analysis is therefore widely accepted in research and practice. Each year, market researchers conduct thousands of conjoint analyses (Orme 2013). Some researchers believe that conjoint analysis is ‘the most significant development in marketing research methodology over the last 40 years’ (Rao 2014a, p. 47; 2014b).

Various textbooks (e. g., Baier and Brusch 2009; Gustafsson et al. 2013; Orme 2013; Rao 2014b) and textbook chapters (e. g., Backhaus et al. 2015a, 2015b; Hair et al. 2010) exist that give an introduction to conjoint analysis, to compositional as well as hybrid preference measurement approaches. These publications mainly focus on how to define an experimental design (the alternatives evaluated within the conjoint task), how to collect preference data, and how to estimate the levels’ utility (the levels’ part-worths). Moreover, commercial software (e. g., see [www.sawtoothsoftware.com](http://www.sawtoothsoftware.com)) enables researchers to easily master these steps without much effort. Today, reasonable conjoint analysis estimates seem to require just a few mouse clicks.

But despite all these publications and easy to use software, many conjoint analysis studies fail (e. g., Brzoska 2003; Drechsler et al. 2013; Fine 2009; Johnson and Amrose 2009; Schlag 2008). There are numerous reasons for such failure. In line with Eggers et al. (2016) or Orme (2013), we argue that ‘craft factors’ (how to design conjoint studies) should be more prominently debated (and empirically tested) than currently done in most textbooks or papers. These factors can strongly influence the validity of preference measurement results.

Our first objective is to help users selecting an appropriate preference measurement approach for the decision context they would like to investigate. Previous comparative studies (e. g., Louviere and Woodworth 1983; Natter and Feurstein 2002; Kamakura and Ozer 2000; Vriens et al. 1996) suggest that no dominant preference measurement approach exists. By discussing advantages and disadvantages of established techniques and also referring to more recently developed approaches, our user’s guide helps in this selection process in a specific research con-

text. Our second objective is to bring three craft factors into focus: We deliberate on the selection of attributes and levels which is the essential first step when setting up a preference measurement study. We then discuss how the decision context as well as the attribute and levels can be adequately introduced to respondents. Our third objective is to provide users with more practical guidance how to interpret conjoint results, evaluate the quality of empirical conjoint data (the reliability, validity and applicability), set up market simulations and measure the external validity. Our paper targets researchers and practitioners that have some fundamental understanding about conjoint analysis (e. g., they know common textbooks on conjoint analysis) but still lack sufficient experience when deciding on a specific preference elicitation technique, when setting up a study, or when interpreting conjoint results and predicting future market share.

In summary, our paper aims to sensitize researchers towards potential pitfalls that could negatively affect or even jeopardize the validity of their preference measurement results. Tab. 1 provides an overview of all steps necessary when conducting a conjoint analysis. It also provides information on this paper’s focus (left column, steps 1, 2, 4, 7, 8, and 9) and the focus of many textbooks (right column, steps 1, 3, 5, and 6). This does not mean that previous publications ignored some steps (nor do we), however, our goal is to discuss those steps in more detail that, from our perspective, have been underweighted previously.

The remainder of this paper is structured as follows. The next two sections introduce conjoint analysis, by first explaining the value of conjoint analysis for research and practice (Section 2) and then briefly introducing the basic measurement idea (Section 3). In line with the first objective, Section 4 presents commonly used decompositional, hybrid, and compositional preference measurement approaches. In line with the second objective, three important craft factors are subsequently discussed in Sections 5 to 7. Section 5 considers the determination of attributes and levels. Section 6 explores the explication of attributes and levels and the use of warm-up tasks. Section 7 includes a description of reliability and validity issues. In line with the third objective, Section 8 illustra-

tes the interpretation of part-worth utilities and Section 9 demonstrates how to conduct market simulations and describes factors that influence the external validity. Section 10 offers concluding remarks.

## 2. Understanding the Value of Conjoint Analysis

Researchers and practitioners using preference measurement methods aim at exploring the current decision making of consumers and predicting future decision behaviour. Understanding consumers' decision making is essential for manufacturers, policymakers, and researchers.

Conjoint analytic approaches are largely used in the early stages of the product development process, even before firms develop prototypes of their products (Green et al. 1997). They provide managers with important information on (i) the value consumers assign to certain levels (levels' part-worth utilities) and (ii) the relevance of certain attributes when making the decision (attribute importance weights). Further, this information is frequently used to (iii) identify consumer segments and (iv) predict products' market share (what-if scenarios). These conjoint analysis results are easy to communicate to management and engineers and provide actionable insight (Chapman et al. 2008).

Knowledge about customer needs help manufacturers design better products. Assessing the perceived value of product characteristics from a consumer perspective increases the possibility of consumers liking and buying the product, rendering market introduction successful and, thus, reducing the risk of market failure (Markham and Lee 2013; Schneider and Hall 2011). Firms can also avoid failures if they have better information on target groups' size, consumers' willingness to pay and their potential reactions to changes in the competitive environment (e. g., price changes or introduction of new products by competitors). These estimates also play a crucial role in forecasting the impact of potential cannibalization of line extensions, brand value and the impact of other marketing mix changes (e. g., changes in communication or distribution strategies; Cattin and Wittink 1982; Hartmann and Sattler 2002; Wittink and Cattin 1989). Conjoint analysis approaches have also been adopted to aid consumers' decision making since preference data can be applied in recommender systems (Backhaus et al. 2010; De Bruyn et al. 2008) and to assess preferences on non-linear pricing models such as the two-part tariff structures of the 'Bahn Card' (Lilien et al. 2013). Conjoint analytic approaches are also commonly applied to analyse individual as well as group decision making (Krishnamurthi 1988).

Furthermore, conjoint analysis is not only relevant in the marketing field but also commonly used in areas such as healthcare (e. g., to facilitate doctor-patient communication), economics (e. g., to assess public transport prefer-

ences), human resources (e. g., to develop incentive schemes, define desired characteristics for new employees and understand job choice; Huber et al. 1971; Netzer et al. 2008) and computer science and machine learning (Netzer et al. 2008; Orme 2013). Lawyers apply conjoint analysis to assess damages from patent infringement by valuing patents and copyrights (Cameron et al. 2013). For example, conjoint analysis was central in Apple's 2.5 billion USD suit against Samsung. Apple conducted two studies (one on smartphones and the other on tablets) to quantify the value of features that Samsung incorporated into their products and protected by Apple patents. Predictions of customers' willingness to pay for these features were then presented in court (Eggers et al. 2016). The royalty rates to stream music were determined in a manner similar to that utilized by the US Copyright Royalty Board (McFadden 2014). Due to its wide range of applications, conjoint analysis is one of the most often used techniques to assess consumers' needs (Lilien et al. 2013; Netzer et al. 2008).

## 3. Basic Idea of Conjoint Analysis

This section provides an overview of conjoint analysis and presents assumptions and recent research on how respondents answer conjoint questionnaires. Conjoint analysis is **decompositional** since (as described earlier) it asks respondents to evaluate products described by all respective attribute levels (the so-called 'full-profile presentation') and then decomposes these evaluations to obtain information on the value consumers derive from specific features.

Products assessed within conjoint analysis can be defined by simply identifying all potential level combinations. However, the number of product evaluations will increase exponentially with the number of attributes and levels (e. g., each respondent would be required to evaluate 729 alternatives when considering six attributes with each three levels). Thus, experimental designs aim at reducing the number of alternatives respondents must evaluate to reduce their burden. In a traditional conjoint analysis, such experimental designs can be generated on the basis of the Latin square or Addelman plans (Addelman 1962).[1] In all experimental designs, each level of an attribute is presented equally often together with every other level of the other attributes in the study. Moreover, when applying traditional conjoint analysis approaches, all respondents evaluate the same experimental design. To limit the number of products that respondents must evaluate, these designs often focus on the main effects by assuming that the levels of different attributes are uncorrelated. Respondents are then asked to evaluate the products e. g. based on a ranking, ratings, paired-comparisons etc. (traditional conjoint analysis) or choices (choice-based conjoint analysis).

Before estimating part-worth utilities, market researchers must define a utility model for each attribute. The utility



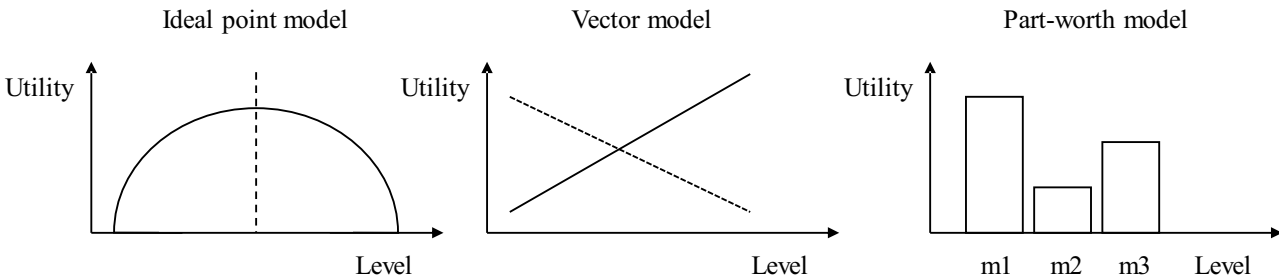


Fig. 1: Utility models

model determines the relationship between attribute levels and consumer utility. Figure 1 presents three common utility models. An **ideal point model** assumes that an attribute has an ‘optimal’ level that maximizes consumer utility; for example, a chocolate bar should not be too sweet or not sweet enough. A **vector model**, on the other hand, assumes that higher (or lower) levels of an attribute increase consumer utility; for example, a faster computer processor should be perceived as beneficial (continuous line of the vector model in Fig. 1). Similarly, a lower-priced computer (dotted line in Fig. 1) should increase consumer utility. Finally, the **part-worth model** does not make any a priori assumptions regarding the utility caused by specific levels. Thus, the part-worth model is the most flexible approach. It is most often applied in practice. However, when applying the part-worth model, a higher number of parameters must be estimated (Green and Srinivasan 1978; 1990).

Researchers can estimate part-worths based on estimation techniques such as regression analysis (ratings-based data) or multinomial logistic regression (choice data). The total value of an alternative can then be computed as the sum of its respective part-worth utilities (see Equation 1). Almost all published research on conjoint analysis and preference measurement is based on the assumption of a linear additive utility model (Cui and Curry 2005).

$$U_{im} = \sum_{j=1}^J \sum_{k=1}^K \beta_{ijk} * y_{ijkm} \quad (1)$$

with:

$U_{im}$ : total utility respondent  $i$  derives from alternative  $m$

$\beta_{ijk}$ : part-worth utility for level  $j$  of attribute  $k$  for respondent  $i$

$y_{ijkm}$ : binary coding for level  $j$  of attribute  $k$  for alternative  $m$  and respondent  $i$ , which takes the value of 1 if level  $j$  is present and 0 otherwise

$J$ : number of attribute levels

$K$ : number of attributes

The assumption of a linear additive utility model has an important consequence: additive utility models are compensatory, that is, they assume that a favourable level of one attribute can compensate for the less favourable level of another attribute (and vice versa). This assumption is also grounded in the idea that humans make computer-like calculations, which may be questionable at least in purchase situations wherein the respondent is unwilling or does not have the time to consider all attribute levels.

Nonetheless, research has shown that the assumption of a linear additive model is surprisingly robust to violations of this assumption (for an overview, see Cui and Curry 2005).

In contrast to such unsystematic errors, systematic deviations from the linear additive model bias conjoint estimates. Attribute levels that are completely unacceptable are an example of such systematic biases. To elaborate, respondents may accept, to a certain degree, a higher product price if it can be justified with higher quality. However, consumers might be unwilling to accept a product that is priced higher than a certain price threshold. If so, consumers will not be willing to buy the product, regardless of how favourable other product features are. Such systematic **non-compensatory decision rules** cannot be represented using a linear additive utility model and thus, linear additive utility functions are not suitable (for an overview, see Steiner et al. 2016).

Part-worth utilities provide information on the value of the respective levels and enable researchers to predict a product’s total utility. From a managerial perspective, of high significance are the **importance weights**, that is, the relevance of the attributes for the respondents when making decisions (Green and Krieger 1995). Conjoint analysis does not directly survey the attribute importance weights. Instead, importance weights are derived from part-worth utilities. Attribute importance denotes the degree of change in alternative’s utility value when the attribute improves from the least to the most preferred level (the so-called ‘bandwidth’). Thus, the importance weight for attribute  $j$  ( $w_j$ ) can be computed from the bandwidth of an attribute (i. e., the most minus the least preferred level) divided by the sum of all bandwidths for all attributes (see Equation 2; Cattin and Wittink 1982).

$$w_j = \frac{\max \beta_{ij} - \min \beta_{ij}}{\sum_{j=1}^J (\max \beta_{ij} - \min \beta_{ij})} \quad (2)$$

The resulting importance weights of all attributes  $J$  add up to 100 percent. In general, attribute importance weights are first computed at individual levels and then, aggregated across respondents (Orme 2013).

It is noteworthy that the definitions of the attribute set (attributes and levels surveyed in a conjoint analysis) and range of attribute levels influence attribute importance

weights. A narrow bandwidth of levels (e. g., narrow price range) will result in a low importance of that attribute (e. g., price) while a broader bandwidth will increase it (see also Section 3 and Orme (2013)).

To develop a deeper understanding of how respondents make decisions in conjoint tasks, eye tracking can be used to monitor respondents' decision-making processes (Meißner and Decker 2010). Meißner et al. (2016), for example, analysed respondents' attention towards alternatives and attributes in three CBC eye-tracking studies and found a close relationship between attention and preference. First, **alternative focus** predicts that people pay more attention to alternatives that have high utility values. Second, **attribute focus** assumes that people focus more on important attributes. To assess the potential influence of the decision context, Meißner et al. (2016) further relied on two potential influence factors. First, respondents might pay more attention to information that is centrally presented (centrality bias) and second, decision makers may focus more on attributes and products that they look at first in a task. The authors found that central information and information that is perceived first receives more attention. Importantly, however, this increase in attention did not increase the probability of respondents choosing these alternatives. This finding is important because it suggests that conjoint analysis results are not influenced by incidental perceptions resulting from the conjoint tasks' layout. Moreover, as the study also demonstrated that alternative and attribute focus become stronger in later choice tasks, the conjoint task layout seems to facilitate more effective decision making owing to practice, that is, fewer fixations are necessary before respondents can make their decisions.

A key characteristic of CBC is its lattice structure. More specifically, attributes are arranged in a matrix and their order is identical for all alternatives, which enables respondents to easily compare different options. Further, the number of attributes is commonly limited to avoid overwhelming the respondents and alternatives are often distributed across different sub-tasks. Decision makers are repeatedly asked to evaluate alternatives or choose one of them. These layout features are likely to enable learning with conjoint exercises.

## 4. Methods and Experimental Design

Green and Rao (1971) introduced the concept of conjoint analysis more than 45 years ago. Since then, researchers have developed numerous methods to improve data quality or reduce respondent effort when participating in preference elicitation tasks. This paper aims to provide a brief overview of preference measurement techniques with focus on different types of conjoint analysis. However, such an overview cannot be all-encompassing. Nevertheless, the objective is to provide an overview of the bandwidth of possible approaches that can be applied to gain insight into consumer preferences. *Tab. 2* summarizes the approaches considered in the present discussion.

### 4.1. Compositional Approaches

Compositional approaches ask respondents to directly evaluate attributes and/or levels. **Direct ratings** ask respondents to assess the attributes' importance without considering the respective attribute levels. Such direct importance ratings are often uninformative (Krosnick and Alwin 1988) because respondents tend to rate all attributes as important. Unweighted self-explicated approaches, in contrast, focus on asking respondents to assess the attribute levels' desirability but neglect differences in attribute importance. These approaches implicitly assume that all attributes are equally important. Both approaches do not allow researchers to assess the total utility of an alternative.

**Weighted self-explicated approaches** ask respondents to evaluate both, attributes and levels. Often, respondents are first asked to evaluate the desirability of each attribute level. For example, respondents could be asked to determine the most preferable level of an attribute, which is assigned 10 points, followed by the least preferable level, which is given 0 points. The remaining levels are assigned between 0 and 10 points. On the basis of this first stage, the bandwidth of levels (most and least preferred level) can be determined.

This bandwidth is then presented in the second stage, where respondents are asked to evaluate the attributes' importance. Here as well, respondents define the most

Compositional approaches	Hybrid approaches	Decompositional approaches
<ul style="list-style-type: none"><li>■ Self-explicated approaches:<ul style="list-style-type: none"><li>– Direct rating</li><li>– Unweighted self-explicated approach</li><li>– Weighted self-explicated approach</li><li>– Conjunctive compensatory self- explicated approach</li><li>– Adaptive Self-Explicated approach and its extension, pre-sorted self- explicated approach</li></ul></li><li>■ Analytic Hierarchy Process and its adaptation for preference measurement</li><li>■ Max-Diff Scaling / Best-Worst Scaling</li><li>■ Restricted-Click-Stream analysis</li></ul>	<ul style="list-style-type: none"><li>■ Traditional Hybrid Conjoint Analysis</li><li>■ Sawtooth Software Adaptive Conjoint Analysis (ACA)</li><li>■ Sawtooth Software Adaptive Choice-Based Conjoint Analysis (ACBC)</li></ul>	<ul style="list-style-type: none"><li>■ Traditional Conjoint Analysis (full-profile approaches vs. trade-off matrices; simultaneous presentation of all product concepts vs. paired-comparisons)</li><li>■ Choice-Based Conjoint analysis (CBC)</li></ul>

Tab. 2: Compositional, decompositional and hybrid preference measurement approaches

important attribute (i. e. the attribute for which an improvement from the least to the most preferred level results in the highest utility increase). This attribute is valued at 10 points. The importance of the remaining attributes is then determined relative to the most important one. A product's total utility is determined by multiplying the attribute importance weights with the desirability ratings of the respective levels and then, summing up these values across all attributes. Thus, total utility is composed of its part-worth utilities (Netzer and Srinivasan 2011; Srinivasan 1988). This method is known as **conjunctive compensatory self-explicated approach** (or also Casemap). Variants of it also enable researchers to account for potential unacceptable levels by asking respondents to eliminate these unacceptable levels prior to rating the levels.

A major disadvantage of self-explicated methods is that the use of rating scales does not require respondents to make trade-off decisions between attributes (Green and Srinivasan 1990). Nevertheless, self-explicated approaches have been frequently used and results show a surprising robustness compared to common conjoint analysis approaches (Srinivasan and Park 1997). Moreover, such compositional methods enable researchers to assess many attributes, while common conjoint analysis approaches are limited to six or eight attributes (see below). The ability to consider many attributes is increasingly important for market research practices since products and services are becoming more and more complex and consumers have become more informed than ever (Netzer and Srinivasan 2011).

Drawing on the criticism that self-explicated approaches do not explicitly consider trade-offs between attributes, Netzer and Srinivasan (2011) proposed an **Adaptive Self-Explicated** approach (ASE). In the first step, respondents are asked to evaluate the attribute levels' desirability on an 11-point rating scale (0 = 'not at all desirable' and 10 = 'extremely desirable'). However, the attributes in the second step are not evaluated using rating scales. Instead, Netzer and Srinivasan (2011) proposed the use of a constant-sum scale that forces decision makers to engage in trade-offs. Because evaluating attributes using constant-sum scales is cognitively demanding and might overwhelm respondents, the process of constant-sum evaluations across all attributes is broken down into constant-sum allocations between attribute pairs. As a consequence, applying the approach increases the number of pairwise comparisons for the respondent.

To reduce this number, researchers first ask respondents to rank the attributes in terms of importance. Thus, in the second stage, the number of pairwise comparisons is reduced by asking respondents to rank attributes before conducting the comparisons. In the third stage, respondents then allocate the 100 points using a sliding bar. Thus, constant-sum evaluations are adapted to respondents' rankings. In total, three constant-sum questions are constructed that comprise the most and least impor-

tant attributes and those in the middle. Then, importance weights are computed for these attributes and interpolated to obtain an importance score for the remaining attributes.

Netzer and Srinivasan (2011) compared ASE to a weighted self-explicated approach and CBC using state-of-the-art estimation techniques (e. g. hierarchical Bayes and fast polyhedral conjoint) as well as Adaptive Conjoint Analysis. Based on two empirical studies, they demonstrated that ASE offered reasonable utility estimates and improved the predictive validity (see below).

ASE, however, is based on the assumption that the initial ranking tasks provide reliable information on attribute importances. Errors at this early stage might negatively affect the validity of the preference measurement results. Therefore, Schlereth et al. (2014) proposed a pre-sorted adaptive self-explication approach that includes a rating task preceding the ranking. This initial rating of attribute importance should ease the subsequent ranking. On the basis of the two empirical studies, Schlereth et al. (2014) demonstrated that his extension reduces the cognitive burden of the entire evaluation task and increases the predictive validity of the approach.

The major benefit of all self-explicated techniques is that they enable researchers to consider a higher number of attributes than standard conjoint analytic approaches. *Tab. 2* presents the most common self-explicated approaches. All recent approaches are structurally based on Srinivasan (1988). The reader is referred to Schlereth et al. (2014) for an extensive overview of self-explicated approaches.

Other studies have tested the applicability of the **Analytic Hierarchy Process** (AHP) and shown that it could outperform conjoint analytic approaches even in moderately complex decision contexts. For example, Scholl et al. (2005) demonstrated that AHP outperforms traditional conjoint analysis when assessing an evaluation task described by six attributes. Helm et al. (2008) showed that a traditional conjoint analysis seems more suitable for simple decision contexts (consisting of four attributes), while AHP is preferable for more complex evaluation tasks.

Scholz et al. (2010) proposed the **Pairwise Comparison-based Preference Measurement** (PCPM) approach which is based on AHP. The authors found that PCPM outperformed an Adaptive Conjoint Analysis (ACA; described below) and a weighted self-explicated approach for highly complex products (conducting two studies, they assessed 10 and 15 attributes). The PCPM approach was further extended by Meißner et al. (2011), who suggested an algorithm that adaptively selects pairwise comparisons on the basis of a respondent's previous answers to increase the approach's efficiency.

Finally, **Max-Diff Scaling** is a relatively new approach to assess consumers' preferences as long as researchers are only interested in obtaining attribute importance



weights. Within a max-diff scaling (also known as maximum difference scaling or a version of Best-Worst Scaling), a list of items is defined and then subsets (choice sets) are created. Within the survey, respondents are asked to identify the most (best) and least (worst) important item from each subset that are sequentially presented. This data enables researchers to compute importance weights for each item (attribute). Max-Diff Scaling enables researchers to assess more attributes compared to conjoint-analytic approaches. Moreover, identifying the least and most relevant item from a subset of attributes requires little cognitive effort (Louviere 1991; Sawtooth Software 2013a).

The **restricted-click-stream analysis** is based on the general idea of information display boards (such as mouselab). Within these approaches, products are presented to respondents, however, their respective attribute levels are covered. Respondents can click on the respective fields to reveal a product's feature and then select the preferred alternative. This idea of a mouselab experiment is adapted to gain insight into the attributes' perceived importance by adjusting two key characteristics. First, any feature that was uncovered by a respondent remains visible and second, the number of features that respondents can uncover in each choice set is restricted. Thus respondents need to deliberately decide on the attribute for which they need more information before making a choice. Schlereth and Schulz (2014) compare importance weights derived from conjoint analysis with those from a restricted-click-stream analysis and demonstrate that the attention that respondents pay to a certain attribute is a feasible proxy to assess its importance.

4.2. Decompositional Approaches

A common characteristic of compositional approaches is that they directly survey respondents' evaluations of attributes and their respective levels. In addition, they are less cognitively demanding and enable researchers to assess a higher number of attributes and levels. However, they are used less often than decompositional approaches such as conjoint-analysis, whose tasks allow a more realistic presentation of the stimuli. As noted, conjoint analysis asks respondents to evaluate multi-attribute alternatives and then decomposes these product evaluations to estimate the levels' part-worth utilities.

The two types of conjoint analytic approaches that are frequently used are traditional conjoint analysis and CBC.

4.2.1. Traditional Conjoint Analysis

Conjoint analysis was first introduced by Green and Rao (1971) in the field of marketing and based on an idea published earlier by Luce and Tukey (1964). This first version of conjoint analysis presented respondents with 27 product profiles that were described by all respective levels (**full profile approach**). They were then asked to define four piles of cards (poor, fair, good and excellent).

Price	Top speed (MPH)		
	130	100	70
2,500 USD	1	2	5
4,000 USD	3	4	6
6,000 USD	7	6	9

Tab. 3: Example for trade-off matrix (Johnson 1974)

Within each pile, respondents sorted all products from the worst to best. Using this rank order, Green and Rao (1971) then computed part-worth utilities for all levels.

Three years later, Johnson (1974) published a similar approach drawing on trade-off matrices. In this approach, respondents were not required to evaluate complete products but attribute pairs, and thus, it is known as a **partial-profile approach**. Tab. 3 provides an example for such a trade-off matrix based on a ranking. In this small data example, the respondent prefers a car at the lowest price (2,500 USD) and the highest speed (130 MPH); this alternative ranked first. Similarly, the car that ranked second had the lowest price (2,500 USD) and a medium top speed (100 MPH).

Trade-off matrices are easier to answer than full profile approaches since they limit the decision task to attribute pairs. However, evaluating whole multi-attribute alternatives is more realistic. As a result, trade-off matrices have little relevance in marketing research practice, whereas full-profile approaches have become widespread in research and practice (Green and Srinivasan 1978; Johnson 2001; Oppewal et al. 1994).

The literature contains various traditional conjoint analyses, which differ by the scales used. Traditional conjoint-analysis can be based on a rating, ranking or dollar metrics. Moreover, market researchers may present all stimuli simultaneously, separately or in pairs. Presenting all products simultaneously may overwhelm respondents and evoke the use of simplifying decision rules.

Thus, researchers have argued that the paired comparisons of product profiles ease decision making (Johnson 2001). Two types of paired comparisons have been commonly used. Dichotomous paired-comparisons ask respondents to select the preferred alternative within each choice set, although these approaches survey little information. Therefore, conjoint approaches based on paired comparisons commonly rely on graded paired comparisons, that is, respondents not only state their preferred alternative but also indicate preference strength on a rating scale. While paired comparisons ease decision making, they require numerous more evaluations compared to conjoint analysis approaches that simultaneously present all stimuli (for the development of experimental designs, see Green and Devita 1974; Hausruckinger and Herker 1992)

As noted, the number of alternatives that can be considered in a conjoint analysis is rather limited. Green et al. (1972), therefore, proposed to split all attributes into subgroups (partial profile design). Respondents were told that the attributes excluded from a subset do not differ in

their respective levels and thus, do not influence evaluations. Oppewal et al. (1994) extended this idea of partial profile designs and proposed a **Bridging Conjoint Analysis**. Similar to Green et al. (1972), the total number of attributes is split into different sub-designs; however, one attribute (bridging attribute) is used in all sub-sets. The part-worth utilities of the remaining attributes are then scaled with respect to the bridging attribute to increase the comparability of estimates across attribute sub-sets. Such partial-profile designs seem beneficial since they enable researchers to assess a higher number of attributes without requiring respondents to consider them simultaneously. However, the approach is rarely used when conducting a traditional conjoint analysis because it increases the number of evaluations. Thus, while partial profiles might be less relevant for traditional conjoint approaches, similar ideas are common for CBC and are presented below.

Traditional conjoint analysis does not provide information on whether the conjoint stimuli, that is, the products being evaluated by respondents, are acceptable from the respondents' perspective. Voeth and Hahn (1998), therefore, proposed a **Limit Conjoint Analysis**. In the first step, respondents are asked to rank all stimuli of a traditional conjoint analysis. In the second step, respondents must then add a limit card that denotes the point that separates acceptable and non-acceptable alternatives. This information can be used when conducting market simulations, that is, non-acceptable alternatives have a 0 % choice probability.

The estimation technique used to estimate part-worth utilities depends on the scale used. Data surveyed on the basis of rankings can be assessed using MONANOVA or LINMAP. OLS regression is used for rating data (Green and Srinivasan 1978).

#### 4.2.2. Choice-based Conjoint Analysis

The development of the CBC represents another important milestone in the history of conjoint analysis. The use of discrete choices was proposed by Louviere and Woodworth (1983). Instead of ranking or rating product alternatives, respondents are asked to choose the preferred alternative from sets of alternatives, the so-called 'choice sets'. Choice designs commonly comprise up to 12 choice tasks, each with 2–7 alternatives (Johnson and Orme 2007; Orme 2013). In its core, CBC combines Green and Rao's (1971) idea of using systematic experimental designs to reduce the number of product evaluations and McFadden's multinomial logistic regression that relates utilities to choice probabilities (MNL; McFadden 1976). CBC is the most often applied type of conjoint analysis (Orme 2013).

CBC has several advantages but it also has disadvantages compared to a traditional conjoint analysis. First, asking respondents to choose the preferred alternative is often assumed to better resemble decisions consumers make in the marketplace. Moreover, choosing is often perceived

as less cognitively demanding than asking respondents to rate or rank alternatives. Rating scales are often interpreted differently by various respondents, while choices are unambiguous. Furthermore, within CBC the experimental design is pooled across respondents. This enables researchers to use more flexible experimental designs (e. g., consider interaction effects). Finally, CBC allows researchers to include a 'none' option that enables them to predict product acceptance. Previous research suggests that even simple aggregate CBC models seem to provide reasonable market share predictions (Chapman et al. 2009; Louviere and Woodworth 1983; Natter and Feurstein 2002). Natter and Feurstein (2002), therefore, argue that more advanced estimation techniques might not always be beneficial.

However, numerous other studies have demonstrated that traditional conjoint analysis approaches (Benbenisty 1983; Tschulin 1991) and even approaches based on direct evaluations (Netzer and Srinivasan 2011; Srinivasan and deMaCarty 1999) can predict real consumer behaviour surprisingly well. Different conjoint analysis approaches can yield similar performance and even simple approaches might be reasonable (Kamakura and Ozer 2000). There are even numerous studies that demonstrate CBC's weak ability to predict real choices (e. g., Brzoska 2003; Drechsler et al. 2013; Fine 2009; Johnson and Amrose 2009; Schlag 2008). For example, Schlag (2008) compared the external validity (ability to predict consumers' real willingness to pay for a product) of a traditional conjoint analysis and a CBC using two research objects. He noted that external validity was poor for both conjoint analysis types, although all his studies present high internal validity.

The most obvious disadvantage of CBC is that choices only capture information on the preferred alternative and not evaluations of the remaining alternatives in a choice set; for example, there is no information on the rank order of the remaining alternatives or strength of preference. Thus, CBC surveys less information.

Fast polyhedral adaptive conjoint analysis approaches aim at addressing this problem and improving CBC by dynamically adjusting the choice sets surveyed (Toubia et al. 2003). Alternatives are created in a way to avoid dominated or dominating alternatives by rendering the stimuli as similar as possible in their likelihood of being chosen (Huber and Zwerina 1996; for applications, see also, for example, Johnson et al. 2003; Johnson et al. 2004). However, these approaches assume that people make consistent and compensatory decisions, which is not always true. As a result, these experimental designs may not be more efficient (Eggers and Sattler 2011).

Following data collection, the part-worth utilities can be estimated at an aggregate level using McFadden's multinomial logistic regression, at target group level by applying a latent class analysis or at individual level adopting a hierarchical Bayes approach. These estimation techniques are described in Elshiewy et al. (2017).



Other studies have suggested employing estimation techniques based on machine learning, such as support vector machine algorithms, to better account for non-compensatory decision behaviour (Cui and Curry 2005; Evgeniou et al. 2005). For example, Evgeniou et al. (2005) demonstrated that support vector machine algorithms handle noise better than other benchmark estimation techniques such as logistic regressions, hierarchical Bayes and fast polyhedral methods.

### 4.3. Hybrid Approaches

Researchers commonly struggle with two major challenges. First, unacceptable levels that may make the preference measurement results less reliable and second, researchers may need to consider more attributes and levels than traditional or CBC approaches can handle.

Hybrid approaches aim at handling these problems by surveying consumer preferences directly (and also enables researchers to directly identify unacceptable levels, preferred levels and attribute importance weights) and accordingly, using this data to individualize conjoint analysis. In other words, hybrid approaches commonly adopt a compositional first step and then use this data to, for example, develop an experimental design for conjoint analysis that focuses on the most important attributes (Johnson 2001). The two most commonly used hybrid approaches are the **Adaptive Conjoint Analysis (ACA)** that is based on a traditional conjoint analysis and the **Adaptive Choice Based Conjoint Analysis (ACBC)**.

**ACA** is concerned with identifying the most important attributes from a set of attributes at the individual level. In the first step, respondents are asked to evaluate the levels and attributes on the basis of a weighted self-explicated approach (see above). The attributes which are most important for the respondent are then included in the subsequent conjoint analysis task. Here, respondents evaluate pairs of alternatives using a rating scale. These paired comparisons are adjusted to the results from the first compositional step and the experimental design of the conjoint task is adjusted to previous evaluations in the conjoint task. Similar to the previously mentioned fast polyhedral conjoint designs, ACA seeks a utility balance to survey respondents' preferences more efficiently. When estimating part-worth utilities, ACA combines respondents' answers from both steps. ACA's main benefit is the number of attributes that can be considered: while common ACA studies include 15 attributes, it can handle up to 30 (Orme 2013; Sawtooth Software 2007).

**ACBC** aims at identifying levels that are relevant from a respondent's perspective. These levels are then surveyed in the subsequent CBC task. Thus, ACBC's primary goal is not to handle a larger number of attributes, but to improve the input quality of the answers by integrating several additional tasks aimed at identifying the individual's **evoked set** that comprises the acceptable product levels. In the first step, the respondent can define an 'ideal' product in a configurator-like 'build your own' task.

Then, using the answers, several alternatives are constructed in a second 'screening section'. Here, the respondent evaluates the acceptability of several alternatives. If a consumer always rejects an alternative which includes a specific level (or only accepts products with a specific level), these levels are directly evaluated. The respondent is asked to evaluate whether a specific level is unacceptable (or a must-have level). In the default setting, the respondent can define up to five levels as unacceptable and four as must-haves. Respondent's answers from all steps are used in estimating part-worth utilities. Since ACBC surveys more information at individual level, its estimates are assumed to be more stable than CBC part-worth utilities (Sawtooth Software 2009).

Previous research suggests that respondents perceived ACBC questionnaires to be more engaging and both the questionnaires and products to be more realistic (since ACBC individualizes its levels). Moreover, sample size requirements are lower since ACBC surveys more information. However, part-worth utilities were similar to CBC results (Orme 2013).

## 5. Definition of Attribute Sets for Preference Measurement

An attribute set comprises all attributes and their respective levels to be assessed within preference measurement. Previous research suggests that many conjoint analysis studies include attributes and levels that do not significantly influence consumer preferences (Louviere et al. 2005). Many researchers, therefore, believe that the definition of an appropriate attribute set is one of the most (Weiber and Mühlhaus 2009), if not the most, important step when measuring preferences. However, most scholars pay little attention to this step. Thus, many researchers believe that their colleagues should put more effort into the selection of an appropriate attribute set (Bradlow 2005; Helm et al. 2004; Keeney and Gregory 2005; Lee and Bradlow 2011; Lilien et al. 2013; Louviere et al. 2005; Wittink et al. 1982). According to Orme (2013). Researchers should consider the following requirements when defining an attribute set:

- attribute levels must be *mutually exclusive*
- there should be *no interaction* effects between attribute levels when conducting a traditional conjoint analysis
- attributes and levels must be *understandable*, concrete and unambiguously interpretable
- attributes must be *important* and attribute levels must be *relevant and acceptable*

*Mutually exclusive* indicates that an alternative can only comprise one level of each attribute. For example, imagine the attribute 'add-on features' for laptops has the levels 'no add-ons', 'docking station' and 'digital pen'. These attribute levels are not mutually exclusive because the market also consists of laptops that have a docking

station and digital pen (e. g., the Microsoft Surface laptops). In this case, researchers would be unable to compute the total utility of a laptop that has both features. Thus, it is necessary to include these levels as separate attributes, each described by two levels (yes/no), or extend the number of levels in the above example by adding the level 'docking station and digital pen'. By doing so, the attribute would cover all possible level combinations, as suggested by Orme (2013).

Experimental designs for traditional conjoint analysis approaches often focus on estimating the main effects. Thus, potential **interactions effects** cannot be considered. Interaction effects are relevant if the level of one attribute influences consumer preferences for a level of another attribute. Assume, for example, that a consumer prefers a car that is 'grey' in colour. However, when thinking about a Ferrari sports car, he/she might prefer the colour red. Thus, the brand influences preferences for the colour attribute. Specific experimental designs enable researchers to estimate a priori known interaction effects (e. g., Carmone and Green 1981; Green and Srinivasan 1990). Alternatively, researchers can define 'compound attributes' to handle potential interaction effects for a traditional conjoint analysis. For example, rather than having brand and colour as separate attributes, one can combine them (e. g., 'red Ferrari', 'grey Ferrari', 'red Audi' or 'grey Audi'; Green and Srinivasan 1990).

An advantage of using CBC is that it can handle interaction effects because it pools an entire experimental design across all respondents. Market researchers are thus also able to identify interaction effects post hoc, that is, after conducting a study. However, CBC is based on choices instead of more informative rankings or ratings, and compared to traditional approaches, it requires a higher number of respondents to derive stable estimates. This need to survey more respondents is even higher when expecting potential interaction effects for two reasons. First, a greater number of parameters must be estimated and second, a sufficient number of observations with specific level combinations are needed that potentially cause the interaction effect. Surveying an insufficient number of respondents may result in unstable and implausible interaction effects.

Attributes and levels must also be *understandable*, concrete and unambiguously interpretable; for example, attribute levels such as 'very expensive laptop' or 'weight is 1–2 kilos' are likely to be interpreted ambiguously, which leads to imprecise results and unreliable part-worth utilities. Researcher should thus, use concrete values such as 'price is 1,600 EUR', 'weight is 1.2 kilos' (Keeney and Gregory 2005; Orme 2013). Market researchers should also ensure that all respondents interpret the given information in the same way. It is, therefore, important to express attributes and levels in consumer-speak (see below).

If unsure that respondents understand attributes and their levels, researchers should explain them on additional

screens before the preference measurement task. Attributes that are difficult to describe using text, such as product designs, can be presented using images, films, multimedia presentations, virtual reality or real product prototypes (Dahan and Srinivasan 2000; Ernst and Sattler 2000; Green and Srinivasan 1990; Jasper 2015; Looschilder et al. 1995; Orme 2013; Stadie 1998).

Finally, levels must be *acceptable* and attributes *important* for decision making. Conjoint analysis approaches should avoid unacceptable attribute levels because their existence would violate the assumption of a linear additive utility model (see above). As noted above, some methods (ACBC and the conjunctive compensatory self-explicated approach) enable researchers to identify unacceptable levels. This information is then used to avoid biased estimates.

Moreover, attributes should be *important* and its levels *relevant* since respondents would otherwise simply ignore them or apply non-compensatory decision rules that could bias part-worth utilities. When identifying important attributes, it is essential to remember that conjoint analysis is based on the fundamental assumption that an attribute's importance is determined by the range of its respective levels relative to those of all other attributes (see Equation 2). For example, the price of a product may not have any importance, per se, and might be completely irrelevant if all available alternatives have identical prices. However, the broader the bandwidth of the attribute levels (i. e., difference between minimal and maximal price), the more important an attribute (Cattin and Wittink 1982; Goodwin and Wright 2000; Orme 2013; Steiner et al. 2016; Wilkie and Pessemier 1973). Thus, the importance of an attribute cannot be assessed without considering its respective levels.

Methods to determine potentially important attributes can be assigned to three groups: (i) approaches not based on a pre-study (ii) approaches based on qualitative pre-studies and (iii) approaches based on quantitative pre-studies.

First, methods that are not based on any pre-study share the advantages that attribute sets can be easily and quickly defined. As noted, the attribute set could simply be based on researchers or managers' instincts (Cattin and Wittink 1982). However, such an approach may only confirm or refute experts' extant beliefs. Further, it will provide limited new insight from a consumer perspective and important attributes or levels may be neglected (Gibson and Marder, 2002; Lee and Bradlow, 2011). Alternatively, market researchers could use an attribute set determined in a previous study or rely on secondary data such as advertising leaflets, websites or product reviews.

The quality of the resulting attribute set is not necessarily inferior. For example, in the case of managers, the resulting quality heavily depends on the experts' knowledge. Such an approach, however, does not guarantee that respondents understand the attributes and level descripti-

ons (Louviere 1988; for an extensive discussion of these approaches, see Steiner et al. 2016).

Second, qualitative pre-studies can also be used to define attributes for preference measurement. Qualitative methods include process tracking (i. e., respondents describe how they made a decision), projective methods (e. g., laddering interviews), in-depth personal interviews, the elicitation technique ('name all attributes that come to mind when thinking of buying a ...') and focus groups (Adamowicz et al. 2008; Cattin and Wittink 1982). Qualitative methods provide deeper insight into reasons underpinning decision making and how people come to a decision. However, the number of respondents surveyed is often too small to make predictions about attributes relevant to a specific market. Qualitative interviews, thus, may represent the first step to identifying attributes since qualitative studies do not generate results that are representative for the target population (Adamowicz et al. 2008; Steiner et al. 2016).

Third, quantitative methods include approaches such as repertory-grid, direct ratings and dual questioning. The repertory-grid technique aims at revealing similarities and dissimilarities between products. However, previous research has demonstrated that the attributes used to make judgements on similarities or dissimilarities are not identical to those used when deciding between options (Lefkoff-Hagius and Mason 1993). The direct rating of attributes ('please rate the importance of the following attributes') and dual questioning (respondents are not only asked to rate the importance but also to assess the perceived distinctiveness of attributes) do not survey information on levels that respondents considered when evaluating the attributes. Such evaluations of attribute importance weights without surveying the respective levels cannot be used to explain decision making (Goodwin and Wright 2000; Wilkie and Pessemier 1973).

In sum, none of the currently applied methods was developed to define a reasonable attribute set for conjoint analysis. In addition, no method describes the attributes on the basis of its levels. Thus, there is scope for methodological developments.

After defining a (long) list of potentially important attributes, researchers must define the **number** of attributes that respondents need to evaluate in a conjoint-analysis task. On the one hand, considering few attributes leads to less reliable results since respondents are likely to infer missing information from other attributes; for example, they may infer missing quality information from a brand attribute (Islam et al. 2007). Similarly, Gibson (2001) suggests that 'all the attributes and levels that could affect choice should be included' (p. 18) in preference measurement. Thus, the number of attributes should not be too small. On the other hand, conjoint analysis asks respondents to simultaneously consider all attributes. Respondents are more likely to use non-compensatory decision rules (e. g., by simply ignoring attributes) when being asked to evaluate alternatives that are described using

too many attributes. As a result, the utility estimates will be less reliable (Green and Srinivasan 1978).

Therefore, researchers often suggest limiting the number of attributes to a maximum of six (Lilien et al. 2013) or in certain cases, eight (Orme 2013). However, researchers conducting conjoint studies should also consider respondents' expertise and involvement with the research object. Fewer attributes should be considered if respondents' have limited expertise on the research object since the perceived complexity of the product will be high and respondents are more likely to use simplification strategies (Lines and Denstadli 2004; McCullough 2002; Payne et al. 1999). Researcher can reduce this perceived complexity by using illustrative figures (see above) or keeping the attribute and level descriptions in the conjoint tasks as brief as possible. This enables researchers to survey a higher number of attributes (Orme 2013). Moreover, fewer attributes should be assessed if respondents are less involved with the product category since this will increase the probability of them using simplifying decision rules when assessing many attributes (Curry 1997; Lines and Denstadli 2004). Thus, a pre-study on respondents' expertise and involvement can help define a reasonable number of attributes.

A common practice among researchers and managers is to define the **attribute levels** after selecting the attributes to be used in their conjoint study. In this case, a key question is how researchers select appropriate levels. Some researchers suggest that the levels cover a typical range of levels (Eggers and Sattler 2011). Lilien et al. (2013) suggested that attribute levels should cover extreme values in the relevant market. Similarly, Orme (2013) proposed that 'attribute levels should cover the full range of possibilities for relevant existing products as well as products that may not yet exist, but that you want to investigate' (p. 54).

However, including a bandwidth of levels that is too wide increases the risk of including unacceptable levels (see above). Some approaches enable researchers to identify unacceptable levels at an individual level before or during the preference measurement task and unacceptable levels can then be excluded (see above). These approaches aim at inferring respondents' decision strategies from their responses and then appropriately adjust the range of levels for the attributes. Other researchers suggested avoiding unacceptable levels at the outset to avoid as many unnecessary evaluations as possible. For example, Urban and Hauser (1980) suggested that if 'you ask consumers to evaluate products ... and you want these answers to be relevant, you must limit your questions to the evoked set' (p. 178).

Steiner et al. (2016) therefore, proposed an approach aimed at avoiding major problems in current approaches to defining attribute sets. Respondents begin with a warm-up task, in which they are asked to inform themselves about the alternatives currently on the market and asked to focus on those they are willing to buy (**evoked**



set). Next, the respondents are asked to name the products that seem acceptable and describe them on the basis of these attributes and levels they would consider to make a final decision. The resulting number of potential levels can be high. Researchers could, therefore, skip rarely mentioned levels. The emerging attribute set is then evaluated using a simplified self-explicated approach (see above). The result is a list of commonly used levels and attributes importance weights determined from the respective levels. Moreover, the approach uses the wordings consumers commonly apply when choosing a product and thus, respondents can more easily understand the resulting attribute set. In summary, this approach combines a qualitative and quantitative step and requires researchers to conduct two pre-studies, rendering it considerably more time-consuming than one-step qualitative or quantitative pre-tests.

Another alternative approach to define attribute sets was recently proposed by Rex Yuxing et al. (2015). The authors suggested using product related *search keywords* from Google Trends to determine important attributes. Their empirical findings revealed that search keywords are reasonable indicators of attributes' importance in consumers' decision-making processes. In addition, they recommended using the search keywords to identify relevant attribute levels. For example, a search for specific brand names provides information on brands that are most popular among consumers.

Finally, market researchers must decide on the number of levels to describe each attribute. Researchers are often tempted to use a large number of levels to, for example, better understand the non-linear effects of product price (e. g. price thresholds).

However, when applying a traditional conjoint analysis, the **number of levels** and attributes influence how many alternatives a respondent needs to evaluate. Therefore, researchers must limit the number of levels to avoid respondent fatigue. CBC pools the experimental design across respondents, making it possible to survey a higher number of levels. However, increasing the number of levels reduces the number of observations for each level. As a result, a researcher must either increase the sample size, which increases costs (Orme 2013), or accept that the utility estimates will be less precise and include more noise.

Many researchers have suggested selecting between two and five levels for each attribute. For quantitative levels, such as price, a researcher may interpolate between levels (Lilien et al. 2013; Orme 2013). The literature also suggests using equally spaced levels (e. g., price levels such as 1.49 EUR, 1.99 EUR, 2.49 EUR) when defining quantitative attribute levels such as price (Darmon and Rouziès 1989; Huber et al. 1992). Moreover, researchers should consider potential biases if the attributes differ by the number of levels. For example, Currim et al. (1981) found that respondents pay more attention to attributes with a higher number of levels. Thus, the higher the

number of levels, the greater the importance assigned to the attribute. This effect is known as the **number-of-levels effect**.

Therefore, market researchers should attempt to even the number of levels across all attributes (Orme 2013; Wittink et al. 1982). However, using the same number of levels for every attribute might also make the preference measurement less realistic because important levels might be missing in the preference measurement design (Wittink et al. 1989). Consequently, researchers using conjoint analysis face a trade-off decision because they want to describe the attributes using all potentially relevant levels (e. g., include all important brands) as well as avoid overestimating the importance of the attributes with a higher number of levels.

## 6. Sample Selection, Explicating the Decision Context and Warm-up Tasks

Before conducting a preference measurement study, market researchers must decide how to select a **sample**. For most products, the main goal of the sampling process is to draw a sample which is representative of the target population. Depending on the product, however, the goal might also be to select consumers potentially interested in buying the product or who have the required expertise in the product category. Thus, when conducting a preference measurement study for a new sports car, wealthy consumers and consumers who currently own a sports car of the brand in question should be over-represented in the sample because how these potential buyers evaluate its features is of particular interest.

The focus on potential buyers is important since, an immediate need for a product or service is also likely to alter the evaluation of the product features. Previous research, for example, has suggested that respondents are less risk averse and consider fewer attributes when the choice is less imminent (Wright and Weitz 1977). Moreover, respondents that currently do not plan to buy a product are likely to underestimate the importance of the price (Bornemann and Homburg 2011).

For more complex products which involve currently unknown features, respondents may need sufficient deliberation time before their preferences can be reliably measured. It is, therefore, advisable to inform respondents about the relevant attributes and features before measuring preferences to anticipate the learning process from the consumer perspective. Otherwise, respondents might be unable to understand all attributes and benefits of the respective levels. In addition, respondents should have sufficient experience with the investigated product category (American Marketing Association 1992; Huber et al. 1971; Steiner 2007). If they do not, they are likely to rely on fewer information cues and overweigh negative levels (Wright and Kriewall 1980; Wright and Weitz 1977).

Therefore, before actually measuring consumer preferences, (i) **explicating the decision context** and (ii) including **warm-up** or **practice tasks** are common to help respondents understand subsequent preference measurement tasks:

(I) When **explaining** attributes and levels to the respondents the goal is to give respondents some information about the specific **decision context**. The description of the decision situation is important because preferences are often context-dependent. For example, preferences differ depending on who will use the product (own purchase vs. gift purchase) and where the product is purchased (supermarket vs. specialty store).

Attribute and level **descriptions** can be provided using **text** (Louviere 1988, Steiner 2007) or pictures and video (Helm et al. 2012; Urban et al. 1996). Alternatively, respondents could be asked to **experience** attribute levels' benefits before measuring preferences. For example, Jasper (2015) used virtual reality so that respondents could better understand the benefits of different car headlight systems.

Information on attributes and its respective levels is often presented immediately before preference measurement (e. g. Huber et al. 1993; Jaeger et al. 2001; Louviere 1988; Steenkamp and Wittink 1994; Steiner 2007). Some other researchers provide respondents with such information few days before the measurement and ask them to inform themselves about the benefits that the levels evoke (Helm et al. 2012; Steiner 2007; Wright and Kriewall 1980). Wright and Kriewall (1980) tested the effect of such a procedure and observed that respondents who received information few days before the study provided better answers in the evaluation task than those who did not, that is, the predictive validity was higher.

In general, explicating attributes and levels helps researchers reduce respondents' uncertainty about the meaning of attribute levels and enables mental simulation (i. e., it helps decision makers imagine using the products and predicting its benefits; Jaeger et al. 2001; Wright and Kriewall 1980). Introducing attributes and levels to respondents is particularly relevant in the case of complex products with many attributes, but less relevant in decision contexts in which they already have considerable experience. For example, Jaeger et al. (2001) assessed consumers' preferences for apples (e. g., granny smith or red delicious) and observed little influence of attribute explication on predictive validity. Respondents in this study were consumers who regularly bought apples, that is, they already held stable preferences.

(II) Researchers frequently use **warm-up tasks** which are indistinguishable from the conjoint tasks used for utility estimation (Meißner et al. 2016). The main reason for including warm-up evaluation tasks is they enable respondents to practice trade-offs (Jaeger et al. 2001; Meißner et al. 2016; Steenkamp and Wittink 1994).

Elrod et al. (1992) suggested systematically designing the alternatives in such warm-up evaluation tasks. They defined alternatives that were assumed to represent the least, average and most preferred alternative, that is, respondents should be made aware of the bandwidth of levels assessed in a subsequent conjoint analysis. In addition, they recommended using an answering format which differs from the choice context but prepares respondents in making trade-off decisions. Jaeger et al. (2001), for example, presented 10 alternatives and asked respondents to categorize them into three groups (definitely buy, consider buying and would not buy) before answering the main conjoint tasks.

## 7. Validity and Reliability

Reliability and validity measures are commonly applied to evaluate the quality of preference measurement results. In the following sub-sections, we provide a brief overview of the commonly used measures.

### 7.1. Reliability

A respondent's reliability can be tested, for example, by asking respondents to evaluate the same choice tasks twice during a conjoint interview. Researchers use intervening tasks to prevent respondents from noticing such reliability checks. An alternative is asking respondents to evaluate the stimuli in a second interview few days after the initial interview is conducted. The test-retest reliability denotes consistency between both evaluations (Green and Srinivasan 1978).

### 7.2. Internal and Predictive Validity

Internal validity can be used to test the **consistency** of respondents' evaluations in the conjoint analysis task. The use of a specific internal validity measure depends on the estimation technique applied. For example, when estimating part-worth utilities using a traditional conjoint analysis and OLS regression,  $R^2$  provides a reasonable measure for respondent consistency. Alternatively, Pearson's correlation coefficients (for ratings-based evaluations), Kendall's Tau or Spearman's correlation coefficients (rankings-based evaluations) can be used to compare the observed with predicted preferences for the tested stimuli (Green and Srinivasan 1978). When applying CBC, the root likelihood value (RLH) can be used to assess how well the estimates fit the data (Orme 2013).

The most dependable option to test the quality of a method's estimates is to assess its ability to predict real choices (see the section on external validity). However, it is often impossible to survey real purchase decisions, and therefore, market researchers commonly focus on predictive validity as a surrogate. Predictive validity is analysed by including additional evaluation tasks in the questionnaire that are used as a benchmark. Two types of benchmark are common: hold-out tasks and reference methods.

Item	References
Amount of information, information content (1 = too less, 9 = sufficient)	Day 1975; Smead et al. 1981; Bettman and Zins 1979; León 1997; Steenkamp and Van Trijp 1997; Bottomley et al. 2000
Degree of realism (1 = low, 9 = high)	Smead et al. 1981; Bettman and Zins 1979; Green and Srinivasan 1978; León 1997; Steenkamp and Van Trijp 1997
Certainty when evaluating alternatives (1 = low, 9 = high)	Smead et al. 1981; Bettman and Zins 1979; León 1997; Bottomley et al. 2000
Difficulty of evaluation (1 = high, 9 = low)	Day 1975; Smead et al. 1981; Bettman and Zins 1979; Johnson 1987; León 1997; Bottomley et al. 2000; Steenkamp and Van Trijp 1997
Enjoyment (1 = low, 9 = high)	León 1997; Bottomley et al. 2000; Steenkamp and Van Trijp 1997
Perceived diversity (1 = boring, 9 = diversified)	Green and Srinivasan 1978; Johnson 1987; Steenkamp and Van Trijp 1997

Tab. 4: Frequently used applicability measures

Hold-out tasks are evaluation tasks that are indistinguishable from the conjoint tasks but are not used when estimating utilities (Orme et al. 1997). Reference methods ask respondents to perform evaluation tasks differently from those of conjoint tasks. Thus, reference methods aim at assessing respondents' consistency not only with respect to a specific evaluation task type but also across different task types (Steiner 2007). The alternatives used in hold-out tasks and reference methods are often randomly designed (Green et al. 1993; Orme et al. 1997).

Market researcher can apply several **hit rates** to assess how well the estimated part-worth utilities predict evaluations in the hold-out tasks or reference method. The most often used is the 'first choice hit rate' (HR1): it tests how well part-worth utilities predict the preferred alternative in a given hold-out task. HR12 and HR123 indicate how often the first and second or the first, second and third alternative are correctly predicted when comparing the estimated total utilities with a reference method (Scholl et al. 2005).

Internal and predictive validity are important measures that assess respondents' consistency. For example, researchers can use consistency measures to identify respondents who randomly answered the conjoint tasks, following highly inconsistent respondents can be removed from the dataset. However, consistency measures do not provide information on whether the additivity assumption of the additive utility model is appropriate. For example, respondents who focus on subset of attributes while ignoring others (i. e., those applying a simplifying decision rule) might give highly consistent answers. A respondent might always choose an alternative with the lowest price and thus is highly consistent in his/her answers. However, the respondent's answers provide little insight how the respondent will trade-off attributes.

7.3. Applicability measures

Applicability measures are commonly used to obtain information on respondents' subjective interview experience. It is important to consider respondents' perceptions of a preference measurement study because subjective reactions may affect the data quality; in particular, negative perceptions are likely to induce biased estimates

(Day 1975; Bettman and Zins 1979; McDaniel et al. 1985).

Information on the perceived applicability of an evaluation task tends to be surveyed directly after the preference measurement, which is based on direct evaluations using a rating scale (Smead et al. 1981). Tab. 4 presents items used frequently to assess the applicability of alternative preference measurement approaches.

*Amount of information* does not refer to the absolute number of information cues but to the ability of respondents to choose an alternative that fits their needs. Thus, it relates to the researcher's definition of the attribute set, which should allow consumers to select alternatives they consider important (Day 1975; Bottomley et al. 2000).

Unrealistic stimuli (low *degree of realism*) should be avoided to reduce respondent burden arising from unnecessary evaluations (Mehta et al. 1992). However, it is often impossible to avoid unrealistic alternatives (e. g., high-priced, low-quality cars) when creating experimental designs. Some studies indicate that unrealistic alternatives might have little influence on predictive validity (Moore and Holbrook 1990).

*Certainty* measures perceived confidence during the evaluation of stimuli within a conjoint analysis. *Perceived difficulty* assesses the cognitive effort required when evaluating stimuli. Both constructs are related and low certainty and high perceived difficulties are indicators of the potential degree of respondents' dependence on simplifying decision rules (Hoeffler and Ariely 1999; Thaler and Sunstein 2009; Spiegler 2011). Moreover, applicability measures often address items that are likely to influence respondents' motivation (*enjoyment* of evaluation task and *perceived diversity*).

7.4. Face validity

Face validity assesses whether estimated part-worth utilities 'look right' from a researcher's perspective (Churchill 1979), because researchers and experts will have a priori expectations regarding part-worth utilities (Acito and Jain 1980; Green and Srinivasan 1978; Scholl et al. 2005). For example, it is reasonable to assume that respondents prefer lower prices over higher ones (if all



Tab. 5: Example dummy coding of part-worth utilities

Colour		Brand	
White	0	A	0
Black	20	B	-20
Grey	40	C	20
Red	10	D	10

other attribute levels of an alternative are equal). Obviously, it is not possible to test all attributes' face validity since there are often attributes without an unambiguous a priori preference order for their levels (e. g., preference for a specific colour). Market researchers need to focus on attributes with a clear a priori ranking of their levels to assess the face validity of the data.

8. Interpretation of Part-Worth Utilities

Conjoint analysis utility estimates are interval scaled. Before interpreting the data, it is important to note that the utility estimates do not include a 'natural' zero point, which means rescaling each part-worth utility value will not influence the results' interpretation (we discuss exceptions below). This insight is important when interpreting utility estimates because the data might be coded in different ways.

For traditional conjoint analysis approaches, **dummy coding** is the most common coding approach. For example, when applying a regression analysis to estimate part-worth utilities, one level of each attribute is set to '0'. All other estimated part-worth utilities can then be interpreted in reference to this level.

Let us explore the following example which presents part-worth utilities from a traditional conjoint analysis estimated using a regression analysis (see Tab. 5). In regression analysis, a constant term and attribute levels' part-worth utilities are estimated (assume the constant term is 100). The constant term includes the utility a respondent derives from attribute levels not considered within a conjoint task and all levels a researcher defines as reference level (here, the colour white and brand A). As a consequence, the part-worth utilities of the reference levels remain unknown. For example, with respect to colour, we do not know the utility a consumer derives from 'white'.

The part-worth utilities for the remaining colours (black and grey) represent the change in utility from the reference (white) to a respective level. For example, utility increases from white to grey by 40 units. Thus, it is important to note that these 40 units do not represent the total utility of the colour grey. The total value of 'grey' would be the sum of utility for 'white' (unknown) and part-worth utility of 'grey'. Similarly, part-worth utilities from the 'brand' attribute indicate that the respondent preferred brand C over A and brand B was the least preferable brand. Here, utility increases by 40 units when switching from brand B to C.

Thus, part-worth utilities provide information on the change in utility induced by attribute levels. However, they provide no insight on the total utility consumers derive from a level. This is important since this implies that researchers cannot simply compare part-worth utilities between attributes. Thus, it would be wrong to imply, for example, that brand C provides twice as much value compared to the colour red. Moreover, researchers cannot conclude that brand C provides a higher utility than the colour red.

Another coding type commonly used with CBC is **effects coding**. When applying effects coding, each attribute's part-worth utilities are scaled to sum to zero. Positive values denote levels that were preferred, while negative ones indicate less preferable levels. However, negative values do not mean that the respective levels have no (or a negative) value to the respondent. The negative values are merely a result of the effects coding as the sum of an attribute's part-worth utilities must equal zero. Similar to dummy coding, part-worth utilities do not provide information on the absolute utility of a level but on an increase or decrease caused by a certain level compared to another level of the same attribute. Here as well, comparisons between attributes are not possible (Orme 2013).

As noted, a **traditional conjoint** analysis based on ratings suffers from the possibility of respondents differently interpreting rating scales. For example, some respondents might avoid extreme points on rating scales. **Standardizing** such data aligns the scale across respondents and enables researchers to aggregate the data (Sawtooth Software 2009; Steiner 2007).

For a traditional conjoint analysis, Equation 3 represents a common approach to standardizing part-worth utilities that result from measurements with rating scales (several other approaches can be applied as well; for an overview, see Steiner 2007). Commonly, researcher scale part-worth utilities and set the least preferred level to zero to ease interpretation (see Equation 3; Green and Krieger 1991; Green et al. 2001; Johnson 1987). Equation 3 also divides the respective part-worth utilities by the sum of all part-worth utilities of all attributes. As a result, the sum of all standardized part-worth utilities equals one for all respondents.

$$\beta_{ijk}^{std} = \frac{\beta_{ijk} + |\min \beta_{ij}|}{\sum_{j=1}^J \sum_{k=1}^{K_j} (\beta_{ijk} + |\min \beta_{ij}|)}$$
 (3)

As noted, for CBC, part-worth utilities are commonly zero-centred at the respondent level. For each attribute,

part-worth utilities are scaled to sum to zero. Moreover, across attributes, differences between most and least preferred levels average 100. Researchers that fear the misinterpretation of negative values by the management could reset the least preferred level to zero (see denominator in Equation 3).

9. Market Simulations and External Validity

Part-worth utilities provide information on how desirable attribute levels are. However, from a managerial perspective, predictions of consumers' potential future behaviour are of greater interest. Market simulations provide information on the relative share of respondents who prefer predefined products in a certain competitive environment. They enable managers to test alternative market scenarios. For example, managers can assess price-demand curves, the impact of product adjustments, competitive pressure between products and potential cannibalization effects of line extensions (Green and Krieger 1988). Conducting a market simulation begins with defining relevant products. Then, the total utility of these products is computed at the individual or target group level (Green and Krieger 1988). As described in Equation 1, the total product utility is the sum of its part-worth utilities.

Researchers conducting a CBC may compare the total utility value of products to that of the 'none' option. The higher the difference between the total utility of an alternative and the utility of the 'none' option, the more likely it is for respondents to accept the alternative. A product's total utility below the 'none' option's value indicates that respondents are more likely to not accept the offer.

CBC is based on choices, and thus, researchers can easily apply the **logit model** to estimate market shares. Market shares are predicted by simply exponentiating the total utility of a product ( $U_{im}$  is the total utility of  $U$  with  $m$  levels for respondent  $i$ ) and then dividing this value by the sum of all products' exponentiated values and the none option (see Equation 4 and Tab. 6 for an example; Green and Krieger 1988; McFadden 1976; Sawtooth Software 2013b).

$$C_{im}^{prob} = \frac{\exp(U_{im})}{\sum_{m=1}^M \exp(U_{im})} \tag{4}$$

Researchers that apply a traditional conjoint analysis need to transform total product utilities into choices using common decision rules. In addition, researchers' applying effects coding (see above) should first normalize part-worth utilities. **Normalizing** the data is necessary since applying effects coding may result in negative total utilities (product 3 in Tab. 6) and as a result, market share predictions would be negative. Data is often normalized by assigning a zero to the least preferred level of each attribute (Green and Krieger 1988; also see above).

Moreover, depending on the estimation technique used (e. g. OLS regression), a researcher may also estimate a constant term. This constant term must also be considered when computing the products' total utility because part-worth utilities only provide insight into the increase in utility relative to a reference level (see above dummy or effect coding). While some researchers ignore the constant term (e. g. Lund et al. 1988), Green and Krieger (1988) demonstrated the effect of (not) accounting for the constant term (see below for an example).

As noted, researchers then need to select a **decision rule** that converts utility values into choice probabilities. Two types of rules are common: deterministic and probabilistic rules. Deterministic rules assign a choice probability of 100 % to a product, while all other alternatives are chosen with a probability of 0 %. The total market share then represents the share of decision makers who selected the respective product.

The simplest decision rule is the **'first choice'** rule (or max utility rule). This rule assumes that a respondent will always choose an alternative with the highest utility. The first choice rule, thus, only considers the frequency of preferring an alternative while ignoring the utilities of all others. Moreover, the first choice rule assumes perfectly rational decision makers, that is, consumers will choose an alternative with the highest utility with a probability of 100 %, even if the alternatives only differ slightly in total utility value. However, in reality, humans often behave differently and thus, do not always choose products that have the highest utility. People make choices with variability even when asked to select the preferred alternative from identical choice sets (Huber and Miller 1999). As a result, first choice models tend to overestimate the market shares of preferred products (Johnson and Orme 2003; Orme and Baker 2000). Therefore, some researchers have suggested discontinuing the use of the first choice model (Sawtooth Software 2003).

	Total utility	Exp. (total)	Market Share
Product 1	0,800	2,226	0,353
Product 2	0,500	1,649	0,262
Product 3	-0,500	0,607	0,096
None	0,600	1,822	0,289
Sum		6,303	1

Tab. 6: Market share estimation for total utilities based on CBC

		Total utility			Choice			First Choice	Market Share				Logit			
		R1	R2	R3	R1	R2	R3		BTL				R1	R2	R3	Aggregated
Scenario 1	Product 1	0.8	0.3	0.1	100%	0%	0%	33%	80%	27%	10%	39%	65%	38%	31%	44%
	Product 2	0.2	0.8	0.9	0%	100%	100%	67%	20%	73%	90%	61%	35%	62%	69%	56%
Scenario 2	Product 1	3.8	3.3	3.1	100%	0%	0%	33%	54%	46%	44%	48%	65%	38%	31%	44%
	Product 2	3.2	3.8	3.9	0%	100%	100%	67%	46%	54%	56%	52%	35%	62%	69%	56%
Scenario 3	Product 1	2.4	0.9	0.3	100%	0%	0%	33%	80%	27%	10%	39%	86%	18%	8%	37%
	Product 2	0.6	2.4	2.7	0%	100%	100%	67%	20%	73%	90%	61%	14%	82%	92%	63%

Tab. 7: Market share estimation based on first choice, BTL and logit rules

Despite its potential problems, researchers commonly rely on the first choice rule when testing a conjoint analysis' external validity and/or benchmarking alternative preference measurement techniques (see, for example, Chapman et al. 2009; Lund et al. 1988; Srinivasan and deMaCarty 1999).

In contrast, probabilistic (share of utility) rules assign choice probabilities to all alternatives. For example, the **BTL model** (Bradley and Terry 1952; Luce 1959) assumes that any option has some probability of being chosen. Its most common version is identical to the **share of preference** model (Orme and Baker 2000). The probability of an alternative  $m$  being chosen by respondent  $i$  is the utility of  $m$  divided by that of all alternatives (see Equation 5). Choice probabilities at the individual level are then averaged across respondents to compute overall market share values.

$$C_{im}^{prob} = \frac{\exp(U_{im})}{\sum_{m=1}^M \exp(U_{im})} \quad (5)$$

The **logit rule** described above can also be applied to predict market shares for estimates using a traditional conjoint analysis. The BTL and logit rule often result in similar market share estimates (Green and Krieger 1988).

It is noteworthy that the BTL and logit rule are sensitive to the approach used to standardize data. The example in Tab. 7 presents the utilities of three respondents (R1, R2, and R3). It is similar to that in Green and Krieger (1988) and demonstrates certain potential effects of standardizing data (i. e. considering an additive or multiplicative constant). Scenario 1 denotes a base scenario while in scenario 2, a constant term of 3 is added to all utilities. In scenario 3 all values are multiplied by 3. Market share estimates based on the first choice rule are not influenced by such changes since standardizing data does not alter the rank of the preferred alternative. However, adding a constant term impacts market share estimates for the BTL rule. Adding a constant term reduces differences in the total product utilities. As a result, differences in market share estimates diminish. Adding a constant term does not influence market share estimates that are based on a logit rule. In contrast, multiplying all utility values

(scenario 3) does not influence BTL market share estimates but heightens differences for market share estimates derived by the logit rule. In summary, market researchers' methods of standardizing data influences market share estimates for BTL and logit choice rules.

Moreover, both BTL and logit rules suffer from the 'share inflation' problem if similar products exist in the market. A commonly cited example is the **red bus/blue bus problem**. Imagine a market with two options, cars and red busses, both of which have a market share of 50 %. Now assume that a new option, a blue bus, is introduced. The logit and BTL rules predict that this new option (blue bus) will obtain an equal share as those of the red bus and cars. As a result, the market share for busses will be 67 % and those for cars will shrink to 33 %. Such predictions are, however, unrealistic. It is more reasonable to assume that introducing a new colour for busses should not influence its market share, which should, remain close to 50 %. In other words, market share estimates should be **independent from irrelevant alternatives** (IIA problem). As noted, BTL and logit rules are influenced by IIA, while this problem does not hold true for the first choice rule (Huber and Miller 1999; Johnson and Orme 2003; Orme and Baker 2000).

In summary, first choice rules seem unrealistic since they assume 'perfectly rational' decision makers. For traditional conjoint analysis approaches, researchers may influence market share estimates for BTL and logit rules depending on how they standardize data. Moreover, BTL and logit rules suffer from IIA. However, in practice, these biases can be avoided, to a certain extent, when using individual- or at least, group-level preference data. However, there is no 'gold standard' rule when predicting market shares. Researchers commonly apply different approaches and select a method that produces share estimates that fit real market shares. Moreover, market shares can be 'tuned' to better represent real market shares, for example, by using a multiplier, i. e., defining a value with which all estimates are multiplied to better reflect real market shares (Huber and Miller 1999; Sawtooth Software 2003). Alternatively, researchers can test which rule best predicts hold-out sets and use this rule to predict market share.



All three types: the first choice rule, BTL rule, and the logit rule are commonly applied in research and practice. The randomized first-choice model (RFC) represents a technique that, thus far, has been used less often. RFC is based on the traditional first choice rule but accounts for the fact that any consumer choice is, to a certain degree, influenced by variability (by adding a random error). In essence, RFC is based on repeated market share predictions at the individual (or group) level while adding product variation and attribute variation in all iterations. Here, hundreds or thousands of iterations are drawn. Market share estimates are then computed as an average of all iterations.

RFC adds attribute variability (a random error to part-worth utilities) since respondents may not always pay the same attention to specific features (e. g., respondents may pay more attention to selected features on certain occasions and to price on some others). Moreover, adding an identical random error to products that share the same levels results in a stronger correlation and therefore, stronger competition. This avoids IIA biases (Huber and Miller 1999; Orme 2013; Orme and Baker 2000).

Finally, preference measurement data can be used to define an 'optimal' product or product portfolios. These approaches do not only rely on preference data but also consider costs and can be used to increase profitability (for more details, see Sawtooth Software 2003).

Market researchers can also use estimates to evaluate the **external validity** of preference measurement, that is, the ability to predict real choice behaviour. To evaluate the performance of preference measurement approaches, researchers compare stated preferences with subsequent **real choices** (Louviere and Timmermans 1992; Louviere and Woodworth 1983; Tscheulin 1991; Krishnamurthi 1988; Srinivasan and Park 1997) or willingness to pay (Brzoska 2003; Schlag 2008) at individual levels. All of these studies are based on a within-subject design. Alternatively, other studies have assessed the ability to predict market shares (Lund et al. 1988; Louviere and Timmermans 1992; Louviere and Woodworth 1983; Natter and Feurstein 2002; Parker and Srinivasan 1976). Thus, they aim to provide insight on whether revealed purchase intentions within surveys can be used to predict future market success beyond the sample surveyed.

As noted, numerous studies have demonstrated the general ability of common **preference measurement approaches** to predict real market behaviour, that is, preference elicitation techniques can have a high external validity. The extant literature has also suggested that traditional conjoint analysis methods well predict real choices (Benbenisty 1983; Krishnamurthi 1988; Parker and Srinivasan 1976; Robinson 1980; Srinivasan and Park 1997; Tscheulin 1991), while the same holds true for CBC (Chapman et al. 2009; Louviere and Timmermans 1992; Natter and Feurstein 2002). Chapman et al. (2009) compared ACBC with CBC and observed similar predictions with ACBC, revealing market share estimates that were

only slightly closer to reality. However, even simple compensatory preference measurement approaches seem to provide reasonable estimates (Srinivasan and Park 1997; Srinivasan and de MaCarty 1999).

With respect to specific **estimation techniques**, previous research suggests that even simple aggregate models seem to provide reasonable market share predictions and that there is little difference between the methods applied (Natter and Feurstein 2002). Natter and Feurstein (2002) used CBC data; computed part-worth utilities using an aggregate model, latent-class approach and a hierarchical Bayes approach. They then compared market share predictions with real market shares. They concluded that none of the three estimation techniques outperformed the other and that all provided good estimates. Similar results were found for different estimation techniques used to analyse data on the basis of traditional conjoint analysis approaches (Kamakura and Ozer 2000; Vriens et al. 1996).

In summary, conjoint analysis and even simple approaches seem feasible methods to predict real choices. However, sometimes they fail to do so. The reasons underpinning potential differences between predicted and real market behaviour are multifold. As noted in the previous sections, the selection of attributes, the experimental design, and the presentation format also influence conjoint analysis estimates. For example, conjoint analysis results can only reasonably predict real behaviour if marketers considered all relevant attributes and their respective levels in the study. Moreover, market simulators are based on the assumption that only product features (levels) and their respective prices influence product choice. As such, market simulators do not consider the potential effects of other marketing-mix elements. For example, market simulators do not account for differences arising from distribution (e. g., limited availability due to stock-outs, differences in sales force effectiveness, ubiquity of product and market coverage) and communication (differences in product or brand awareness). Moreover, conjoint market simulations do not consider different market-entry strategies (entering a market earlier or later influences sales; Sawtooth Software 2006).

Wong and Sheth (1985) proposed a framework that helps explaining differences between intentions and actual behaviour. It considers four main dimensions: unexpected events, personal characteristics, social environment, and involvement. 'Unexpected events' refers to the influence of the in-store choice environment on decision making. For example, product unavailability (stock-outs), sales promotions and other price changes and time pressure might influence ultimate choices. Moreover, Wong and Sheth (1985) suggested that the choice environment's impact is influenced by personal characteristics. For example, certain people are less effective in self-regulating their purchase behaviour and tend to engage more in impulsive decision making. Impulsive decision making is less predictable than deliberate choices; thus, the gap

between choice predictions and real behaviour might be higher. Social environment denotes the influence of others on consumers' decision making. For example, the brand people buy as gifts might differ from one they commonly buy for themselves. Finally, decision makers' product involvement also influences the relationship between intentions and real behaviour. The fundamental reasoning is similar to personal characteristics; that is, consumers will engage in more extensive decision making if the choice task involves a financial, functional, or social risk. Thus, decisions tend to be based on product features. However, if a choice task involves limited risk, consumers are likely to engage in impulsive decision making.

Finally, and in line with all empirical works, sampling also influences results' generalizability. To predict real market behaviour, respondents must be representative of the population or target group of interest (Sawtooth Software 2006).

In sum, conjoint analysis predictions might fail because product choice is not only influenced by factors that can be considered within conjoint analysis studies (e.g., product features, price, experimental design and conjoint analysis type) but also numerous other factors. Researchers, therefore, suggested that conjoint simulation results should not be interpreted as predictions of market shares but as the **share of interest** a product is likely to spark in the market (Sawtooth Software 2006). Market researchers who aim at predicting market shares and know that the above-mentioned assumptions are violated may adjust market share predictions using the above mentioned multipliers to better capture differences in, for example, brand awareness, distribution level, and differences in purchase frequency or amount (for a detailed description, see Sawtooth Software 2006).

## 10. Conclusions

As recently noted by Eggers et al. (2016), determining the factors that influence the validity of preference measurement results should be a key goal for researchers in the field. For example, realistic images of the attributes and levels as well as video instructions may increase the validity of a preference measurement study. The influence of these craft factors might, in some cases, affect the results' validity more strongly than the selection of the preference measurement approach. We, therefore, argue that more research should be warranted to determine the craft factors that need special attention. This paper explicates important craft factors such as the definition of attribute sets, the explication of decision context and use of warm-up tasks. We, therefore, contribute to this important discussion, although without empirically testing these potential craft factors.

Our paper is intended to serve as a user's guide to the 'galaxy' of consumer preference measurement. To elabo-

rate, we highlight topics we consider important for both academics and practitioners, but neglect others because of space restrictions. We highlight potential pitfalls on the basis of our own experience with preference measurement studies and hope that our guide will help potential users of preference measurement approaches to make better decisions when setting up their preference measurement studies.

## Notes

- [1] Addelman plans are implemented in SPSS. Conjoint experimental designs can be easily developed using the Orthoplan command.

## References

- Acito, F., & Jain, A.K. (1980). Evaluation of Conjoint Analysis Results: A Comparison of Methods. *Journal of Marketing Research*, 17(1), 106–112.
- Adamowicz, W., Bunch, D., Cameron, T. A., Dellaert, B. G. C., Hanneman, M., Keane, M., Louviere, J., Meyer, R., Steenburgh, T., & Swait, J. (2008). Behavioral Frontiers in Choice Modeling. *Marketing Letters*, 19(3/4), 215–228.
- Addelman, S. (1962). Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments. *Technometrics*, 4(2), 21–46.
- American Marketing Association (1992). *Conjoint Analysis: A Guide for Designing and Interpreting Conjoint Studies*, Chicago: American Marketing Association.
- Backhaus, K., Becker, J., Beverungen, D., Frohs, M., Müller, O., Weddeling, M., Knackstedt, R., & Steiner, M. (2010). Enabling Individualized Recommendations and Dynamic Pricing of Value-Added Services through Willingness-To-Pay Data. *Electronic Markets*, 20(2), 131–146.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2015a). *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. 14th Ed., Berlin: Springer.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2015b). *Fortgeschrittene multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. 3rd Ed., Berlin: Springer.
- Baier, D., & Brusch, M. (2009). *Conjointanalyse: Methoden-Anwendungen-Praxisbeispiele*. Heidelberg: Springer.
- Bangcuyo, R. G., Smith, K. J., Zumach, J. L., Pierce, A. M., Guttman, G. A., & Simons, C. T. (2015). The Use of Immersive Technologies to Improve Consumer Testing: The Role of Ecological Validity, Context and Engagement in Evaluating Coffee. *Food Quality and Preference*, 41, 84–95.
- Benbenisty, R. L. (1983). Attitude Research, Conjoint Analysis Guided Ma Bell's Entry Into-Terminal Market. *Marketing News*, 17(10), 12.
- Bettman, J. R., & Zins, M. A. (1979). Information Format and Choice Task Effects in Decision Making. *Journal of Consumer Research*, 6(2), 141–153.
- Bornemann, T., & Homburg, C. (2011). Psychological Distance and the Dual Role of Price. *Journal of Consumer Research*, 38(3), 490–504.
- Bottomley, P. A., Doyle, J. R., & Green, R. H. (2000). Testing the Reliability of Weight Elicitation Methods: Direct Rating versus Point Allocation. *Journal of Marketing Research*, 37(4), 508–513.
- Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs, I. The Method of Paired Comparisons. *Biometrika*, 39(3/4), 324–345.
- Bradlow, E. T. (2005). Current Issues and a "Wish List" for Conjoint Analysis. *Applied Stochastic Models in Business & Industry*, 21(4/5), 319–323.
- Brzoska, L. (2003). *Die Conjoint-Analyse als Instrument zur Prognose von Preisreaktionen*, Hamburg: Verlag Dr. Kovac.

- Cameron, L., Cragg, M., & McFadden, D. (2013). *The Role of Conjoint Surveys in Reasonable Royalty Cases*. Retrieved May 9, 2017, from <https://www.law360.com/articles/475390/the-role-of-conjoint-surveys-in-reasonable-royalty-cases>.
- Carmone, F. J., & Green, P. E. (1981). Model Misspecification in Multiattribute Parameter Estimation. *Journal of Marketing Research*, 18(1), 87–93.
- Cattin, P., & Wittink, D. R. (1982). Commercial Use of Conjoint Analysis: A Survey. *Journal of Marketing*, 46(3), 44–53.
- Chapman, C. N., Love, E., & Alford, J. L. (2008). Quantitative Early-Phase User Research Methods: Hard Data for Initial Product Design. *Proceedings of the 41<sup>st</sup> Hawaii International Conference on System Sciences* Waikoloa, Hawaii: University of Hawaii.
- Chapman, C. N., Alford, J. L., Johnson, C., Weidmann, R., & Lahav, M. (2009). CBC vs. ACBC: Comparing Results with Real Product Selection. *Sawtooth Software Research Paper*, Sequim: Sawtooth Software.
- Churchill Jr., G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, 16(1), 64–73.
- Cui, D., & Curry, D. (2005). Prediction in Marketing Using the Support Vector Machine. *Marketing Science*, 24(4), 595–615.
- Currim, I. S., Weinberg, C. B., & Wittink, D. R. (1981). Design of Subscription Programs for a Performing Arts Series. *Journal of Consumer Research*, 8(1), 67–75.
- Curry, J. (1997). After the Basics. *Marketing Research*, 9(1), 6–11.
- Dahan, E., & Srinivasan, V. S. (2000). The Predictive Power of Internet-Based Product Concept Testing using Visual Depiction and Animation. *Journal of Product Innovation Management*, 17(2), 99–109.
- Darmon, R. Y., & Rouziès, D. (1989). Assessing Conjoint Analysis Internal Validity: The Effect of Various Continuous Attribute Level Spacings. *International Journal of Research in Marketing*, 6(1), 35–44.
- Day, G. S. (1975). The Threats to Marketing Research. *Journal of Marketing Research*, 12(4), 462–467.
- De Bruyn, A., Liechty, J. C., Huizingh, E. K. R. E., & Lilien, G. L. (2008). Offering Online Recommendations with Minimum Customer Input through Conjoint-based Decision Aids. *Marketing Science*, 27(3), 443–460.
- Drechsler, W., Natter, M., & Leeftang, P. S. H. (2013). Improving Marketing's Contribution to New Product Development. *Journal of Product Innovation Management*, 30(2), 298–315.
- Eggers, F., Hauser, J. H., & Selove, M. (2016). The Effects of Incentive Alignment, Realistic Images, Video Instructions, and Ceteris Paribus Instructions on Willingness to Pay and Price Equilibria. *Proceedings of the Sawtooth Software Conference*, Orem: Sawtooth Software.
- Eggers, F., & Sattler, H. (2011). Preference Measurement with Conjoint Analysis: Overview of State-Of-The-Art Approaches and Recent Developments. *GfK-Marketing Intelligence Review*, 3(1), 36–47.
- Elrod, T., Louviere, J. J., & Davey, K. S. (1992). An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models. *Journal of Marketing Research*, 29(3), 368–377.
- Elshiewy, O., Guhl, D., & Boztuğ, Y. (2017). Multinomial Logit Models in Marketing – From Fundamentals to State-of-the-Art. *Marketing ZFP – Journal of Research and Management*, 39(3), 32–49.
- Ernst, O., & Sattler, H. (2000). Multimediale versus traditionelle Conjoint-Analysen. *Marketing ZFP*, 22(2), 161–172.
- Evgeniou, T., Boussios, C., & Zacharia, G. (2005). Generalized Robust Conjoint Estimation. *Marketing Science*, 24(3), 415–429.
- Fine, L. M. (2009). The Bottom Line: Marketing and Firm Performance. *Business Horizons*, 52(3), 209–14.
- Gibson, L. D. (2001). What's Wrong with Conjoint Analysis? *Marketing Research*, 13(4), 16–19.
- Gibson, L., & Marder, E. (2002). Revisiting "What's Wrong...". *Marketing Research*, 14(4), 47–48.
- Goodwin, P., & Wright, G. (2000). *Decision Analysis for Management Judgment*. New York: Wiley.
- Green, P. E., & Devita, M. T. (1974). A Complementarity Model of Consumer Utility for Item Collections. *Journal of Consumer Research*, 1(3), 56–67.
- Green, P. E., & Krieger, A. M. (1988). Choice Rules and Sensitivity Analysis in Conjoint Simulators. *Journal of the Academy of Marketing Science*, 16(2), 114–127.
- Green, P. E., & Krieger, A. M. (1991). Segmenting Markets with Conjoint Analysis. *Journal of Marketing*, 55(4), 20–31.
- Green, P. E., Krieger, A. M., & Agarwal, M. K. (1993). A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences. *Marketing Letters*, 4(4), 369–380.
- Green, P. E., & Krieger, A. M. (1995). Attribute Importance Weights Modification in Assessing a Brand's Competitive Potential. *Marketing Science*, 14(3), 253–270.
- Green, P. E., Krieger, A. M., & Vavra, T. G. (1997). Evaluating New Products – Conjoint Analysis offers a Level of Detail that few Concept Testing Methods can match. *Marketing Research*, 9(4), 12–21.
- Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty Years of Conjoint-Analysis: Reflections and Prospects. *Interfaces*, 31(3), 56–73.
- Green, P. E., & Rao, V. R. (1971). Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research*, 8(3), 355–363.
- Green, P. E., & Srinivasan, V. S. (1978). Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research*, 5(2), 103–123.
- Green, P. E., & Srinivasan, V. S. (1990). Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *Journal of Marketing*, 54(4), 3–19.
- Green, P. E., Wind, Y., & Jain, A. K. (1972). Preference Measurement of Item Collections. *Journal of Marketing Research*, 9(4), 371–377.
- Gustafsson, A., Herrmann, A., & Huber, F. (2013). *Conjoint Measurement – Methods and Applications*. Berlin: Springer.
- Jair, J. F.; Black, W. C.; Babin, B. J., & Anderson, R. E. (2009) *Multivariate Data Analysis*. Harlow: Pearson Education.
- Hartmann, A., & Sattler, H. (2002). Preference Measurement in Marketing Research Practice. *Proceedings of the 31. Annual Conference of the European Marketing Academy*, Rom: European Marketing Academy.
- Hausruckinger, G., & Herker, A. (1992). Die Konstruktion von Schätzdesigns für Conjoint-Analysen auf Basis von Paarvergleichen. *Marketing ZFP*, 14(2), 99–110.
- Helm, R., Bühren, C., & Pabst, R. (2012). Using Warm-Up Stages Before Measuring Preferences for Innovative Products. *International Journal of Product Development*, 15(4), 227–248.
- Helm, R., Scholl, A., Manthey, L., & Steiner, M. (2004). 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. (2008). A Comparative Empirical Study on Common Methods for Measuring Preferences. *International Journal of Management and Decision Making*, 9(3), 242–265.
- Hoeffler, S., & Ariely, D. (1999). Constructing Stable Preferences: A Look into Dimensions of Experience and Their Impact on Preference Stability. *Journal of Consumer Psychology*, 8(2), 113–139.
- Huber, G. P., Daneshgar, R., & Ford, D. L. (1971). An Empirical Comparison of Five Utility Models for Predicting Job Preferences. *Organizational Behavior & Human Performance*, 6(3), 267–282.
- Huber, J., Orme, B., & Miller, R. (1999). Dealing with Product Similarity in Conjoint Simulations, In Sawtooth Software (Ed.), *Sawtooth Software Conference Proceedings*, Sequim: Sawtooth Software.



- Huber, J., Wittink, D. R., Fiedler, J. A., & Miller, R. (1993). The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice. *Journal of Marketing Research*, 30(1), 105–114.
- Huber, J., Wittink, D. R., Johnson, R. M., & Miller, R. (1992). Learning Effects in Preference Tasks: Choice-Based Versus Standard Conjoint. *Sawtooth Software Research Paper*, Sequim: Sawtooth Software.
- Huber, J., & Zwerina, K. (1996). The Importance of Utility Balance in Efficient Choice Designs. *Journal of Marketing Research*, 33(3), 307–317.
- Islam, T., Louviere, J. J., & Burke, P. F. (2007). Modeling the Effects of Including/Excluding Attributes in Choice Experiments on Systematic and Random Components. *International Journal of Research in Marketing*, 24(4), 289–300.
- Jaeger, S. R., Hedderley, D., & MacFie, H. J. H. (2001). Methodological Issues in Conjoint Analysis: A Case Study. *European Journal of Marketing*, 35(11/12), 1217–1237.
- Jasper, J. (2015). *Virtual Reality-Based Product Representations in Conjoint Analysis*, Hamburg: Verlag Dr. Kovac.
- Johnson, G. J., & Ambrose, P. J. (2009). Welcome to the Bazaar. *Marketing Research*, 21(4), 18–24.
- Johnson, R. M. (1974). Trade-Off Analysis of Consumer Values. *Journal of Marketing Research*, 11(2), 121–127.
- Johnson, R. M. (1987). Adaptive Conjoint Analysis. In Sawtooth Software (Ed.), *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, Ketchum: Sawtooth Software, 253–265.
- Johnson, R. M. (2001). History of ACA. *Sawtooth Software Research Paper*, Sequim: Sawtooth Software.
- Johnson, R. M., Huber, J., & Bacon, L. (2003). Adaptive Choice-Based Conjoint. In Sawtooth Software (Ed.), *Proceedings of the Sawtooth Software Conference*, Ketchum: Sawtooth Software, 333–343.
- Johnson, R. M., Huber, J., & Orme, B. (2004). A Second Test of Adaptive Choice-Based Conjoint Analysis (The Surprising Robustness of Standard CBC Designs). In Sawtooth Software (Ed.), *Proceedings of the Sawtooth Software Conference*, Ketchum: Sawtooth Software, 217–234.
- Johnson, R. M., & Orme, B. K. (2003). Getting the Most from CBC. *Sawtooth Software Research Paper*, Sequim: Sawtooth Software.
- Johnson, R. M., & Orme, B. K. (2007). A New Approach to Adaptive CBC. *Sawtooth Software Research Paper*, Sequim: Sawtooth Software.
- Kamakura, W., & Ozer, M. (2000). A Multi-Trait Multi-Method Validity Test of Partworth Estimates. In A. Gustafsson, A. Herrmann, & F. Huber (Eds.), *Conjoint Measurement – Methods and Applications*, Berlin: Springer, 251–277.
- Keeney, R. L., & Gregory, R. S. (2005). Selecting Attributes to Measure Achievement of Objectives. *Operations Research*, 53(1), 1–11.
- Krishnamurthi, L. (1988). Conjoint Models of Family Decision Making. *International Journal of Research in Marketing*, 5(3), 185–198.
- Krosnick, J. A., & Alwin, D. F. (1988). A Test of the Form-Resistant Correlation Hypothesis: Ratings, Rankings, and the Measurement of Values. *Public Opinion Quarterly*, 52(4), 526–538.
- Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–94.
- Lefkoff-Hagius, R., & Mason, C. H. (1993). Characteristic, Beneficial, and Image Attributes in Consumer Judgment of Similarity and Preference. *Journal of Consumer Research*, 20(1), 100–110.
- León, O. G. (1997). On the Death of SMART and the Birth of GRAPA. *Organizational Behavior and Human Decision Processes*, 71(3), 249–262.
- Lilien, G. L., Rangaswamy, A., & DeBruyn, A. (2013). *Principles of Marketing Engineering*, 2<sup>nd</sup> Ed., Pennsylvania: DecisionPro.
- Lines, R., & Denstadli, J. M. (2004). Information Overload in Conjoint Experiments. *International Journal of Market Research*, 46(3), 297–310.
- Loosschilder, G. H., Rosbergen, E., Vriens, M., & Wittink, D. R. (1995). Pictorial Stimuli in Conjoint Analysis – To Support Product Styling Decisions. *Journal of Market Research Society*, 37(1), 17–34.
- Louviere, J. J. (1988). *Analyzing Decision Making – Metric Conjoint Analysis*, Iowa: Sage Publishing.
- Louviere, J. J. (1991). Best-Worst Scaling: A Model for the Largest Difference Judgments. *Working Paper*, Alberta: University of Alberta.
- Louviere, J. J., Eagle, T. C., & Cohen, S. H. (2005). Conjoint Analysis: Methods, Myths and Much More. *CenSoC Working Paper*.
- Louviere, J. J., & Timmermans, H. J. (1992). Testing the External Validity of Hierarchical Conjoint Analysis Models of Recreational Destination Choice. *Leisure Sciences*, 14(3), 179–194.
- Louviere, J. J., & Woodworth, G. (1983). Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. *Journal of Marketing Research*, 20(4), 350–367.
- Luce, R. D. (1959). *Individual Choice Behaviours: A Theoretical Analysis*, New York: J. Wiley.
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. *Journal of Mathematical Psychology*, 1(1), 1–27.
- Lund, D. B., Malhotra, N. K., & Smith, A. E. (1988). Field Validation Study of Conjoint Analysis Using Selected Mail Survey Response Rate Facilitators. *Journal of Business Research*, 16(4), 351–368.
- Markham, S. K., & Hyunjung, L. (2013). Product Development and Management Association's 2012 Comparative Performance Assessment Study. *Journal of Product Innovation Management*, 30(3), 408–429.
- McCullough, D. (2002). A User's Guide to Conjoint Analysis. *Marketing Research*, 14(2), 18–23.
- McDaniel, S. W., Verille, P., & Madden, C. S. (1985). The Threats to Marketing Research: An Empirical Reappraisal. *Journal of Marketing Research*, 22(1), 74–80.
- McFadden, D. L. (1976). Quantal Choice Analysis – A Survey. *Annals of Economic and Social Measurement*, 5(4), 363–390.
- McFadden, D. L. (2014). *In the Matter of Determination of Rates and Terms for Digital Performance in Sound Recordings and Ephemeral Recordings (WEB IV)*, Washington: Library of Congress. Docket No. 14-CRB-0001-WR, October 6.
- Mehta, R., Moore, W. L., & Pavia, T. M. (1992). An Examination of the Use of Unacceptable Levels in Conjoint Analysis. *Journal of Consumer Research*, 19(3), 470–476.
- Meißner, M., & Decker, R. (2010). Eye-Tracking Information Processing in Choice-based Conjoint Analysis. *International Journal of Market Research*, 52(5), 591–610.
- Meißner, M., Decker, R., & Scholz, S. W. (2011). An adaptive Algorithm for Pairwise Comparison-Based Preference Measurement. *Journal of Multi-Criteria Decision Analysis*, 17(5/6), 167–177.
- Meißner, M., Musalem, A., & Huber, J. (2016). Eye-Tracking Reveals a Process of Conjoint Choice that is Quick, Efficient and Largely Free from Contextual Biases. *Journal of Marketing Research*, 53(1), 1–17.
- Moore, W. L., & Holbrook, M. B. (1990). Conjoint Analysis on Objects with Environmentally Correlated Attributes: The Questionable Importance of Representative Design. *Journal of Consumer Research*, 16(4), 490–497.
- Natter, M., & Feurstein, M. (2002). Real World Performance of Choice-Based Conjoint Models. *European Journal of Operational Research*, 137(2), 448–458.
- Netzer, O., & Srinivasan, V. (2011). Adaptive Self-Explication of Multiattribute Preferences. *Journal of Marketing Research*, 48(1), 140–156.

- Netzer, O., Toubia, O., Bradlow, E., Dahan, E., Evgeniou, T., Feinberg, F., Feit, E. M., Hui, S. K., Johnson, J., Liechty, J. C., Orlin, J. B., & Rao, V. (2008). Beyond Conjoint Analysis: Advances in Preference Measurement. *Marketing Letters*, 19(3/4), 337–354.
- Oppewal, H., Louviere, J. J., & Timmermans, H. J. (1994). Modeling Hierarchical Conjoint Processes with integrated Choice Experiments. *Journal of Marketing Research*, 31(1), 92–105.
- Orme, B. K. (2013). *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, 3<sup>rd</sup> Ed., Madison: Research Publishers.
- Orme, B. K., Alpert, M. I., & Christensen, E. (1997). Assessing the Validity of Conjoint Analysis – Continued. *Sawtooth Software Research Paper*, Sequim: Sawtooth Software.
- Orme, B. K., & Baker, G. C. (2000). Comparing Hierarchical Bayes Draws and Randomized First Choice for Conjoint Simulations. *Sawtooth Software Working Paper*, Sequim: Sawtooth Software.
- Parker, B. R., & Srinivasan, V. (1976). A Consumer Preference Approach to the Planning of Rural Primary Health-Care Facilities. *Operations Research*, 24(5), 991–1025.
- Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring Constructed Preferences: Toward a Building Code. *Journal of Risk and Uncertainty*, 19(1), 243–270.
- Rao, V. R. (2014a). Conjoint Analysis. In R. S. Winer, & S. A. Neslin (Eds.), *The History of Marketing Science*, Denvers: World Scientific Publishing, 47–76.
- Rao, V. R. (2014b). *Applied Conjoint Analysis*, Heidelberg: Springer.
- Rex Yuxing, D., Ye, H., & Damangir, S. (2015). Leveraging Trends in Online Searches for Product Features in Market Response Modeling. *Journal of Marketing*, 79(1), 29–43.
- Robinson, P. J. (1980). Applications of Conjoint Analysis to Pricing Problems. In D. R. Wittink (Ed.), *Market Measurement and Analysis: Proceedings of ORSA/TIMS Special Interest Conference*, Cambridge: Marketing Science Institute, 183–205.
- Sawtooth Software. (2003). *Advanced Simulation Module (ASM) for Product Optimization v1.4 Technical Paper*. Retrieved May 9, 2017, from <https://www.sawtoothsoftware.com/download/techpap/asmtech.pdf>.
- Sawtooth Software. (2006). *External Effect Adjustments in Conjoint Analysis*. Retrieved May 9, 2017, from <https://www.sawtoothsoftware.com/download/techpap/externaleffects.pdf>.
- Sawtooth Software. (2007). *The ACA/Web v6.0 Technical Paper*. Retrieved May 9, 2017, from <https://www.sawtoothsoftware.com/support/technical-papers/aca-related-papers/aca-technical-paper-2007>.
- Sawtooth Software. (2009). *Scale Factor and Conjoint Results*. Retrieved May 9, 2017, from <http://www.sawtoothsoftware.com/about-us/news-and-events/sawtooth-solutions/ss30-cb/1092-scale-factor-and-conjoint-results>.
- Sawtooth Software. (2013a). *The MaxDiff System Technical Paper*. Retrieved May 9, 2017, from <https://www.sawtoothsoftware.com/download/techpap/maxdifftech.pdf>.
- Sawtooth Software. (2013b). *The CBC System for Choice-Based Conjoint Analysis Version 8*. Retrieved May 9, 2017, from <https://www.sawtoothsoftware.com/download/techpap/cbctech.pdf>.
- Sawtooth Software. (2016). *Results of 2016 Sawtooth Software User Survey*. Retrieved May 9, 2017, from <http://www.sawtoothsoftware.com/about-us/news-and-events/news/1693-results-of-2016-sawtooth-software-user-survey>.
- Schlag, N. (2008). *Validierung der Conjoint-Analyse zur Prognose von Preisreaktionen mithilfe realer Zahlungsbereitschaften*, Lohmar: EUL Verlag.
- Schlereth, C., Eckert, C., Schaaf, R., & Skiera, B. (2014). Measurement of Preferences with Self-Explicated Approaches: A Classification and Merge of Trade-Off- and Non-Trade-Off-Based Evaluation Types. *European Journal of Operational Research*, 238(1), 185–198.
- Schlereth, C., & Schulz, F. (2014). Schnelle und einfache Messung von Bedeutungsgewichten mit der Restricted-Click-Stream Analyse: Ein Vergleich mit etablierten Präferenzmessmethoden. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung (ZfbF)*, 66(8), 630–657.
- Schneider, J., & Hall, J. (2011). Why most Product Launches Fail. *Harvard Business Review*, 89(4), 21–23.
- Scholl, A., Manthey, L., Helm, R., & Steiner, M. (2005). Solving Multiattribute Design Problems with Analytic Hierarchy Process and Conjoint Analysis: An Empirical Comparison. *European Journal of Operational Research*, 164(3), 760–777.
- Scholz, S. W., Meissner, M., & Decker, R. (2010). Measuring Consumer Preferences for Complex Products: A Compositional Approach Based on Paired Comparisons. *Journal of Marketing Research*, 47(4), 685–698.
- Smead, R. J., Wilcox, J. B., & Wilkes, R. E. (1981). How Valid are Product Descriptions and Protocols in Choice Experiments?. *Journal of Consumer Research*, 8(1), 37–42.
- Spiegler, R. (2011). *Bounded Rationality and Industrial Organization*, Oxford: Oxford University Press.
- Srinivasan, V. S. (1988). A Conjunctive-Compensatory Approach to the Self-explication of Multiattributed Preferences. *Decision Sciences*, 19(2), 295–305.
- Srinivasan, V. S., & de Ma Carty, P. (1999). Predictive Validation of Multiattribute Choice Models. *Marketing Research*, 11(4), 28–34.
- Srinivasan, V. S., & Park, C. S. (1997). Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement. *Journal of Marketing Research*, 34(2), 286–291.
- Stadie, E. (1998). *Medial gestützte Limit Conjoint-Analyse als Innovationstest für technologische Basisinnovationen – Eine explorative Analyse*, Münster: LIT Verlag.
- Steenkamp, J. E., & Van Trijp, H. (1997). Attribute Elicitation in Marketing Research: A Comparison of Three Procedures. *Marketing Letters*, 8(2), 153–165.
- Steenkamp, J. E., & Wittink, D. R. (1994). The Metric Quality of Full-Profile Judgements and the Number-of-Attribute-Levels Effect in Conjoint Analysis. *International Journal of Research in Marketing*, 11(3), 275–286.
- Steiner, M. (2007). *Nachfrageorientierte Präferenzmessung: Bestimmung zielgruppenspezifischer Eigenschaftssets auf Basis von Kundenbedürfnissen*, Wiesbaden: DUV Verlag.
- Steiner, M., Helm, R., & Hüttl-Maak, V. (2016). A Customer-based Approach for Selecting Attributes and Levels for Preference Measurement. *International Journal of Product Development*, 21(4), 233–266.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge – Improving Decisions about Health, Wealth and Happiness*, London: Penguin Books.
- Toubia, O., Simester, D. I., Hauser, J. R., & Dahan, E. (2003). Fast Polyhedral Adaptive Conjoint Estimation. *Marketing Science*, 22(3), 273–303.
- Tscheulin, D. K. (1991). Ein empirischer Vergleich der Eignung von Conjoint-Analyse und „analytic hierarchy process“ (AHP) zur Neuproduktplanung. *Zeitschrift für Betriebswirtschaftslehre*, 61(1), 1267–1280.
- Urban, G. L., & Hauser, J. R. (1980). *Design and Marketing of New Products*, New Jersey: Prentice-Hall.
- Urban, G. L., Weinberg, B. D., & Hauser, J. R. (1996). Premarket Forecasting of Really-New Products. *Journal of Marketing*, 60(1), 47–60.
- Voeth, M., & Hahn, C. (1998). Limit Conjoint-Analyse. *Marketing ZFP*, 20(2), 119–132.
- Vriens, M., Wedel, M., & Wilms, T. (1996). Metric Conjoint Segmentation Methods: A Monte Carlo Comparison. *Journal of Marketing Research*, 33(1), 73–85.
- Weiber, R., & Mühlhaus, D. (2009). Auswahl von Eigenschaften und Ausprägungen bei der Conjointanalyse. In D. Baier, & M. Brusch (Eds.), *Conjointanalyse*, Berlin Heidelberg: Springer, 43–58.

- Willkie, W. L., & Pessemier, E. A. (1973). Issues in Marketing's Use of Multi-Attribute Attitude Models. *Journal of Marketing Research*, 10(4), 428–441.
- Wittink, D. R., & Cattin, P. (1989). Commercial Use of Conjoint Analysis: An Update. *Journal of Marketing*, 53(3), 91–96.
- Wittink, D. R., Krishnamurthi, L., & Nutter, J. B. (1982). Comparing Derived Importance Weights across Attributes. *Journal of Consumer Research*, 8(4), 471–474.
- Wittink, D. R., Krishnamurthi, L., & Reibstein, D. J. (1989). The Effect of Differences in the Number of Attribute Levels on Conjoint Results. *Marketing Letters*, 1(2), 113–123.
- Wong, J. K., & Sheth, J. N. (1985). Explaining Intention-Behavior Discrepancy-A Paradigm. *Advances in Consumer Research*, 12(1), 378–384.
- Wright, P., & Kriewall, M. A. (1980). State-of-Mind Effects on the Accuracy with Which Utility Functions Predict Marketplace Choice. *Journal of Marketing Research*, 17(3), 277–293.
- Wright, P., & Weitz, B. (1977). Time Horizon Effects on Product Evaluation Strategies. *Journal of Marketing Research*, 14(4), 429–443.

## Keywords

Preference Measurement, Conjoint Analysis, Discrete Choice, Attribute Set, Market Simulation.

## MARKETING

ZFP – Journal of Research and Management

**Editor-in-Chief:** Prof. Dr. Bernhard Swoboda, Chair for Marketing and Retailing, University of Trier, Universitätsring 15, D-54296 Trier, Phone: +49 651 201 3050, Fax: +49 651 201 4165, Email: editor@marketing-zfp.de

**Senior Editors:** Prof. Dr. Heribert Gierl, University of Augsburg, Prof. Dr. Andrea Gröppel-Klein, Saarland University, Prof. Dr. Lutz Hildebrandt, Humboldt-University of Berlin, Prof. Dr. Hans Mühlbacher, International University of Monaco, Prof. Dr. Henrik Sattler, University of Hamburg, Prof. Dr. Udo Wagner, University of Vienna.

**Manuscripts:** We ask all authors who would like to submit a paper to send this paper to the editor-in-chief Bernhard Swoboda via mail: editor@marketing-zfp.de. Neither the publisher nor the editors assume any liability for unsolicited manuscripts. Unsolicited manuscripts will only be returned if accompanied by return postage. The acceptance of a contribution has to be in writing.

**Copyright:** Upon acceptance for publication the author transfers to the C.H.BECK the exclusive copyright of his

or her contribution for the duration of the copyright as laid down by law. The copyright covers the exclusive right and licence to reproduce, publish, distribute and archive the article in all forms and media of expression now known or developed in the future, including reprints, translations, photographic reproductions, microform, electronic form (offline and online) or any other reproduction of similar nature. The author's second window right after the expiry of 12 months after first publication, as laid down in article 38(4) German Copyright Law, remains unaffected. All articles published in this journal are protected by copyright law. Without first obtaining permission from the publisher, no material published in this journal may be reproduced, distributed, performed or displayed publicly, or made accessible or stored in electronic databases or reproduced, distributed or utilized electronically, outside the narrow limitations of copyright law.

**Publisher:** C.H.BECK oHG, Wilhelmstr. 9, 80801 München; postal address: P.O.Box 40 03 40, 80703 München; phone: +49 89 38189 0; fax: +49 89 38189 398, Bank account: Postbank München IBAN: DE82 7001 0080 0006 2298 02, BIC: PBNKDEFFXXX.

The publisher is a *offene Handelsgesellschaft* under German law with Dr. Hans Dieter Beck and Dr. h.c. Wolfgang Beck as partners.

**Subscription:** An annual subscription to the journal comprises four issues.

**Subscription rates 2018:** € 225 (VAT incl.) annual subscription rate, campus licence € 349 (VAT incl.). Single Issue: € 63,50 (VAT incl.), shipping charges have to be added to the rates. Subscription and rate include print issue and a licence for the online archive. The components cannot be cancelled separately. Complaints about copies not received must be lodged within 6 weeks starting at the end of the quarter.

**Subscription service:** Please order with either the publisher or any book shop.

**CustomerServiceCenter:** Phone: +49 89 38189 750, Fax: +49 89 38189 358, E-Mail: kundenservice@beck.de

**Cancellation:** The subscription may be cancelled in writing 6 weeks before the end of a calendar year.

**Citation:** Marketing ZFP – Journal of Research and Management, number of volume(number of issue), year, page.

**Typesetting:** FotoSatz Pfeifer GmbH, 82152 Krailling.

**Printing:** Kessler Druck und Medien GmbH & Co. KG, Michael-Schäffer-Straße 1, 86399 Bobingen.