

Compensatory versus noncompensatory models for predicting consumer preferences

Anja Dieckmann*
Basic Research
GfK Association

Katrin Dippold
Department of Marketing
University of Regensburg

Holger Dietrich
Basic Research
GfK Association

Abstract

Standard preference models in consumer research assume that people weigh and add all attributes of the available options to derive a decision, while there is growing evidence for the use of simplifying heuristics. Recently, a greedoid algorithm has been developed (Yee, Dahan, Hauser & Orlin, 2007; Kohli & Jedidi, 2007) to model lexicographic heuristics from preference data. We compare predictive accuracies of the greedoid approach and standard conjoint analysis in an online study with a rating and a ranking task. The lexicographic model derived from the greedoid algorithm was better at predicting ranking compared to rating data, but overall, it achieved lower predictive accuracy for hold-out data than the compensatory model estimated by conjoint analysis. However, a considerable minority of participants was better predicted by lexicographic strategies. We conclude that the new algorithm will not replace standard tools for analyzing preferences, but can boost the study of situational and individual differences in preferential choice processes.

Keywords: Conjoint analysis, greedoid algorithm, choice modeling, lexicographic heuristics, noncompensatory heuristics, consumer choice, consumer preferences.

1 Introduction

How do customers choose from the abundance of products in modern retail outlets? How many attributes do they consider, and how do they process them to form a preference? These questions are of theoretical as well as practical interest. Gaining insights into the processes people follow while making purchase decisions will lead to better informed decision theories. At the same time, marketers are interested in more realistic decision models for predicting market shares and for optimizing marketing actions, for example, by adapting products and advertising materials to consumers' choice processes.

In consumer research, decision models based on the idea of utility maximization predominate to date, as expressed in the prevalent use of weighted additive models derived from conjoint analysis to capture preferences (Elrod, Johnson & White, 2004). At the same time, judgment and decision making researchers propose alternative decision heuristics that are supposed to provide psychologically more valid accounts of human decision making,

and gather evidence for their use (e.g., Bröder & Schiffer, 2003a; Gigerenzer, Todd & the ABC Research Group, 1999; Newell & Shanks, 2003). Recently, the field of judgment and decision making has been equipped with a new tool, a greedoid algorithm to deduce lexicographic decision processes from preference data, developed independently by Yee et al. (2007) and Kohli and Jedidi (2007).

We aim to bring together these two lines of research by comparing the predictive performance of lexicographic decision processes deduced by the new greedoid algorithm to weighted additive models estimated by full profile regression-based conjoint analysis as a standard tool in consumer research. We derive hypotheses from the theoretical framework of adaptive decision making about when which approach should be the better-suited tool, and test them in an empirical study.

1.1 The standard approach to model preferences in consumer research

Conjoint analysis is based on seminal work from Luce and Tukey (1964). Green developed the method further and adapted it to marketing and product-development problems (e.g., Green & Rao, 1971; Green & Wind, 1975). Today, conjoint analysis is regarded as the most prevalent tool to measure consumer preferences (Wittink

*We thank Jörg Rieskamp, Jonathan Baron and two anonymous reviewers for helpful comments on earlier versions of this manuscript, and those who kindly volunteered to participate in the study. Address: Anja Dieckmann, Basic Research, GfK Association, Nordwestring 101, 90319 Nürnberg, Germany. E-mail: anja.dieckmann@gfk.com. The GfK Association is the non-profit organization of the GfK Group. Its activities include noncommercial, fundamental research in close cooperation with scientific institutions.

& Cattin, 1989; Wittink, Vriens & Burhenne, 1994). In a survey among market research institutes, 65% of the institutes indicated having used conjoint analysis within the last 12 months, and growing usage frequency was forecasted (Hartmann & Sattler, 2002). Conjoint analysis is used to analyze how different features of products contribute to consumers' preferences for these products. This is accomplished by decomposing the preference for the whole product into partitions assigned to the product's constituent features. The established way to collect preference data is the full-profile method. Product profiles consisting of all relevant product features are presented to respondents. These profiles are evaluated either by rating or ranking or by discrete choice (i.e., buy or non-buy) decisions¹.

The assumption behind the decompositional nature of conjoint analysis is that people weigh and add all available pieces of product information, thus deriving a global utility value for each option as the sum of partworth utilities. Options with higher utility are preferred — either deterministically or probabilistically — over options with lower utility. Clearly, this assumption rests on traditional conceptions of what constitutes rational decision making. Homo economicus is assumed to carefully consider all pieces of information and to integrate them into some common currency, such as expected utility, following a complex weighting scheme.

For rating- and ranking-based conjoint methods,² the basic weighted additive model (WADD) can be stated as follows:

$$r_k = \beta_0 + \sum_{j=1}^J \sum_{m=1}^M \beta_{jm} \cdot x_{jm} + \varepsilon_k \quad (1)$$

with

r_k = response for option k ;

β_{jm} = partworth utility of level m of attribute j ;

$x_{jm} = 1$ if option k has level m on attribute j ;

else $x_{jm} = 0$; and

ε_k = error term for response for option k .

The partworth utilities are estimated, usually by applying multiple regression, such that the sum of squares between empirically observed responses r_k (ratings or rankings) and estimated responses \hat{r}_i is minimal.

¹More elaborate methods that ask for self-explicated attribute preferences such as ACA (Johnson, 1987; Green, Krieger & Agarwal, 1991) and HILCA (Wildner, Dietrich & Hölscher, 2007) have been developed over the years, but the full-profile method still remains the basic principle.

²Basically, one can distinguish between rating- and ranking-based conjoint using regression analysis as estimation method and choice-based conjoint with multinomial logit estimation methods (e.g., Green & Srinivasan, 1978; Elrod, Louviere & Davey, 1992).

1.2 Simple decision heuristics

The traditional view of rational decision making as utility maximization has been challenged in the judgment and decision making literature. Many authors propose alternative accounts of human decision making processes and argue that people are equipped with a repertoire of decision strategies from which to select depending on the decision context (e.g., Beach & Mitchell, 1978; Einhorn, 1971; Gigerenzer, Todd, & the ABC Research Group, 1999; Payne, 1976, 1982; Payne, Bettman, & Johnson, 1988, 1993; Rieskamp & Otto, 2006; Svenson, 1979). According to Payne et al. (1988, 1993), decision makers choose strategies adaptively in response to different task demands, and often apply simplified shortcuts — heuristics — that allow fast decisions with acceptable losses in accuracy. Moreover, simple heuristics are often more or at least equally accurate in predicting new data compared to more complex strategies (e.g., Czerlinski, Gigerenzer & Goldstein, 1999; Gigerenzer, Czerlinski & Martignon, 1999). The explanation is that simple heuristics are more robust, extracting only the most important and reliable information from the data, while complex strategies that weigh all pieces of evidence extract much noise, resulting in large accuracy losses when making predictions for new data — a phenomenon called overfitting (Pitt & Myung, 2002).

Lexicographic strategies are a prominent category of simple heuristics. A well-known example is Take The Best (TTB; Gigerenzer & Goldstein, 1996), for inferring which of two alternatives has a higher criterion value by searching sequentially through cues in the order of their validity until one discriminating cue is found. The alternative with the positive cue value is selected. TTB is “noncompensatory” because a cue cannot be outweighed by any combination of less valid cues, in contrast to “compensatory” strategies, which integrate cue values (e.g., the WADD model). Applied to a consumer choice context, a lexicographic heuristic would prefer a product that is superior to another product on the most important aspect³ for which the two options have different values, regardless of the aspects that follow in the aspect hierarchy.

³Following Yee et al.'s (2007) terminology, the levels of attributes (e.g., color, size) are called aspects (e.g., red, green, blue, small, big). The TTB heuristic is formulated for dichotomous attributes, or cues (e.g., feature present or not). Typical profiles used in conjoint analysis are often characterized by multi-level attributes (e.g., different price levels). To apply TTB, these levels can be transformed into dichotomous aspects. For instance, a three-level attribute is translated into three dichotomous aspects (e.g., low price present or not, intermediate price present or not, high price present or not).

1.3 Inferring lexicographic decision processes

Choices can be forecast by linear compensatory models even though the underlying decision process has a different structure (e.g., Einhorn, Kleinmuntz & Kleinmuntz, 1979). A WADD model, for example, can theoretically reproduce a non-compensatory decision process if, in the ordered set of weights, each weight is larger than the sum of all weights to come (e.g., aspect weights of 2^{1-n} with $n = 1, \dots, N$ and $N =$ number of aspects; Martignon & Hoffrage, 1999, 2002). Despite its flexibility in assigning weights, however, Yee et al. (2007) showed in Monte Carlo simulations that WADD models fall short of capturing the non-compensatory preference structure and are outperformed by lexicographic models when the choice is made in a perfectly non-compensatory fashion. Moreover, the goal is not only to generate high prediction performance but also insight into the process steps of decision making. Although conclusions from consumer self-reports and process tracing studies are limited (see below), several such studies suggest that only a minority of participants use a weighted additive rule, thus questioning the universal application of conjoint analysis (Denstadli & Lines, 2007; Ford, Schmitt, Schechtman, Hults & Doherty, 1989).

Most users of conjoint models are well aware of their status as “as if” models (Dawkins, 1976), and do not claim to describe the underlying cognitive process but only aim to predict the outcome. Consequently, many researchers call for psychologically more informed models (e.g., Bradlow, 2005; Louviere, Eagle & Cohen, 2005). However, these rightful claims suffer from the lack of data analysis tools that estimate heuristics based on preference data. Self-reports (e.g., Denstadli & Lines, 2007) are an obvious tool for tracking decision strategies but have questionable validity (Nisbett & Wilson, 1977). A more widely accepted way to deduce heuristics from people’s responses are process tracing techniques, such as eye tracking and mouse tracking (e.g., Payne et al., 1988, 1993; see Ford et al., 1989, for a review), or response time analyses (e.g., Bröder & Gaissmeier, 2007). However, these techniques are very expensive for examining large samples, as often required in consumer research. Moreover, data collection methods such as information boards tend to interfere with the heuristics applied and might induce a certain kind of processing (Billings & Marcus, 1983). Finally, it is unclear how process measures can be integrated into mathematical prediction models, as the same processing steps can be indicative of several strategies (Svenson, 1979).

In inference problems where the task is to pick the correct option according to some objective external criterion, such as inferring which of two German cities is

larger, heuristics can be deduced by using datasets with known structure — in this case, a data set of German cities including their description in terms of features such as existence of an exposition site or a soccer team in the major league (Gigerenzer & Goldstein, 1996). Based on the data set, one can compute the predictive validity of the different features, or cues, and thus derive cue weights. This way, competing inference strategies that process these cues in compensatory or noncompensatory fashions, including their predictions, can be specified a priori. These predictions can then be compared to the observed inferences that participants have made and the strategy that predicts most of these responses can be determined (see, e.g., Bröder, 2000; Bröder & Schiffer, 2003b; Rieskamp & Hoffrage, 1999, 2008).

In preferential choice, by contrast, the individual attribute weighting, or ordering structure, does not follow some objective outside criterion but depends on subjective preference and has to be deduced in addition to the decision strategy people use. Standard conjoint analyses estimate individual weighting structure assuming a weighted additive model, as laid out above. Approaches assuming alternative models are rare. Gilbride and Allenby (2004) model choices in a two-stage-process and allow for compensatory and noncompensatory screening rules in the first stage. Elrod et al. (2004) suggest a hybrid model that integrates compensatory decision strategies with noncompensatory conjunctive and disjunctive heuristics. However, these approaches are basically modifications of the standard WADD model, allowing for noncompensatory weighting and conjunctions and disjunctions of aspects. This model, however, is not a valid representation of human decision making; its flexibility not only makes psychologically implausible computational demands, but also technically requires huge processing capacity. In contrast, the greedoid algorithm we focus on is intriguingly simple. It incorporates the principles of lexicography and noncompensatoriness rather than just adapting weighting schemes to imitate the output of lexicographic heuristics.

Yee et al. (2007)⁴ developed the greedoid algorithm for deducing lexicographic processes from observed preference data, applicable to rating, ranking and choice alike. The algorithm rests on the assumption that the aspects of different options are processed lexicographically. It discloses the aspect sorting order that best replicates the observed (partial) preference hierarchy of options. Generally, the algorithm can be used to estimate various lexicographic heuristics. By introducing specific restrictions, aspects can be treated as acceptance or elimination criteria to model the well-known elimination-

⁴We follow Yee et al.’s (2007) formulation of the algorithm but note that Kohli and Jedidi (2007) have independently developed a greedy algorithm to estimate lexicographic processes.

by-aspects heuristic (Tversky, 1972). For our purposes, we will implement the most flexible lexicographic-by-aspects (LBA) process that allows aspects from different attributes to be freely ranked as acceptance or elimination criteria.

The lexicographic-by-aspects process can be illustrated by a simple example. Given a choice between holiday options differing in travel location, with the three aspects Spain, Italy, and France, and means of transport, with the two aspects plane and car, a person may express the following preference order:

- (1) Spain by plane;
- (2) Spain by car;
- (3) France by plane;
- (4) Italy by plane;
- (5) France by car;
- (6) Italy by car.

The person's preferences in terms of location and transport are quite obvious. She prefers Spain regardless of how to get there. Means of transport becomes determining when considering other countries, with a preference for flying. A restrictive lexicographic-by-attributes process could not predict this preference order without mistake. Using country as first sorting criteria, with the aspect order [Spain — France — Italy], would produce one mistake (i.e., wrong order for options 4 and 5), and sorting by means of transport, with the aspect order [plane — car], would produce two mistakes (i.e., wrong order for options 2 and 3 as well as 2 and 4). A lexicographic-by-aspects process, in contrast, by allowing aspects from different attributes to be ordered after each other, can predict the observed preference ranking perfectly. This way, the sorting order [Spain — plane — France] becomes possible, reproducing the observed order without mistakes.⁵ This is the result that would be produced by the lexicographic-by-aspects implementation of the greedoid algorithm.

The algorithm is an instance of dynamic programming with a forward recursive structure. As goodness-of-fit criterion, a violated-pairs metric is used, counting the number of pairs of options that are ranked inconsistently in the observed and the predicted preference order. Basically, the algorithm creates optimal aspect orders for sorting alternatives — for example, various product profiles — by proceeding step-by-step from small sets of aspects to larger ones until the alternatives are completely sorted. First, the algorithm determines the inconsistencies that would be produced if the alternatives were ordered by one single aspect. This is repeated for each aspect. Then, starting from a set size of $n = 2$ aspects, the algorithm determines the best last aspect within each set, before mov-

⁵Note that with dichotomous attributes only, lexicographic-by-aspects processes lead to the same result as lexicographic-by-attributes processes.

ing forward to the next larger set size, and so forth until the set size comprises enough aspects to rank all options. Maximally, all 2^N possible sets of N aspects are created and searched through (only if the set of profiles cannot be fully sorted by fewer aspects than N). This sequential procedure exploits the fact that the number of inconsistencies induced by adding an aspect to an existing aspect order depends only on the given aspects within the set and is independent from the order of those aspects. Compared to exhaustive enumeration of all $N!$ possible aspect orders, dimensionality is reduced to the number of possible unordered subsets, 2^N , decreasing running time by a factor of the order of 10^9 (Yee et al., 2007).

In an empirical test of the new algorithm, participants had to indicate their ordinal preferences for 32 Smart-Phones (Yee et al., 2007). The products were described by 7 attributes; 3 attributes had 4 levels, 4 attributes had 2 levels, resulting in 20 aspects.⁶ The Greedoid approach was compared to two methods based on the weigh-and-add assumption: hierarchical Bayes ranked logit (e.g., Rossi & Allenby, 2003) and linear programming (Srinivasan, 1998).⁷ Overall, the greedoid approach proved superior to both benchmark models in predicting hold-out data,⁸ and thus seems to represent a viable alternative to standard estimation models. We aim to find out more about the conditions under which this methodology can be fruitfully applied, as well as about its limitations.

1.4 External and internal factors affecting strategy selection

According to the adaptive-strategy-selection view (Payne et al., 1993), people choose different strategies depending on characteristics of the decision task. For instance, simple lexicographic heuristics predict decisions well when

⁶Splitting dichotomous attributes into two aspects results in redundant information — if one aspect is present the other aspect has to be absent and vice versa. But as we do not know which of the two aspects is preferred, we include both aspects in the input for the algorithm. Note that Yee et al. (2007) suggest a more frugal way of treating dichotomous attributes as one aspect by allowing for flipping within the algorithm's code. The result will be the same as including two (redundant) aspects.

⁷Linear Programming is a non-metric method for ranking data. It is based on linear optimization. The criterion to be minimized is the sum of metric corrections necessary to force the estimated stimuli values into the observed rank order. Hierarchical Bayes logit is a powerful but complex way to estimate individual utilities. It combines population level assumptions on utility distributions with individual level choices and estimates individual utilities in an iterative process. Hierarchical means that distribution assumptions are made on a global as well as on an individual level.

⁸The term hold-out refers to data that are withheld from parameter estimation. It is used for testing how well the model derived from the calibration data, which are used to estimate parameters, predicts new data. One aim of using hold-outs to determine predictive accuracy is to avoid favoring models that overfit the calibration data, leading to high fit values but low predictive accuracy for new data.

retrieval of information on the available options is associated with costs, such as having to pay for information acquisition (Bröder, 2000; Newell & Shanks, 2003; Newell, Weston & Shanks, 2003; Rieskamp & Otto, 2006), deciding under time pressure (Rieskamp & Hoffrage, 1999, 2008), or having to retrieve information from memory (Bröder & Schiffer, 2003a). The exact same studies, however, show that, when participants are given more time and can explore information for free, weighted-additive models usually outperform lexicographic strategies in predicting people's decision making. Information integration seems to be a default applied by most people when faced with new decision tasks unless circumstances become unfavorable for extensive information search (Dieckmann & Rieskamp, 2007).

Additionally, the mode in which options are presented as well as the required response mode affect strategy selection (for a review, see Payne, 1982). Simultaneous display of options facilitates attribute-wise comparisons between alternatives.⁹ In contrast, sequential presentation promotes alternative-wise, and thus more holistic, additive processing, as attribute-wise comparisons between options become difficult and would require the retrieval of previously seen options from memory or the application of internal comparison standards per attribute (Dhar, 1996; Lindsay & Wells, 1985; Nowlis & Simonson, 1997; Schmalhofer & Gertzen, 1986; Tversky, 1969). Regarding response mode effects, Westenberg and Koele (1992) propose that the more differentiated the required response, the more differentiated and compensatory the evaluation of the alternatives. Following this proposition, ranking — which additionally is associated with simultaneous presentation of options — requires ordinal comparisons between options and is thus supposed to foster lexicographic processing, while rating requires evaluating one option at a time on a metric scale, which should promote compensatory processing. Indeed, there is empirical evidence that people use strategies that directly compare alternatives, such as elimination-by-aspects, more often in choice than in rating (e.g., Billings & Scherer, 1988; Schkade & Johnson, 1989). Note that ranking tasks are often posed as repeated choice tasks, requiring participants to sequentially choose the most preferred option from a set that gets smaller until only the least preferred option remains. For such ranking tasks, we therefore expect similar differences in comparison to rating tasks as for choice tasks, and anticipate higher predictive accuracy of a lexicographic model relative to a compensatory model.

⁹Lexicographic processes are characterized by attribute-wise comparisons; however, there are also compensatory strategies with attribute-wise search, such as the Majority of Confirming Dimensions rule (Russo & Doshier, 1983).

This prediction agrees with another result reported by Yee et al. (2007). Besides their own ranking data, they re-analyzed rating data by Lenk, DeSarbo, Green and Young (1996). Unlike in the ranking case, the Greedoid approach produced slightly lower predictive accuracy for hold-out data than a hierarchical Bayes ranked logit model. However, the two studies differed in several aspects, so the performance difference cannot be unambiguously attributed to the difference in respondents' preference elicitation task.

Among the internal factors affecting the selection of decision strategies are prior knowledge and expertise (Payne et al., 1993). Experts tend to apply more selective information processing than non-experts (e.g., Bettman & Park, 1980; see Shanteau, 1992, for an overview). Shanteau (1992) reports results demonstrating that experts are more able than non-experts to ignore irrelevant information. Ettenson, Shanteau, and Krogstad (1987) found that professional auditors weighted cues far more unequally, relying primarily on one cue, than students. Similarly, in a study on rating mobile phones, participants that reported having used a weighted additive strategy had the lowest scores on subjective and objective product category knowledge compared to other strategy users (Densstadli & Lines, 2007). In short, experts seem to be better able to prioritize attributes, thus possibly giving rise to a clear attribute hierarchy with noncompensatory alternative evaluation. In contrast, non-experts might be less sure about which attribute is most important and therefore apply a risk diffusion strategy by integrating different pieces of information.

To summarize, we compared two models of decision strategies — weighted additive and lexicographic — in terms of their predictive accuracy for ranking versus rating data. Our hypothesis was that, relative to compensatory strategies, lexicographic processes predict participants' preferences better in ranking than in rating tasks. Our second hypothesis was that, regardless of the required response, predictive accuracy of the lexicographic model is higher for experts than for non-experts, because experts are better able to prioritize attributes (Shanteau, 1992).

2 Method

To test our hypotheses, we selected skiing jackets as product category, which can be described by few attributes, thus allowing for acceptable questionnaire length and complexity. The product can be assumed to be relevant for many people in the targeted student population at a southern German university.

2.1 Participants

A sample of 142 respondents, 56% male, with an average age of 23.9 years, was recruited from homepages mainly frequented by business students at the University of Regensburg as well as via personal invitations during marketing classes and emails. For participation, everyone obtained the chance of winning one out of ten coupons worth 10 € for an online bookstore.

2.2 Procedure

Participants filled out a web-based questionnaire with separate sections for rating and ranking a set of skiing jackets. Each product was described by 6 features: price and waterproofness, each with 3 levels, as well as 4 dichotomous variables indicating the presence of an adjustable hood, ventilation zippers, a transparent ski pass pocket, and heat-sealed seams. These 6 features had been identified as most relevant for skiing jackets during exploratory interviews with skiers. The 96 possible profiles of skiing jackets were reduced to a 16-profile fractional factorial design (calibration set) that is balanced and orthogonal.¹⁰ Each respondent, in each task, was shown the 16 profiles plus 2 hold-outs. Respondents were not aware of this distinction, as the 16 calibration profiles were interspersed with the hold-outs; both were presented and evaluated in the same way.¹¹ For the ranking task, all 18 profiles were shown at once. The task was formulated as a sequential choice of the preferred product (“What is your favorite skiing jacket out of this selection of products?”). The chosen product was deleted from the set, and the selection process started all over again until only the least preferred product was left. During the rating task, one profile at a time was presented to respondents. They were asked to assign a value on a scale from 0 to 100 to each profile (“How much does this product conform to your ideal skiing jacket?”). Each participant saw a new random order of profiles; task order was randomized as well. Between these tasks, people completed a filler task in order to minimize the influence of the first task on the second one.¹² The conjoint tasks were enclosed by demographic questions and questions on expertise (e.g., “Are you a trained skiing instructor?”). The survey took approximately 20 minutes.

¹⁰The fractional design is produced by selecting the shortest possible plan from a library of prepared plans and applying a rule based procedure to adapt it to the given number of attributes and aspects (i.e., SPSS Orthoplan). Balanced means that each level of an attribute is shown equally often, and orthogonality avoids correlations between the shown levels of the features.

¹¹Usually, the distinction becomes evident only in data analysis: Calibration profiles are used for model fitting, while hold-outs are used to validate the predictions of the estimated model.

¹²The filler task consisted of items from the domain-specific risk-taking scale (DOSPERT; Johnson, Wilke & Weber, 2004). The results are reported elsewhere (Dippold, 2007).

2.3 Data analysis

As mentioned above, in applications of conjoint analysis in consumer research there usually is an a-priori definition of distinctive sets of calibration profiles used for model fitting and hold-out profiles used for evaluating predictive performance. The set of calibration profiles is designed to ensure sufficient informative data points per attribute aspect by paying attention to balance and orthogonality in aspect presentation across the different choice options. It could be argued, however, that the hold-outs might be peculiar in some way and thus lead to distorted estimates of predictive accuracy. We therefore decided to conduct a full leave-two-out cross-validation. That is, we fitted the model to all 153 possible sets of 16 profiles out of all 18 profiles, and in each run used the remaining two profiles as hold-outs for computing predictive accuracy. This necessarily involves slight violations of orthogonality in many of the 153 calibration sets. We think that these violations are acceptable, for the sake of generality and because there is no reason to expect the two tested models to be differentially affected by potential lack of information on some aspects.

Ranking and rating data were analyzed separately by the greedoid algorithm and by conjoint analysis. Ordinary least squares regression analysis was used to estimate individual-level conjoint models.¹³ The resulting partworth utilities were used to formulate individual WADD models for each cross-validation run for each participant (see Equation 1). For each pair, the option with higher total utility was — deterministically — predicted to be preferred over the option with lower total utility. The outcome of the greedoid algorithm was used to specify individual LBA processes for each cross-validation run for each participant to decide between all possible pair comparisons of options: Aspects were considered sequentially according to the individual aspect order the greedoid algorithm produced. As soon as one aspect discriminated between options, the comparison was stopped, the remaining aspects were ignored, and the option with the respective aspect was predicted to be preferred.

¹³Note that OLS regression is the standard estimation method for rating but not for ranking data where partworths are traditionally estimated by non-metric algorithms such as MONANOVA or LINMAP (Green & Srinivasan, 1978). The use of simpler metric methods instead of more complex non-metric ones has been subject of many empirical and simulation studies which indicate that metric and non-metric estimation procedures provide similar results (e.g., Carmone, Green & Jain, 1978; Wittink & Cattin, 1981), and that non-metric methods do not per se outperform OLS (Cattin & Bliemel, 1978). In fact, sometimes OLS beats non-metric analyses in terms of parameter precision (Mishra, Umesh & Stem, 1989) and predictive validity (Cattin & Wittink, 1977; Wulf, 2008) which can be attributed to the robustness of OLS. Besides, other more powerful (in terms of predictive accuracy) but also more complex estimation procedures exist than OLS (e.g., hierarchical Bayes). Thus, OLS regression is a conservative benchmark for the WADD estimation.

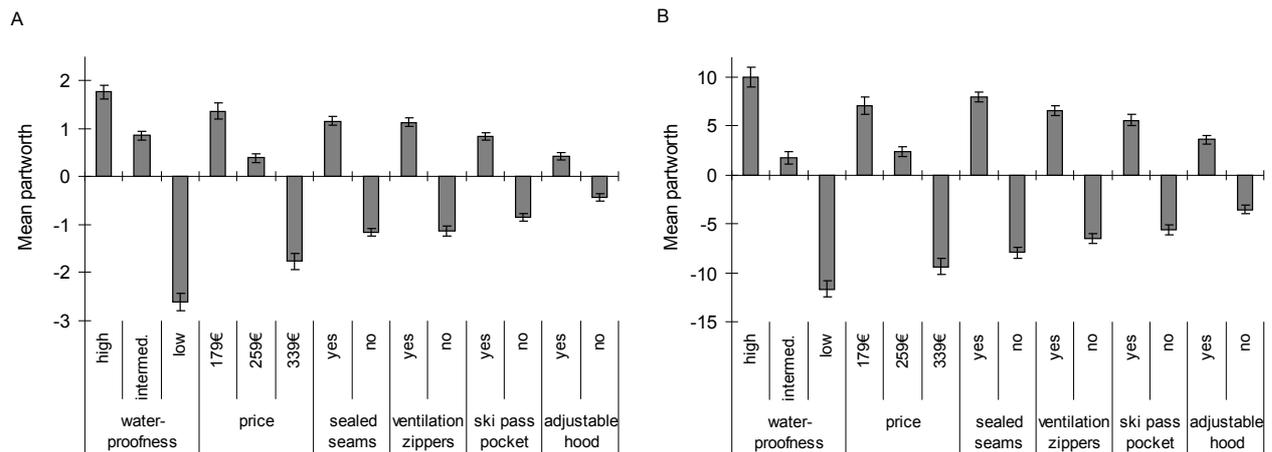


Figure 1: Mean partworths of aspects of the different attributes estimated by least squares regression analysis based on (A) ranking and (B) rating data; attributes ordered by decreasing importance (defined as difference between highest and lowest partworths of its aspects). Error bars represent standard errors.

The models' predictions were compared to empirical rankings or ratings, respectively.¹⁴ Each pair of products for which one model predicts the wrong option to be preferred was counted as one violated pair produced by that model.¹⁵ For each participant, results were averaged across the 153 cross-validation runs. The main focus was on the mean predictive accuracy for hold-outs, that is, pairs of options with at least one option not included in the data fitting process.

3 Results

3.1 Descriptive results

Summarized over all respondents, the aspect partworths resulting from conjoint analyses of the rating task were largely congruent with the partworths resulting from the ranking task (see Figure 1). For both data collection methods, two features (waterproofness and price) received a much higher weight than the four other attributes.¹⁶ These results were comparable to the aspect

orderings disclosed by the greedoid algorithm. Features of the same two attributes dominated decision making, and they did so in rating and ranking (see Figure 2).

3.2 Model fit

The WADD model showed better data fit than the LBA model. For ranking, WADD produced 7.9% violated pairs on average across cross-validation runs and participants ($SD = 6.4$), while LBA produced 10.3% ($SD = 6.0$). For rating, WADD produced 6.5% violated pairs on average ($SD = 5.4$), compared to 8.7% produced by LBA ($SD = 5.3$). Given the high flexibility of the weights of the WADD model that can be adjusted to accommodate highly similar to highly differentiated weighting schemes, this result came as no surprise. The crucial test was how the two models performed when predicting hold-out data.

3.3 Predictive accuracies

3.3.1 Ranking vs. rating

Mean predictive accuracies for hold-out data of the two decision models in terms of percentage of violated pairs, averaged across participants, are summarized in Table 1. Clearly, the WADD model is better than the LBA model at predicting the preferences for the hold-out profiles for both ranking and rating tasks. In line with these descriptive results, a repeated-measurement ANOVA of the dependent variable of individual-level predictive accuracy for hold-out data (in terms of percentage of violated pairs), with the two within-subject factors Task (rating vs. ranking) and Model (WADD vs. LBA), revealed a significant main effect of Model, $F(1,141) = 89.18$, $p < .001$.

¹⁴In the rating data set, participants could assign the same value to two or more options, leading to tied pairs. These ties were eliminated, so that only pairs with a clear preference order were used to evaluate performance of the different models.

¹⁵This violated-pairs metric is chosen because the greedoid algorithm makes only ordinal predictions, that is, its output can be used only to predict whether an alternative is preferred over another, but not how much more attractive it is compared to the other one. But note that the violated-pairs metric, despite its simplicity, is sensitive to error magnitude: The further an alternative is apart in the predicted order from its position in the observed rank order, the more violated pairs will be produced.

¹⁶This effect may be ascribed to the number-of-levels effect (e.g., Wittink, Huber, Zandan & Johnson, 1992), as these two features were the only ones with three instead of two levels.

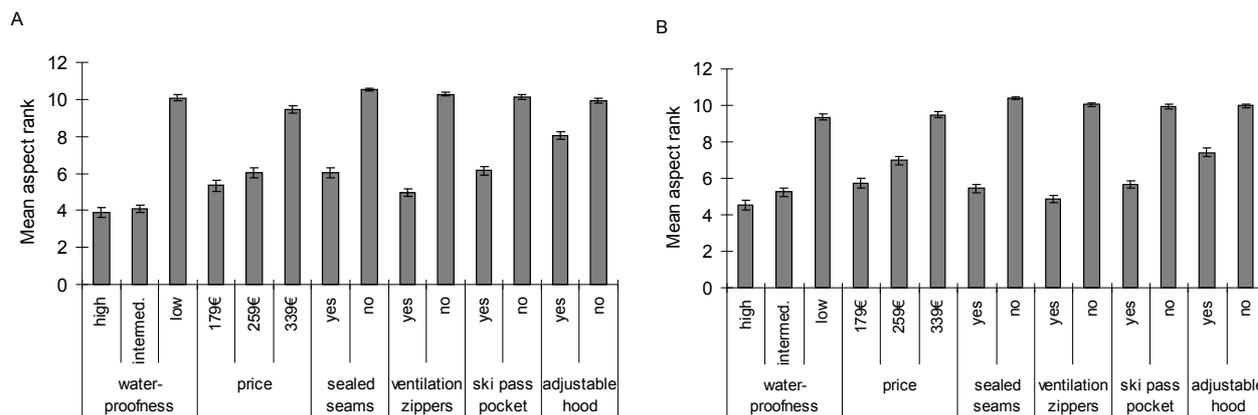


Figure 2: Mean ranks of aspects in the orders resulting from the greedoid algorithm applied to (A) ranking and (B) rating data. Error bars represent standard errors. Aspects that were not included in the aspect order derived from the greedoid algorithm received the mean rank of the remaining ranks (e.g., when the aspect order comprised 6 aspects, all not-included aspects were given rank 10.5, that is, the mean of ranks 7 to 14).

The factor Task did not show a main effect, $F(1,141) = 0.11, p = .746$, but there was a significant interaction of Task x Model, $F(1,141) = 11.60, p = .001$: While LBA produces fewer violated pairs for the ranking compared to the rating task, WADD performs slightly worse for the ranking than for the rating task (see Table 1; the interaction can also be seen within the groups of experts and non-experts in Figure 3). Post-hoc t -tests revealed that predictive accuracy of LBA is marginally higher for ranking than for rating, $t(141) = 1.41, p = .081$. Thus, there is some support for the hypothesis that LBA is better at predicting ranking compared to rating data.¹⁷

One could argue that the violated-pairs metric unfairly favors models that predict many ties, which are not counted as violated pairs. However, the percentages of pair comparisons for which ties are predicted are below 1% for all models, and LBA predicted more ties (0.7% on average for ranking, 0.6% for rating) than WADD (0.1% for ranking, 0.1% for rating), which further backs the general superiority of the compensatory model in our data set.

3.3.2 Experts vs. non-experts

We divided respondents into a non-expert subgroup and an expert subgroup of active skiing instructors. In more detail, only people indicating that they had started or completed training for being a skiing instructor and were

¹⁷To address concerns of task order effects, we repeated the analyses for only the first task. The same pattern of results was found: WADD outperformed LBA for both ranking (15.7% vs. 18.7% violated pairs) and rating (17.3% vs. 22.0% violated pairs), and LBA was better at predicting ranking compared to rating data. When Task Order added as a between-subjects factor in the repeated measurement ANOVA, its main effect was not significant, $F(1,140) = 2.92, p = .090$, nor were there significant interactions.

Table 1: Percentage of violated pairs produced by LBA and WADD for hold-out data.

	Ranking	Rating
WADD	16.1 % (10.8)	15.2 % (8.9)
LBA	18.7 % (11.4)	20.3 % (9.6)

Note: Percentages refer to the proportion of pairs including at least one hold-out option that were wrongly predicted by the respective strategy, averaged across 153 cross-validation runs and across participants ($n = 142$). Pairs of options to which participants had assigned the same value were excluded from the rating data. Standard deviations are given in parenthesis.

active for at least eight days per skiing season were considered experts. According to this criterion, we could identify 27 experts. This sample size is sufficient for statistical group comparisons so we did not have to rely on softer and more subjective ratings of expertise by respondents. Figure 3 shows that experts' stated preferences tended to be generally more predictable than those of non-experts regardless of the model applied. However, when Expertise was added as a between-subjects factor in the repeated measurement ANOVA of individual-level predictive accuracy, with Task and Model as within-subjects factors, the main effect of Expertise was not significant, $F(1,140) = 2.86, p = .093$, not were there signif-

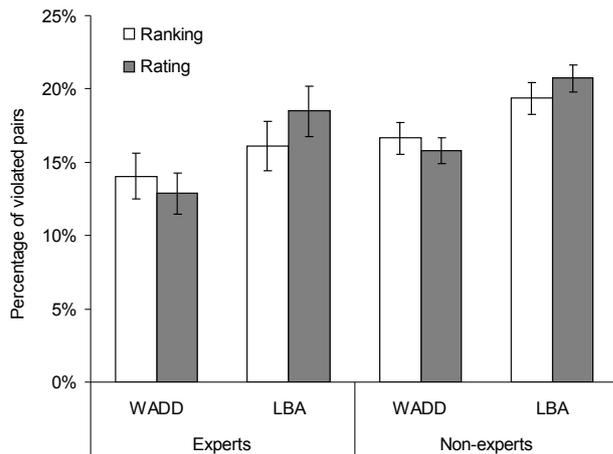


Figure 3: Mean percentage of violated pairs produced by the WADD and LBA models when applied to experts' and non-experts' ranking and rating data. Error bars represent standard errors.

icant interactions (one-way or two-way) between Expertise and Task or Model.

3.3.3 Descriptive analysis of individual differences.

Nevertheless, WADD was not the best model for all participants. For the ranking task, LBA achieved higher mean predictive accuracy than WADD for 35% of the participants ($n = 50$). For the rating task, LBA still achieved higher mean accuracy for 25% of participants ($n = 36$). In Figure 4, the difference in mean percentage of violated pairs between LBA and WADD is plotted for each respondent. Higher values indicate more violated pairs produced by LBA, that is, superiority of the WADD model. Respondents are ordered in decreasing order according to the size of difference between LBA and WADD for the ranking task (plotted as dots). The number of dots below zero is 50, corresponding to the number of respondents for which the LBA model achieved higher accuracy in the ranking task.

The respective difference values for the rating task, for the same participants, are also shown in Figure 4 (plotted as crosses). There are no visible hints that participants strive for inter-task consistency in their strategy use. Indeed, for only 16 of the 50 participants for whom the LBA model achieved higher accuracy for ranking, it also achieved higher accuracy for rating. Note that assuming chance group assignment of the 50 participants, for already 13 participants it can be expected that the LBA model is also superior for rating (i.e., $50/142 * 36$). Similarly, for 72 of the 92 for whom the WADD model achieved higher accuracy for ranking, it also achieved higher accu-

racy for rating, with 69 expected by chance assignment.¹⁸ In line with these findings, the correlation between the two vectors of difference values is very low, with $r = .13$.

4 Discussion

In line with our hypothesis, the lexicographic model was better in predicting ranking compared to rating data. As suggested by other authors, the simultaneous presentation mode and the required ranking response obviously promoted lexicographic processing (e.g., Dhar, 1996; Schkade & Johnson, 1989; Tversky, 1969). However, the compensatory model derived from conjoint analysis proved superior to the lexicographic model based on the aspect orders derived from Yee et al.'s (2007) greedoid algorithm regardless of the response mode in which participants had to indicate their preferences. This result was achieved despite the use of basic, conservative benchmark estimation procedure for the WADD model, that is, ordinary least square regression. Applying more powerful estimation procedures may have resulted in even greater superiority of the WADD model.

The relatively high predictive performance of the WADD model is in stark contrast to the results reported by Yee et al. (2007). One possible reason for the WADD models' inferiority in their study is that the number of options and aspects was higher: Yee et al. used 32 options described on 7 attributes with 20 aspects in total, whereas we used 18 options described on 6 attributes with 14 aspects in total. Thus, their task was more complex, increasing the need for simplifying heuristics. Payne (1976) as well as Billings and Marcus (1983) suggest that a relatively large number of options induces a shift from compensatory to noncompensatory processing with the goal of reducing the number of relevant alternatives as quickly as possible. Some authors also report more use of noncompensatory strategies when the number of attributes increases (Biggs, Bedard, Gaber & Linsmeier, 1985; Sundström, 1987). Thus, the effects of task complexity on the performance of LBA versus WADD models deserve exploration in controlled experiments in the future.

¹⁸In line with these results, predictive accuracies for the two models drop considerably when the model fitted to the responses in one task is applied to predict preferences expressed in the other task. When ranking is predicted by the WADD models fitted to the rating task, the average percentage of violated pairs is 24.4%; when rating is predicted by the WADD models fitted to the ranking task, it is 20.3%. For LBA, the corresponding percentages are 27.2% (ranking predicted by LBA models fitted to rating) and 23.0% (rating predicted by LBA models fitted to ranking). The losses in predictive accuracy when models are applied to predict a different task than the one they are fitted to are thus larger than the accuracy differences between the two models within one task.

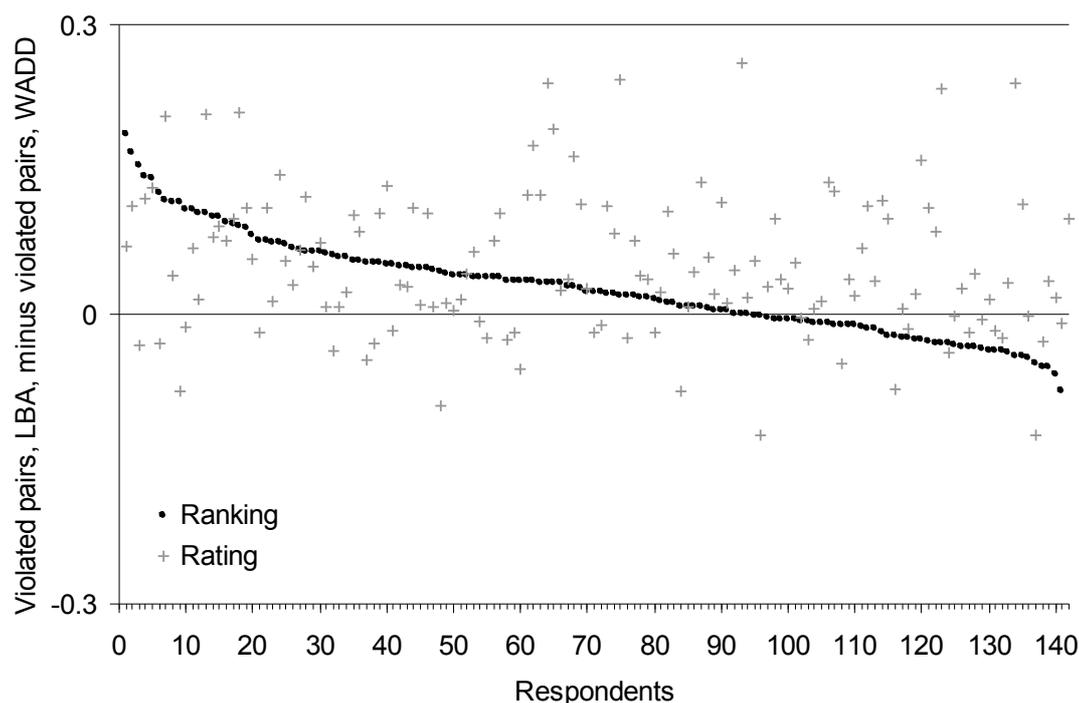


Figure 4: Differences between mean percentages of violated pairs produced by LBA and mean percentages produced by WADD for ranking (black dots) and rating (grey crosses) tasks for each respondent. Respondents are ordered in decreasing order according to the size of difference between LBA and WADD for the ranking task. Values above zero indicate superiority of WADD model (i.e., more violated pairs produced by LBA), and vice versa.

Moreover, two of our attributes seemed to be of utmost importance to many participants (see Figures 1 and 2). Given this importance structure, tradeoffs between the most important attributes seem to be within reach even given limited time and processing capacities. In sum, the circumstances under which the new greedoid approach can be fruitfully applied as a general tool require further exploration. Our research suggests that with few attributes to consider and relatively few options to evaluate, the standard approach will provide higher predictive accuracy on average, for both rating and ranking tasks. However, the WADD model does not outperform LBA for each individual participant. The LBA model is better in predicting the choices of a considerable proportion of people. It might therefore be useful to further study these differences to derive rules for assigning individual participants to certain decision strategies. However, there is little consistency across tasks in terms of which is the more accurate decision model. That is, for a large proportion of people for whom the LBA model was better at predicting the ranking data, the WADD model was better at predicting the rating data, and vice versa. Thus, the results do not seem to be skewed from participants' striving for consistent answers across tasks. But at the same time, the observed inconsistency rules out the assumption of habitual preferences for certain strategies. So, many peo-

ple seem to apply different strategies depending on the preference elicitation method. This diversity in responses to task demands will complicate assignment of participants to strategies.

We hypothesized that expertise would be one individual difference variable that affects strategy selection. However, the lexicographic model achieved only marginally higher accuracy for experts than for non-experts. Also, contrary to expectation, the WADD model still outperformed the lexicographic model in predicting expert decisions. A reason could be that our product category related to a leisure activity for which expertise is likely to be highly correlated with personal interest and emotional involvement. There is empirical evidence that involvement with the decision subject is associated with thorough information examination and simultaneous, alternative-wise processing, while the lack of it leads to attribute-wise information processing (Gensch & Javalgi, 1987). Thus, in addition to the situational factors promoting compensatory decision making over all participants, emotional involvement might have led to relatively high levels of compensatory processing in experts. Future studies should aim at distinguishing between the concepts of expertise and involvement to study their possibly opposite effects on strategy selection.

5 Conclusion

The development of the greedoid algorithm to deduce lexicographic processes offers great potential for the fields of judgment and decision making as well as consumer science. For the first time, a relatively simple and fast tool for deriving lexicographic processes is available, applicable to different kinds of preference data. However, it has to be doubted that the new approach represents a universal tool that will replace the established ones. For decision tasks with relatively low complexity — that is, with few aspects and options — the standard weighted additive model led to superior predictive accuracy for both ranking and rating data compared to the lexicographic model deduced with the greedoid algorithm. To provide advice to practitioners on when the new analysis method might prove useful, we clearly need to find out more about the conditions under which, and the people for whom lexicographic models lead to superior predictions. Based on previous research, situations with significant time pressure, complex decision tasks, or high cost of information gathering might represent favorable conditions for lexicographic processing, and thus for the application of the greedoid algorithm (Bröder, 2000; Payne, 1976; Payne et al., 1988).

People simplify choices in many ways. Verbal protocol studies have revealed many different cognitive processes and rules that decision makers apply, of which non-compensatory lexicographic decision rules are just one example (e.g., Einhorn et al., 1979). There definitely is demand for models that are descriptive of what goes on in decision makers' minds when confronted with the abundant choices their environment has to offer. Combining lexicographic and compensatory processes in one model might be a promising route to follow. Several authors have argued that noncompensatory strategies are characteristic of the first stage of choice, when the available options are winnowed down to a consideration set of manageable size (e.g., Bettman & Park, 1980; Gilbride & Allenby, 2004; Payne, 1976). Once the choice problem has been simplified, people may be able to apply or at least to approximate compensatory processes, which is in line with our results. The prevalence of combinations of lexicographic elimination and additive strategies is further backed by recent evidence from verbal protocol analyses (Reisen, Hoffrage & Mast, 2008). There is preliminary work by Gaskin, Evgeniou, Bailiff & Hauser (2007) trying to combine lexicographic and compensatory processes in a two-stage model, with lexicographic processing, estimated with the greedoid algorithm, on the first stage. We are curious how these approaches will turn out in terms of predictive accuracy.

References

- Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, 3, 439–449.
- Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, 7, 234–248.
- Biggs, S. F., Bedard, J. C., Gaber, B. G., & Linsmeier, T. J. (1985). The effects of task size and similarity on the decision behavior of bank loan officers. *Management Science*, 31, 970–987.
- Billings, R. S., & Marcus, S. A. (1983). Measures of compensatory and noncompensatory models of decision behavior: Process tracing versus policy capturing. *Organizational Behavior and Human Performance*, 31, 331–352.
- Billings, R. S. & Scherer, L. L. (1988). The effects of response mode and importance on decision making strategies: Judgment versus choice. *Organizational Behavior and Human Performance*, 41, 1–19.
- Bradlow, E. T. (2005). Current issues and a “wish list” for conjoint analysis. *Applied Stochastic Models in Business and Industry*, 21, 319–324.
- Bröder, A. (2000). Assessing the empirical validity of the “Take-The-Best” heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332–1346.
- Bröder, A. & Gaissmaier, W. (2007). Sequential processing of cues in memory-based multi-attribute decisions. *Psychonomic Bulletin and Review*, 14, 895–900.
- Bröder, A., & Schiffer, S. (2003a). “Take The Best” versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277–293.
- Bröder, A., & Schiffer, S. (2003b). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making*, 16, 193–213.
- Carmone, F. J., Green, P. E., & Jain, A. K. (1978). Robustness of conjoint analysis: Some Monte Carlo results. *Journal of Marketing Research*, 15, 300–303.
- Cattin, P. & Bliemel, F. (1978). Metric vs. nonmetric procedures for multiattribute modelling: Some simulation results. *Decision Sciences*, 9, 472–480.
- Cattin, P. & Wittink, D. R. (1977). Further beyond conjoint measurement: Toward a comparison of methods. *Advances in Consumer Research*, 4, 41–45.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

- ford University Press.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
- Denstadli, J. M., & Lines, R. (2007). Conjoint respondents as adaptive decision makers. *International Journal of Market Research*, 49, 117–132.
- Dhar, R. (1996). The effect of decision strategy on deciding to defer choice. *Journal of Behavioral Decision Making*, 9, 265–81.
- Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, 35, 1801–1813.
- Dippold, K. (2007). Optimierung der Prognosegüte von Conjoint-Befragungen durch Anwendung des Greedoid-Algorithmus. [Optimization of predictive accuracy of conjoint questionnaires through application of the greedoid algorithm.] Unpublished diploma thesis, University of Regensburg, Germany.
- Einhorn, H. J. (1971). Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance*, 6, 1–27.
- Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, 86, 465–485.
- Elrod, T., Johnson, R. D., & White, J. (2004). A new integrated model of noncompensatory and compensatory decision strategies. *Organizational Behavior and Human Decision Process*, 95, 1–19.
- Elrod, T., Louviere, J. J., & Davey, K. S. (1992). An empirical comparison of ratings-based and choice-based models. *Journal of Marketing Research*, 29, 368–377.
- Ettenson, R., Shanteau, J., & Krogstad, J. (1987). Expert judgment: Is more information better? *Psychological Reports*, 60, 227–238.
- Ford, J. K., Schmitt, N., Schechtman, S. L., Hulst, B. H., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, 43, 75–117.
- Gaskin, S., Evgeniou, T., Bailiff, D., & Hauser, J. (2007). Two-stage models: Identifying non-compensatory heuristics for the consideration set then adaptive polyhedral methods within the consideration set. *Proceedings of the Sawtooth Software Conference*, 13, 67–83.
- Gensch, D. H., & Javalgi, R. G. (1987). The influence of involvement on disaggregate attribute choice models. *Journal of Consumer Research*, 14, 71–82.
- Gigerenzer, G., Czerlinski, J., & Martignon, L. (1999). How good are fast and frugal heuristics? In J. Shanteau, B. Mellers, & D. Schum (Eds.), *Decision research from Bayesian approaches to normative systems: Reflections on the contributions of Ward Edwards* (pp. 81–103). Norwell, MA: Kluwer.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gilbride, T., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, 23, 391–406.
- Green, P. E., Krieger, A. M., & Agarwal, M. K. (1991). Adaptive conjoint analysis: Some caveats and suggestions. *Journal of Marketing Research*, 28, 215–222.
- Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8, 355–363.
- Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, 5, 103–123.
- Green, P. E., & Wind, Y. (1975). New ways to measure consumer judgments. *Harvard Business Review*, 53, 107–117.
- Hartmann, A., & Sattler, H. (2002). *Commercial use of conjoint analysis in Germany, Austria and Switzerland* (Research Papers on Marketing and Retailing No. 6). Germany: University of Hamburg.
- Johnson, J. G., Wilke, A., & Weber, E. U. (2004). Beyond a trait view of risk taking: A domain-specific scale measuring risk perceptions, expected benefits, and perceived-risk attitudes in German-speaking populations. *Polish Psychological Bulletin*, 35, 153–163.
- Johnson, R. M. (1987). Adaptive conjoint analysis. *Proceedings of the Sawtooth Software Conference*, 1, 253–265.
- Kohli, R., & Jedidi, K. (2007). Representation and inference of lexicographic preference models and their variants. *Marketing Science*, 26, 380–399.
- Lenk, P. J., DeSarbo, W. S., Green, P. E., & Young, M. R. (1996). Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science*, 15, 173–191.
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, 70, 556–564.
- Louviere, J. J., Eagle, T. C., & Cohen, S. H. (2005). *Conjoint analysis: Methods, myths and much more* (CenSoC Working Paper No. 05–001). Sydney, Australia: University of Technology, Faculty of Business, Centre for the Study of Choice.
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1–

- 27.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52, 29–71.
- Mishra, S., Umesh, U. N., & Stem, D. E. (1989). Attribute importance weights in conjoint analysis: Bias and precision. *Advances in Consumer Research*, 16, 605–611.
- Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing “one-reason” decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 53–65.
- Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone “takes-the-best.” *Organizational Behavior and Human Decision Processes*, 91, 82–96.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Nowlis, S. M., & Simonson, I. (1997). Attribute-task compatibility as a determinant of consumer preference reversals,” *Journal of Marketing Research*, 34, 205–218.
- Payne, J. W. (1976). Task complexity and contingent processing in decision processing: An information search and protocol analysis. *Organizational Behavior and Human Decision Processes*, 16, 366–387.
- Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, 92, 382–402.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534–552.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, England: Cambridge University Press.
- Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, 6, 421–425.
- Reisen, N., Hoffrage, U., & Mast, F. W. (2008). Identifying decision strategies in a consumer choice situation. *Judgment and Decision Making*, 3, 641–658.
- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.
- Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, 127, 258–276.
- Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236.
- Rossi, P. E., & Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science*, 22, 304–328.
- Russo, J. E., & Doshier, B. A. (1983). Strategies for multi-attribute binary choice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, 676–696.
- Schkade, D. A., & Johnson, E. J. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes*, 44(2), 203–231.
- Schmalhofer, F., & Gertzen, H. (1986). Judgment as a component decision process for choosing between sequentially available alternatives. In B. Brehmer, H. Jungermann, P. Lourens, & G. Sevon (Eds.), *New Directions in Research on Decision Making* (pp. 139–150). Amsterdam: North-Holland/ Elsevier.
- Shanteau, J. (1992). How much information does an expert use? Is it relevant? *Acta Psychologica*, 81, 75–86.
- Srinivasan, V. (1998). A strict paired comparison linear programming approach to nonmetric conjoint analysis. In J. E. Aronson & S. Zionts (Eds.), *Operations Research: Methods, Models, and Applications* (pp. 97–111). Quorum Books, Westport, CT.
- Sundström, G. A. (1987). Information search and decision making: The effects of information displays. *Acta Psychologica*, 65, 165–179.
- Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, 23, 86–112.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31–48.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281–299.
- Westenberg, M., & Koele, P. (1992). Response modes, decision processes and decision outcomes. *Acta Psychologica*, 80, 169–184.
- Wildner, R., Dietrich, H., & Hölscher, A. (2007). HILCA: A new conjoint procedure for an improved portrayal of purchase decisions on complex products. *Yearbook of Marketing and Consumer Research*, 5, 5–20.
- Wittink, D. R., & Cattin, P. (1981). Alternative estimation methods for conjoint analysis: A Monte Carlo study. *Journal of Marketing Research*, 18, 101–106.
- Wittink, D. R., & Cattin, P. (1989). Commercial use of conjoint analysis: An update. *Journal of Marketing*, 53, 91–96.
- Wittink, D. R., Huber, J. C., Zandan, P., & Johnson, R. M. (1992). The number of levels effect in conjoint: Where does it come from, and can it be eliminated? *Sawtooth Conference Proceedings*. Ketchum, ID: Sawtooth Software.

- Wittink, D. R., Vriens, M., & Burhenne, W. (1994). Commercial use of conjoint in Europe: Results and critical reflections. *International Journal of Research in Marketing*, 11, 41–52.
- Wulf, S. (2008). Traditionelle nicht-metrische Conjoint-analyse — ein Verfahrensvergleich. [Traditional non-metric conjoint analysis — a comparison of methods.] LIT Verlag, Hamburg.
- Yee, M., Dahan, E., Hauser, J. R., & Orlin, J. (2007). Greedoid-based noncompensatory inference. *Marketing Science*, 26, 532–549.

Appendix: Table of product profiles

Calibration as well as hold-out profiles were generated with SPSS Orthoplan. Orthogonality of the fractional factorial design of calibration profiles is ensured.

Card ID	Price	Waterproofness	Hood	Sealed seams	Skipass pocket	Ventilation zippers
1 ^a	339€	Up to medium snowfall/ light rain	Non-adjustable	Yes	No	Yes
2	259€	Up to light snowfall	Non-adjustable	Yes	No	No
3	259€	Up to light snowfall	Elasticated	No	Yes	Yes
4	179€	Up to light snowfall	Non-adjustable	No	No	No
5	339€	Up to medium snowfall/ light rain	Elasticated	Yes	No	No
6	179€	Up to light snowfall	Elasticated	Yes	Yes	No
7	259€	Up to strong snowfall/rain	Elasticated	Yes	No	Yes
8	339€	Up to strong snowfall/ rain	Non-adjustable	No	Yes	Yes
9	339€	Up to light snowfall	Elasticated	No	Yes	No
10	179€	Up to medium snowfall/ light rain	Non-adjustable	Yes	Yes	Yes
11	179€	Up to strong snowfall/ rain	Elasticated	No	No	No
12	339€	Up to light snowfall	Non-adjustable	Yes	No	Yes
13	179€	Up to light snowfall	Elasticated	Yes	Yes	Yes
14	259€	Up to medium snowfall/ light rain	Non-adjustable	No	Yes	No
15 ^a	179€	Up to light snowfall	Non-adjustable	No	Yes	No
16	179€	Up to light snowfall	Non-adjustable	No	No	Yes
17	179€	Up to strong snowfall/ rain	Non-adjustable	Yes	Yes	No
18	179€	Up to medium snowfall/ light rain	Elasticated	No	No	Yes

^a Profiles originally included as the only two hold-out profiles (distinction not applicable anymore due to leave-two-out cross-validation procedure); remaining profiles form a balanced and orthogonal design.