# Improving Data Quality, Model Functionalities and Optimizing User Interfaces in Decision Support Systems

---

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vom Fachbereich Rechts- und Wirtschaftswissenschaften

der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades

Doctor rerum politicarum (Dr. rer. pol.)

genehmigte

## Dissertation

von Diplom-Kaufmann Markus Franz

aus Frankfurt am Main

| | |
|---|---|
| Erstreferent: | Prof. Dr. Oliver Hinz |
| Zweitreferent: | Prof. Dr. Alexander Benlian |
| Tag der Einreichung: | 27.06.2016 |
| Tag der mündlichen Prüfung: | 01.12.2016 |
| Erscheinungsort/-jahr: | Darmstadt 2017 |
| Hochschulkennziffer: | D17 |

# Table of Contents

## List of Tables

## List of Figures

## List of Abbreviations

| | |
|---|---|
| AIC | Akaike’s information criterion |
| AICT | Ambiguous Identifier Clustering Technique |
| ARI | Adjusted Rand index |
| BIC | Bayesian information criterion |
| CA | Conjoint analysis |
| CBCA | Choice-based conjoint analysis |
| DR | Direct rating |
| DSS | Decision support system |
| EM | Expectation maximization |
| EUS | End user satisfaction |
| FCHR | First choice hit rate |
| FN | False negative |
| FP | False positive |
| MAU | Multi(ple)-attribute utility |
| MAUT | Multi-attribute utility theory |
| MPE | Mean percentage error |
| NPS | Net promoter score |
| PEOU | Perceived ease of use |
| PU | Perceived usefulness |
| RBCA | Rating-based conjoint analysis |
| RBFN | Radial basis function networks |
| REML | Restricted maximum likelihood |
| RI | Reuse intention |
| SAU | single-attribute utility |
| SVD | Singular value decomposition |
| TN | True negative |
| TP | True positive |
| WTP | Willingness to pay |

# 1 Introduction

## 1.1 Motivation

Decision makers, like managers in corporations or consumers, often face challenging decision making scenarios. Managers are confronted with structured, semi-structured, and unstructured decisions (Alter 2004; Hosack et al. 2012). Examples are make-or-by decisions, software selections, or investment and R&D planning (Turban et al. 2010). In addition to quickly reacting to an ever changing environment these decisions require the processing of large amounts of data and information (Turban et al. 2010). Consumers searching for a product to meet their individual preferences face preferential choice problems (Todd and Benbasat 1994; Xiao and Benbasat 2007). These decisions can be cognitively demanding, since consumers compare product alternatives taking up to eight attributes into consideration (Jacoby et al. 1977; Moorthy et al. 1997; Olson and Jacoby 1972; Sheluga et al. 1979). Furthermore, a reasonable amount of consumers nowadays purchase products online (Schultz and Block 2015), thus having to deal with large assortment sizes. Amazon.de for example, currently offers more than 2,200 digital compact cameras, 3,400 TV sets and even in narrowly defined product categories, like movies released within the last month on Blue-ray, a choice of more than 270 alternatives (Amazon, 2016).

Without assistance, the decisions made in these situations may suffer from reduced quality. Since decision makers have a limited cognitive capacity to process information on decision alternatives (Payne 1982; Payne et al. 1988; Xiao and Benbasat 2007), having more information than can be handled leads to information overload (O'Reilly 1980) and bounded rationality (Simon 1955; Xiao and Benbasat 2007). Chewning and Harrell (1990) found evidence that information to a certain amount is beneficial for decision making performance and information beyond this point leads to information overload and decisions of lower quality. Decision makers then rather aim at making a satisfactory decision, not necessarily an optimal one (Simon 1955; Xiao and Benbasat 2007).

Decision support systems (DSSs) have been largely developed to overcome the problem of information overload and bounded rationality in decision making (Hosack et al. 2012). Basically, DSSs are information systems that take over the processing of decision relevant data in order to recommend possible solutions (Mallach 2000; Turban et al. 2010; Xiao and Benbasat 2007). Evidence on the value of DSSs can be found in prior research. Consumer DSSs, for example, have been shown to reduce the effort and improve the quality of decision making (Häubl and Trifts 2000), or to raise add-on and cross-selling opportunities (Hinz and Eckert 2010). Nowadays, DSSs are used to support decision making not only of consumers and managers, but also in areas like natural sciences or medical decision making (Hosack et al. 2012). The focus of this dissertation, however, is on DSSs supporting managerial and consumer decision making.

Research on DSSs is prospering, but still far from being complete as can be seen in recent literature reviews (e.g., Hosack et al. 2012; Park et al. 2012; Xiao and Benbasat 2007). This dissertation contributes to the literature on DSSs focusing on research domains related to three core elements each DSS consists of (Turban et al. 2010): i) data management, ii) model management, and iii) user interface.

Data management provides decision relevant information to a DSS (Turban et al. 2010). Within data management, this dissertation focuses on the data quality of online transaction data sets. Online transaction data can be used as an input to econometric models supporting decisions on, for example, shipping cost strategies (Frischmann et al. 2012) or optimal auction designs (Shiu and Sun 2014). The quality of online transaction data can suffer from unobserved product heterogeneity (e.g., Einav et al. 2015), which might affect the validity of the results if not properly addressed (Becker et al. 2013) and thus the decisions derived from those results. To address this problem, this dissertation proposes a novel approach for reducing unobserved product heterogeneity in online transaction data and investigates the approach's accuracy and impact on the goodness of econometric models.

The model management then provides specific quantitative models to analyze the decision relevant data (Turban et al. 2010). This dissertation extends the functionality of utility-based models applied to consumer DSSs in two aspects. The first functionality is the elicitation of exponential utility functions for modeling preferences and making product recommendations, which are held to reflect consumer preferences more precisely than linear functions (Van Ittersum and Pennings 2012). The second functionality extension enables the prediction consumers' individual willingness to pay (WTP) for all products within a category. Having reliable estimates of consumers' preferences and individual WTP for products is a valuable input for supporting management decisions (e.g., Jedidi et al. 1996; Rusmevichientong et al. 2010; Schlereth and Skiera 2012; Wu et al. 2014). Therefore, this dissertation further investigates the question, which function (linear vs. exponential) predicts product utilities and consumers' WTP more accurately and shows how these data can be used to support pricing decisions of managers.

User interfaces, which enable the interaction and communication between a DSS user and the system (Turban et al. 2010) frequently apply visual depictions of information to support the decision making of managers and consumers (Dilla et al. 2010; Lurie and Manson 2007; Kelton et al. 2010). Focusing on coordinate systems as visualization method, this dissertation investigates the question whether visual depictions of information in 2D or 3D format supports decision making better when decision making scenarios are simple or complex. This research responds to a long debate on the question which visualization format (2D vs. 3D) supports decision makers best. Answering this question is important, since providing inappropriate decision support might lead to diminished decision quality and acceptance of DSSs (e.g., Xiao and Benbasat 2007), which would in case of this dissertation be due to the use of an inappropriate visualization format for supporting a decision of a particular level of complexity.

The following subsection contains a brief description of the theoretical framework of this dissertation and the gaps addressed in the related literature. The subsection thereafter briefly describes the structure of this dissertation, which consists of three published articles addressing the research gaps described in the following.

## 1.2 Theoretical Framework

The three core elements of a DSS are i) data management, ii) model management, and iii) a user interface (Turban et al. 2010). This dissertation addresses specific gaps in the literature on all three related research domains, which is briefly summarized in Figure 1.

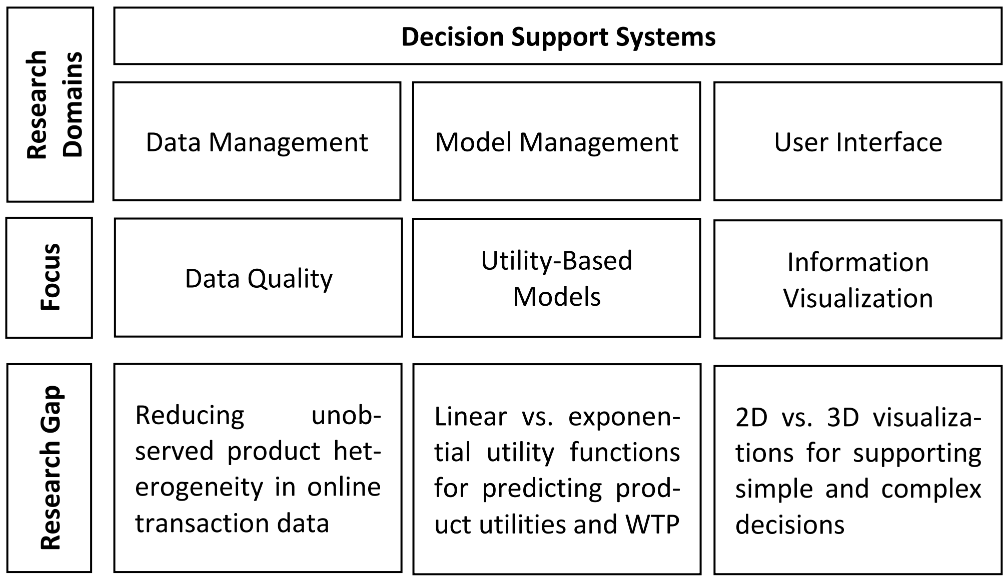



**Figure 1: Theoretical Framework of this Dissertation**

### 1.2.1 Data Management

DSSs contain a data management system including a database storing decision relevant information. Decision relevant information can originate from different sources, especially internal data (e.g., sales forecasts or an e-commerce organization's online transaction data), external data (e.g., data provided by market research) and private data (e.g., guidelines for decision making). Relevant issues in data management concern the security, integration, scalability and quality of the data used and stored for supporting decision making (Turban et al. 2010).

This dissertation addresses a gap in the literature on the quality of online transaction data. Online transaction data are generated when consumers order or bid on products from a seller on the Internet. These data can be used by econometric models to assist decisions with respect to optimal selling strategies, for example, regarding the optimal design of online auctions (Shiu and Sun 2014) or shipping cost strategies (Frischmann et al. 2012). The quality of online transaction data can suffer from unobserved or unexpected heterogeneity stemming from at least three sources: i) consumers (Bapna et al. 2003), ii) sellers (Goes et al. 2013), and iii) products (e.g., Einav et al. 2015). If not properly controlled, unobserved heterogeneity can lead to biased results and misleading conclusions (Becker et al. 2013). Approaches to control for unobserved consumer and seller heterogeneity have already been proposed by prior research (Bapna et al. 2004; Goes et al. 2013). More work is required addressing unobserved product heterogeneity, which is the focus of this dissertation.

Unobserved product heterogeneity occurs when identical or only marginally different products cannot be identified in a set of online transaction data due to a missing product identifier (Einav et al. 2015). In other words, there is no control for effects stemming from distinct products. Existing approaches for reducing unobserved product heterogeneity comprise the restriction of data sets to a few products (Bajari and Hortaçsu 2003; Bapna et al. 2009; Mudambi and Schuff 2010; Shiu and Sun 2014), conducting controlled experiments (Lucking-Reiley 1999; Ostrovsky and Schwarz 2009), manual classification of transaction data (Frischmann et al. 2012) or matching products based on exactly identical product tiles (Einav et al. 2015; Elfenbein et al. 2012). These approaches, however, limit the scope of products, are expensive or suffer from error proneness.

To address this gap in the literature, the authors propose the Ambiguous Identifier Clustering Technique (AICT) for reducing unobserved product heterogeneity in online transaction data sets. It is based on the idea of clustering heterogeneous and unidentified product offerings based on their product title (i.e., a short, textual description) for the purpose of identifying and controlling for (virtually) identical products in the data set. In an empirical study, the authors furthermore investigate the accuracy of the proposed solution and the impact of controlling for product clusters identified by AICT on the goodness of econometric models.

### 1.2.2 Model Management

The model management of DSSs contains quantitative models for analyzing data in order to support decision making (Turban et al. 2010). There are strategic, tactical, operational and analytical models which are designed for supporting top management, middle management or day-to-day decisions in organizations (Turban et al. 2010). Supporting consumer decisions usually relies on choice models aiming at finding products meeting consumers' preferences (Benbasat et al. 1991; Xiao and Benbasat 2007; Zachary 1986). Content-based and collaborative models are frequently applied for this purpose (Xiao and Benbasat 2007). A rather new approach in supporting consumer decision making applies utility-based models which are the focus of this dissertation.

Utility-based models are frequently based on multi-attribute utility theory (MAUT; Huang 2011; Keeney and Raiffa 1993). Since MAUT considers decision alternatives as bundles of attributes, a multi-attribute utility (MAU) function is used to calculate a utility value for each decision alternative in order to recommend products to consumers for purchase (Huang 2011; Scholz Dorner 2012). The MAU function therefore (usually additively) combines: i) single-attribute utility (SAU) functions transforming the level of each attribute into a utility value that are ii) weighted by an individual weight for each attribute (Butler et al. 2008; Huang 2011).

Recent research came up with a number of DSSs applying utility-based models (e.g., Choi and Cho 2004; Guan et al. 2002; Huang 2011; Lee 2004; Liu and Shih 2005; Manouselis and Costopoulou 2007; Schickel-Zuber and Faltings 2005; Schmitt et al. 2002; Scholz and Dorner 2012; Stolze and Stroebel 2003; Tewari et al. 2002). This dissertation addresses two gaps in the literature on utility-based models regarding their ability to elicit and the suitability of using exponential SAU functions to predict product utilities and consumers' individual willingness to pay.

First, SAU functions can have linear (Green et al. 2001; Scholz et al. 2010) and exponential (Harvey 1981; Van Ittersum and Pennings 2012) shapes, which may be different for each consumer and each of the decision alternative's attributes (Van Ittersum and Pennings 2012). Utility-based DSSs should thus elicit flexible SAU functions. Whereas the elicitation of linear SAU functions is rather uncomplicated, exponential SAU functions are rather unpopular for utility based DSSs (Butler et al. 2008; Scholz and Dorner 2012). Approaches like manual selection of appropriate SAU function shapes (e.g., DeSarbo et al. 1995; Green et al. 2001) are not practical for consumer DSSs. Other approaches, like specifying utility values for each attribute level (e.g., Butler et al. 2008) require additional interaction and effort. Consumers though are typically not willing to incur high levels of effort for the purpose of preference elicitation (De Bruyn et al. 2008). This dissertation contributes to the literature of utility-based models by proposing a new approach for eliciting exponential SAU functions at a low level of effort and by empirically investigating the utility accuracy of linear vs. exponential SAU functions.

Second, consumers' preferences and individual WTP for products both are critical inputs for managerial decision models. Examples are pricing strategies (e.g., Schlereth and Skiera 2012; Wu et al. 2014), market share estimation (e.g., Jedidi et al. 1996) or product assortment decisions (e.g., Rusmevichientong et al. 2010). WTP can be defined as a consumer's indifference price regarding a purchasing decision (Moorthy et al. 1997). Methods for estimating consumers' WTP are based on direct WTP elicitation (e.g., Backhaus et al. 2005; Voelckner 2006; Miller et al. 2011), bidding behavior observed in auctions (e.g., Voelckner 2006; Chan et al. 2007; Barrot et al. 2010), direct utility elicitation (e.g., Louviere and Islam 2008; Park et al. 2008), conjoint analyses (e.g. Jedidi and Zhang 2002; Gensler et al. 2012;  Voelckner 2006) or market data (e.g., Kamakura and Russell 1993; Leeflang and Wittink 1992). These methods are rather designed for evaluating consumers' WTP at one point in time. Preferences of consumers may however change. Repeating the measures to keep the estimates current may result in high costs for companies and high effort for consumers. Furthermore, consumers' WTP should be elicited under the actual marketing mix conditions in order to be reliable (Wertenbroch and Skiera 2002). In this dissertation, the authors extend the functionality of utility-based models towards predicting consumers' individual WTP at the time and at the point of purchase for all products within a category. In an empirical study, the authors further investigate the accuracy of the proposed WTP estimation approach when preferences are modeled using linear vs. exponential SAU functions and demonstrate how the data on consumer preferences and individual WTP can be used to support pricing decisions.

### 1.2.3 User Interface

User interfaces take on the task of communicating between users and the DSS. By using interfaces, DSS users can interact with the data and model management systems and access decision relevant data, knowledge and information. The interaction with a DSS can be facilitated by using graphical, auditive or other sensing devices (Turban et al. 2010).

This dissertation addresses a specific gap in the literature on information visualization applied to user interfaces of DSSs. Information visualization (Card et al. 1999) encompasses the process of transforming information (e.g., texts or numbers) into a visual representation (Lurie and Manson 2007). The idea behind information visualization is to free up cognitive resources for solving problems by transferring cognitive load into the human perceptual system and making use of the human capabilities to process and encode visual information (Kosslyn 1994; Lurie and Manson 2007; Tegarden 1999). Information visualization is frequently used to support management and consumer decision making (Dilla et al. 2010; Kelton et al. 2010; Lurie and Manson 2007). Research on information visualization investigates the links between task characteristics, decision maker characteristics, different visualizations and decision making (Dilla et al. 2010). Supporting decision making inappropriately can negatively affect decision quality or the intention to adopt a DSS (Xiao and Benbasat 2007). Therefore choosing an appropriate visualization is an important decision.

This dissertation addresses a debate regarding the question whether 2D or 3D visualizations support decision making best. Previous studies comparing 2D to 3D visualizations mostly focus on decision making performance measures and differ in their suggestion for one particular visualization format (Dull and Tegarden 1999; Kumar and Benbasat 2004; Pilon and Friedman 1998) or find mixed evidence on the value of 2D vs. 3D visualizations (Kim et al. 2011; Nah et al. 2011; Van der Land et al. 2013; Zhu and Chen 2005). Only the studies of Lee et al. (1986) and Tractinsky and Meyer (1999) find support for a particular visualization format depending on certain conditions (i.e., the information format and purpose of the visualization). Supporting decisions can, however, impact on decision making performance and the decision makers' perceptions of the decision support (Xiao and Benbasat 2007). The authors address this gap by investigating the decision support of 2D and 3D visualizations in simple and complex (consumer) decision making scenarios jointly in terms of decision making performance and decision makers' perceptions of the decision support.

## 1.3 Structure of the Dissertation

This dissertation contributes to specific research gaps in the literature on data management, model management and user interfaces of DSSs. This contribution is made in three published articles included in this dissertation (see Table 1).

**Table 1: Articles Included in this Dissertation**

| Article | Article 1 | Article 2 | Article 3 |
|---|---|---|---|
| **Title** | The Ambiguous Identifier Clustering Technique | Measuring Consumers' Willingness to Pay with Utility-Based Recommendation Systems | 2D versus 3D Visualizations in Decision Support – The Impact of Decision Makers' Perceptions |
| **Authors** | Scholz, M. / Franz, M. / Hinz, O. | Scholz, M. / Dorner, V. / Franz, M. / Hinz, O. | Franz, M. / Scholz, M. / Hinz, O. |
| **Research Gap** | Reducing unobserved product heterogeneity in online transaction data | Linear vs. exponential utility functions for predicting product utilities and WTP | 2D vs. 3D visualizations for supporting simple and complex decisions |
| **Study Type** | Field study (N = 5,511) | Laboratory experiment ($N_1$ = 93, $N_2$ = 77) | Laboratory experiment (N = 112) |
| **Main Findings** | AICT is highly accurate, improves goodness of econometric models, reveals impacts on regression estimates and enables creating new variables | Exponential SAU functions better suited for predicting product ranks, linear better for predicting most preferred products and WTP | For a given level of complexity, decision making performance is unaffected by visualization format, but superior perceptions of 2D visualizations in simple scenarios |
| **Publication** | Electronic Markets (2016), 26(2), 143-156 | Decision Support Systems (2015), 72, 60-71 | International Conference on Information Systems (2015), Fort Worth, USA |

Article 1 addresses the research gap regarding unobserved product heterogeneity in online transaction data sets identified in the literature on data management. The extension of utility-based model functionalities, which is related to the literature on model management, is the content of Article 2. Finally, the research gap regarding the use of 2D vs. 3D visualizations to support decision making identified in the literature on user interfaces is addressed in Article 3. Table 1 contains a summary of each article with respect to publishing information, the research gaps addressed, the study type and the study's main finings. The following passages give a brief overview of the three articles.

**Article 1: The Ambiguous Identifier Clustering Technique**

Current approaches addressing unobserved product heterogeneity in online transaction data suffer from some limitations with respect to the scope of products, error proneness or costs. To improve the data quality of online transaction data, the first article of this dissertation proposes the AICT approach for reducing unobserved product heterogeneity. In an empirical setting, the authors further investigate AICT's clustering accuracy, i.e. its ability to correctly identify and group (virtually) identical product offerings, and the impact of controlling for product clusters identified by AICT on the goodness of econometric models.

AICT uses product titles (short, unstructured texts describing a product offering) to identify (virtually) identical products in online transaction data sets with heterogeneous products. Product titles are therefore standardized, divided into charter-based *bi*-grams (sequences of two letters; e.g., Abbasi and Chen 2008), and clustered using agglomerative hierarchical clustering algorithms with average linkage (e.g., Kaufman and Rousseeuw 2005) and multiscale bootstrap resampling (Shimodaria 2004) to determine the optimal clustering solution. The authors use the resulting product clusters to reduce unobserved product heterogeneity by controlling for identified groups of (virtually) identical products in econometric models.

For the empirical investigation, the authors collected and clustered an online transaction data set applying AICT including 5,511 products auctioned on eBay in the categories books, music CDs and digital cameras. A manually generated product identifier serves to evaluate the clustering accuracy of AICT and to compare its accuracy to different clustering specifications with respect to clustering methods, linkage methods, *n*-gram lengths and multiscale bootstrap specifications. To evaluate the goodness of econometric models that control for product clusters, the authors tested four different regression models aiming at predicting auctions' closing prices. All models included the same set of variables that have been identified to impact on an auction's outcome (Shiu and Sun 2014; Hou and Blodget 2010; Tan et al. 2010). AICT's data on product clusters is then used to stepwise extend a baseline model (Model 1) by controlling for cluster-specific intercepts (Model 2) and cluster-specific slopes (Model 3). Clustering products further allows creating a new variable counting auctions offering identical products at an overlapping time span, which is added in Model 4. The authors further evaluate the goodness of Model 4 based on clustering solutions provided by different clustering methods.

Our main results show that, compared to other tested clustering methods and specifications, AICT is highly accurate in identifying (virtually) identical products in online transaction data sets. Reducing unobserved product heterogeneity by controlling for product clusters identified by AICT improves the goodness of econometric models explaining an auction's closing price, also compared to other clustering methods, and affects the respective regression estimates. Finally, clustering products enables the determination of new variables explaining an auction's closing prices, in this case the number of overlapping auctions. Overall, this article provides a novel approach for improving the quality of online transaction data and demonstrates impacts of controlling for unobserved product heterogeneity using AICT on the goodness of econometric models, which can in this case be used to assist the decision regarding an optimal online auction design.

**Article 2: Measuring Consumers' Willingness to Pay with Utility-Based Recommendation Systems**

The second article extends the functionality of utility-based consumer DSS models in two different aspects: i) estimating not only linear, but also exponential SAU functions at low cognitive effort and ii) predicting consumers' willingness to pay for products within a specific category. The authors further compare the accuracy of linear and exponential SAU functions for predicting product utilities and consumers' individual WTP and show how the information on preferences and WTP generated by the extended utility-based model can be used to support managerial decisions with respect to pricing.

The authors develop a low effort approach to enable the elicitation of exponential SAU functions. Compared to eliciting linear SAU functions, the approach requires consumers to additionally specify the average attribute level's utility to determine the function's shape. Based on Butler's et al. (2012) utility exchange approach, the authors further develop an approach to enable the prediction of consumers' individual WTP for all products within a category. The approach requires the elicitation of the price for the best recommended product that makes a consumer indifferent between buying and not buying (i.e. the consumer's WTP for that particular product; Moorthy et al. 1997). The WTP for this particular product gets transformed into a utility threshold value. In this case the utility threshold is a product's utility at which a particular consumer is indifferent between buying and not buying a product in that category. This information is further used to determine the consumer's individual WTP for all products in that category.

These approaches were empirically investigated in two laboratory experiments using between-subject designs. The main experiment was designed to compare the accuracy of linear to exponential utility functions in terms of predicting product utilities and a consumer's willingness to pay. The 93 participants were randomly distributed among two versions of a utility-based DSS to search for a preferred digital camera. The first DSS elicited linear and the second one exponential utility functions. In a two part questionnaire, the authors elicited the participants' WPT and purchase probability for the first ten recommended products, the search effort, user perceptions, e-shop usage and demographics. In a supplementary experiment with 77 participants, the authors tested the fulfillment of basic assumptions of the approaches and the empirical setting.

The main results of this study are that linear utility functions do more accurately predict the consumers most preferred products and, since the WTP estimates are based on a modification of the top recommended product, also more accurately predict a consumers' WTP. Exponential SAU functions in contrast are slightly better at predicting product ranks, that is, the order among the recommended products. Additionally, the authors show how information on utility thresholds and WTP estimates generated by the extended utility-based model can be used to compute revenue-maximizing market prices and individual profit-maximizing product configurations. Overall, this study extends to the functionality spectrum of utility-based models, contributes to the question which SAU function (linear vs. exponential) is more suitable for predicting consumers' WTP and product utilities and illustrates how the data generated by this extended model can be utilized to support pricing decisions.

**Article 3: 2D versus 3D Visualizations in Decision Support – The Impact of Decision Makers' Perceptions**

The third article addresses a gap in the literature on information visualization of DSS interfaces. Most previous studies compared 2D to 3D visualizations in terms of decision making performance measures and found mixed evidence and thus no clear advice regarding the use of 2D vs. 3D visualizations for supporting decision making. Focusing on coordinate systems as visualization method, the study extends previous literature by investigating the question which visualization format, 2D vs. 3D, supports decision makers best in simple and in complex decision making scenarios. The authors further evaluate the decision support of both visualization formats based on a framework from Xiao and Benbasat (2007) not only in terms of decision making performance, but also user perceptions.

To approach the underlying research question, the authors conducted a laboratory experiment on a consumer decision making scenario. The participants were asked to use a particular DSS to search for a preferred digital camera. Decision making complexity was modified by the number of attributes describing a digital camera. The authors created a simple (complex) situation, where digital cameras were described by four (eight) camera attributes. Singular value decomposition (SVD; e.g., Zhang et al. 2007) was used to provide visual information about digital cameras and camera attributes in a 2D or in a 3D coordinate system. To investigate the decision support of a visualization format at a particular level of decision making complexity the authors compared the following four treatments: i) 2D/4 attributes vs. 3D/4 attributes and ii) 2D/8 attributes vs. 3D/8 attributes. 112 students from the University of Passau participated in the experiment and were randomly distributed among the treatments using a 2x2 between-subjects design. By tracking user behavior and a two-part questionnaire the authors collected information on the participants' decision making performance, user perceptions, demographics, psychographics and experience.

This study provides a novel finding: While the decision making performance is unaffected by the interplay of decision making complexity and visualization format, in simple decision making scenarios, when products are described by only a few attributes, users perceive 2D visualizations significantly better than 3D ones. Overall, this study contributes to the understanding of the questions i) which visualization format is appropriate for simple and complex decision making scenarios and ii) which measures should be applied to evaluate the decision support of visualizations in DSS interfaces.

This dissertation proceeds as follows. The study on data quality of online transaction data (Article 1) is included in Section 2. Section 3 comprises Article 2, which focuses on extending the functionality of utility-based models. Finally, Section 4 comprises the study addressing user interfaces (Article 3).

# 2 The Ambiguous Identifier Clustering Technique

**Title:** The Ambiguous Identifier Clustering Technique

**Authors:** Scholz, Michael; University of Passau; Germany
Franz, Markus; TU Darmstadt; Germany
Hinz, Oliver; TU Darmstadt; Germany

**Published:** Electronic Markets (2016), 26(2), 143-156

**Abstract**

Investigations of online transaction data often face the problem that entries for identical products cannot be identified as such. There is, for example, typically no unique product identifier in online auctions; retailers make their offers at price comparison sites hardly comparable and online stores often use different identifiers for virtually equal products. Existing studies typically use data sets that are restricted to one or only a few products in order to avoid product heterogeneity if a unique product identifier is not available. We propose the Ambiguous Identifier Clustering Technique (AICT) that identifies online transaction data that refer to virtually the same product. Based on a data set of eBay auctions, we demonstrate that AICT clusters online transactions for identical products with high accuracy. We further show how researchers benefit from AICT and the reduced product heterogeneity when analyzing data with econometric models.

**Keywords:** Product heterogeneity, clustering, online transaction data, e-commerce

## 2.1 Introduction

The Internet has radically improved the opportunity for IS practitioners and researchers to get access to valuable online transaction data. Such data can be used to get insights into seller strategies and consumer behavior—for example, the determinants of successful online auctions (Shiu and Sun 2014), shipping cost strategies (Frischmann et al. 2012), or the impact of customer reviews on sales (Forman et al. 2008). Empirical investigations rely on comparing transactions of identical products that are offered at different prices, at different times or to different consumer segments. These studies often face the problem of unobserved product heterogeneity: products that are identical or that differ only marginally are not identifiable as such because a unique identifier is missing or the identifier is ambiguous (Einav et al. 2015).

Online stores like Amazon often use unique product identifiers, but product variants such as the Sprint, the Verizon, the AT&T or the unlocked version of Samsung's Galaxy S6 have different identifiers. Customer reviews are assigned to only one of the versions, but almost all customers evaluate only product-specific characteristics, such as battery time, speed or handling. Reviews across these product variants are rather homogeneous and should be considered as reviews for the same product in econometric analyses.

Online auctions at eBay, for example, are described with a title and a product category but not with a product identifier that allows identifying all auctions for the same product. Without a product identifier, as is the case with eBay data, researchers and practitioners either need to restrict their data sets to only one product in order to avoid unobserved product heterogeneity or classify the offers manually.

These two examples capture the problem of unobserved product heterogeneity that often occurs in online transaction data due to missing or ambiguous product identifiers. Recent research has either ignored the problem of unobserved product heterogeneity (Stern and Stafford 2006) or avoided the problem by restricting their data sets to only one or a few products (Bajari and Hortaçsu 2003; Bapna et al. 2009; Mudambi and Schuff 2010) or classified the data by hand (Frischmann et al. 2012). Einav et al. (2015) use a simple strategy to match identical products on eBay. They consider products as identical if they are sold at eBay with the same title in the same category and by the same seller. This strategy helps to identify products that are definitely identical, but it neglects that identical products are typically sold by different sellers in often different categories and also with different titles (e.g., "Michael Jackson: Thriller" vs. "CD Thriller from Michael Jackson").

Identifying (virtually) identical products in online transaction data and hence reducing product heterogeneity in econometric analyses helps researchers to i) estimate demand curves for particular products or product categories, ii) identify product-specific effects on online transaction success variables such as auction success, auction price, conversion rate, or demand, iii) evaluate the effectiveness of transaction parameters such as the buy-it-now option in online auctions, and iv) analyze online consumer behavior.

In this paper, we contribute to existing research by proposing AICT (Ambiguous Identifier Clustering Technique) as, to the best of our knowledge, first method to cluster transaction data based on product titles (i.e., very short texts) to identify offers for virtually identical products. This allows reducing unobserved product heterogeneity in online transaction data sets by identifying and controlling for identical products. AICT clusters online transactions by the titles of the products offered in the transactions and finds those possible cluster solutions that are highly robust without knowing the number of corresponding products—and as we show in an empirical setting—with an outstanding accuracy. We hence provide a research method for helping IS and marketing researchers to better estimate demand curves from online transaction data, to better identify product-specific effects on online transaction success variables, to better evaluate the effectiveness of online transaction parameters, and to better examine online consumers' behavior.

In the next section, we discuss the types and implications of heterogeneity in online transaction data for econometric models. We then draw on recent research on identifying duplicates in databases and develop a novel method (AICT) for identifying duplicates based on unstructured very short texts (i.e., product titles) that might be incorrectly typed, rephrased or slightly different due to product variants. We demonstrate that AICT clusters online transaction data with respect to products with high accuracy and thereafter show how product clusters are applicable to reduce product heterogeneity and improve robustness and prediction accuracy of econometric models. In the final section, we conclude with a discussion of the managerial and research implications.

## 2.2 Heterogeneity in Online Transaction Data

Online transaction data describe events in which a consumer orders or bids on products offered by a seller. Researchers largely investigate these transaction data to get insights into i) the demand of products, ii) sales strategies, and iii) consumer behavior (Liu and Sutanto 2012; Zhou and Hinz 2015). Although collecting these data is fairly easy, analyzing the data is challenging. Transaction data can contain unexpected or unobserved heterogeneity which limits the ability to correctly model economically relevant effects. Both, unexpected and unobserved heterogeneity thus lead to biased parameter estimates and ultimately erroneous conclusions (Becker et al. 2013). Online transaction data can be subject to unexpected or unobserved heterogeneity of at least three sources: i) consumers (e.g., different bidding strategies in online auctions; Bapna et al. 2003), ii) sellers (e.g., different experience levels of sellers; Goes et al. 2013), and iii) products (e.g., substitutes; Einav et al. 2015).

Recent research has proposed methods to control for unobserved consumer and seller heterogeneity (Bapna et al. 2004; Goes et al. 2013). Methods for controlling unobserved product heterogeneity are largely lacking. Products are heterogeneous especially in their value and their market size (Hou and Blodgett 2010). This heterogeneity affects important and frequently used performance variables such as the price in an online transaction. Product heterogeneity is easily manageable if products have a unique identifier and product variants can be matched. This is, however, not the case on several online platforms such as eBay, Pricewatch or Epinions. The problem of unobserved product heterogeneity is typically addressed by restricting data sets to only one or a few products (Bajari and Hortaçsu 2003; Bapna et al. 2009; Mudambi and Schuff 2010; Shiu and Sun 2014), conducting experiments (Lucking-Reiley 1999; Ostrovsky and Schwarz 2009), classifying data by hand (Frischmann et al. 2012), or matching products having exactly identical titles (Einav et al. 2015; Elfenbein et al. 2012).

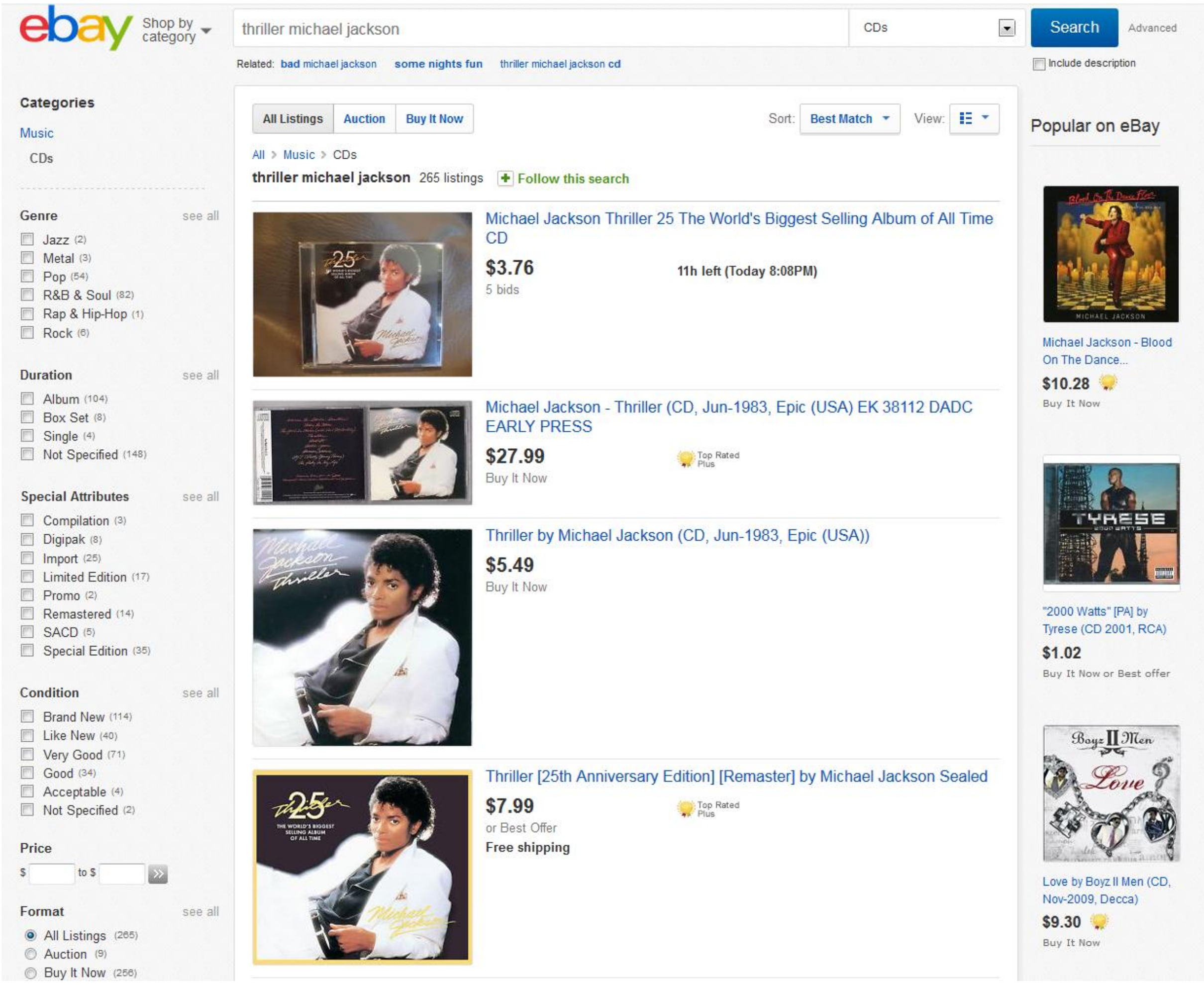



**Figure 2: Example of Identical Products with Different Product Titles and No Product Identifier**

Restricting data sets to only a few products strongly limits the insights of a study. Field experiments suffer from a limited scope of products that can be investigated and high costs for conducting such experiments. Classifying data by hand is time-consuming, costly, and error-prone. Matching products by their title and product category helps to identify products that are definitely identical. Such a strategy, however, is not suitable to identify identical products with different titles or in different categories. Figure 2 shows different auctions for identical products that are described with different titles. These products hence cannot be matched by their title when focusing on exact matches with the identical word order. The products shown in Figure 2 are furthermore sold by different sellers. Empirical strategies such as matching products (Einav et al. 2015) do consider the CDs in Figure 2 as distinct products. Since products such as CDs often do not have a product description but, as shown in Figure 2, offered at very different prices, there is often no other information than the product title for matching virtually identical products in online transaction data.

Analyzing, for example, factors that determine auction success requires knowing how many substitutes are available at which prices. The last CD on Figure 2 is probably not sold due to the availability of substitutes that are offered at lower prices. The starting price hence might determine auction success but this effect is typically moderated by the starting prices and the availability of substitutes.

We contribute to existing research by providing an approach for identifying online transactions that refer to virtually the same product based on product titles – a variable that is available in almost all data sets. Our problem is similar to the identification of duplicates in databases. We hence discuss approaches for identifying duplicates and relate our approach to this literature in the next section.

## 2.3 Identification of Duplicates

The problem of identifying duplicates in databases has long been studied in computer science (Fellegi and Sunter 1969; Jaro 1989; Lim et al. 1993). The general problem is a missing unique identifier, so that the data itself are used to identify duplicates. Existing algorithms belong to one of three possible approaches: rule-based, probabilistic or distance-based.

Rule-based approaches (e.g., Arasu and Kaushik 2006; Lim et al. 1993; Zhao and Ram 2008) aim at applying predefined rules to test if two data entries are duplicates. Such rules might be the removal of word suffixes or the rearrangement of words (Ananthanarayanan et al. 2008). Rule-based approaches are simple to implement, but have been shown to be only moderately accurate (Ananthanarayanan et al. 2008). Differences between two duplicates that are not caused by any systematic (e.g., spelling errors) are not removable with predefined rules; rule-based approaches are thus not appropriate for data that are subject to random noise. Identifying online transaction data for identical products seems not promising with such approaches, due to the high effort of creating and maintaining rules for several product categories and the existence of random noise in product titles (e.g., spelling errors, rephrased titles).

Probabilistic approaches (e.g., Dey 2003; Li et al. 2005; Jaro 1989; Larsen and Rubin 2001; Sadinle and Fienberg 2013) first compare two data entries based on a fixed set of attributes (e.g., name of the artist, name of the song) and thereafter compute the probability that based on the comparison vector the two data entries are duplicates. These approaches are theoretically well founded and have been found to be highly accurate. The requirement of a fixed set of attributes makes these approaches applicable for structured data (e.g., student database) and unstructured data for which latent attributes can be assumed (Bhattacharya and Getoor 2006). Product titles are described with a varying set of attributes such as brand, product name or color and latent attributes may not exist in such short texts. Applying a probabilistic approach to identify duplicates based on product titles is thus not promising.

Distance-based approaches (e.g., Arasu et al. 2006; Cohen 2000; Dey et al. 2002; Tejada et al. 2001) compute the distance for each pair of data entries and identify based on a distance threshold or a cluster technique duplicates among the pairs. These approaches do not rely on rules and do not assume (latent) attributes in the data. Distance-based approaches can hence be applied to identify duplicates in almost all types of data. However, distance-based approaches are subject to two major challenges: first, identifying duplicates in a set of $m$ data entries requires $\binom{m}{2}$ pairwise comparisons and second, clustering data based on distances is often not robust to small data changes (e.g., new data entries). Distances between product titles can be computed based on several textual features (e.g., character *n*-grams, word *n*-grams[1]) and for all categories of products without defining any rules.

In the next section, we propose a new distance-based approach for identifying online transaction data that refer to virtually the same product. Our approach identifies robust clusters of online transaction data based on product titles. We thereafter demonstrate the accuracy of our approach to identify product duplicates in online transaction data and show how our approach helps to control product heterogeneity in econometric analyses.

## 2.4 Ambiguous Identifier Clustering Technique

### 2.4.1 Text Preparation

Product titles are short texts that often consist of less than 50 characters. Methods typically used to prepare texts before comparing them for similarity (e.g., stop word removal, stemming) would fur-

[1] An *n*-gram is a successional subsequence of *n* items (characters or word) of a text. The character-based *n*-grams of size 2 (*bi*-grams) of the word “text” are “te”, “ex”, “xt”. A detailed example of character-based *n*-grams can be found in Table 2.

ther reduce the number of characters. Most titles furthermore consist of merely a brand and/or product name. Part-of-speeches such as verbs or pronouns are unlikely to occur and if they occur, they are almost always a part of either the brand or the product name (e.g., “Eat This Not That”). Removing common words or words with a particular part-of-speech as in other text clustering methods (Cagnina et al. 2014) is hence not recommendable when clustering product titles. In order to improve the accuracy of identifying transactions for virtually identical products, we propose removing all words or phrases of a product title that cannot be ascribed to a particular product. Such words might point to a product category (e.g., "CD Thriller from Michael Jackson") or are used for promotion (e.g., "Best Album ever Released by Michael Jackson: Thriller").

Product titles in online platforms are defined by the seller and might hence vary due to three reasons. First, product titles might include spelling errors such as missing or swapped characters (e.g., "Micheal Jacksen: Thriler"). Second, sellers might rephrase product titles and, for example, include additional words (e.g., "Thriller from Michael Jackson"). And third, sellers might use different versions of a product title (e.g., title vs. title and subtitle) to describe their offers.

Titles of virtually identical products can thus significantly differ from each other. This complicates identifying identical product offerings. Word-based or sentence-based text measures (e.g., the number of equal words) do not tolerate differences such as spelling errors or rephrased titles. Character-based *n*-grams are, in contrast to other textual features such as sentences, words or word-based *n*-grams[2], similar for titles with spelling errors and for rephrased titles.

Separating words into character-based *n*-grams helps finding similar product titles also in the presence of spelling errors, rephrased titles or different versions of product titles. Titles with spelling errors differ only marginally in terms of their character-based *n*-grams from correctly spelled titles. For example, take the product titles for the first two transactions of Michael Jackson's album "Thriller" in Table 2. Although the second title includes three spelling errors, it is characterized by nearly the same *bi*-grams as the first auction title.

In the case of rephrased titles or different versions of a product title, the number of corresponding *n*-grams is unaffected, while the number of *n*-grams per title changes. We propose using *bi*-grams, because spelling errors such as missing or swapped characters have a smaller effect on *bi*- than larger *n*-grams. The *bi*-gram vectors for “dangerous” and “dagnerous” are different in three positions (“an”, “ng”, “ge”) whereas the *tri*-gram vectors for these two words are different in four positions (“dan”, “ang”, “nge”, “ger”). Both words furthermore consist of eight *bi*-grams but only seven *tri*-grams.

[2] An overview of textual features is presented in Abbasi and Chen (2008) and in Zheng et al. (2006).

Product titles of transactions 1 and 3 have the same number of corresponding *bi*-grams as the product title of transaction 1 has with itself. The total number of *bi*-grams is different for transaction 1 and 3. A high proportion of the *bi*-grams of transaction 3 also occur in the *bi*-gram vector for transaction 1 making both *bi*-gram vectors rather similar.

**Table 2: Product Titles of Sample Transactions**

| Transaction | Title | Prepared Title | Bi-grams |
|---|---|---|---|
| 1 | Michael Jackson: Thrill-er | michael jackson thrill | mi, ic, ch, ha, ae, el, l_, _j, ja, ac, ck, ks, so, on, n_, _t, th, hr, ri, il, ll |
| 2 | Micheal Jacksen: Thriler | micheal jacksen thril | mi, ic, ch, he, ea, al, l_, _j, ja, ac, ck, ks, se, en, n_, _t, th, hr, ri, il |
| 3 | Thriller from Michael Jackson | thrill from michael jackson | th, hr, ri, il, ll, l_, _f, fr, ro, om, m_, _m, mi, ic, ch, ha, ae, el, l_, _j, ja, ac, ck, ks, so, on |

To further improve the accuracy of clustering product titles, we recommend stemming those words that remain after stop word removal. We then compute the distance between two prepared product titles.

### 2.4.2 Clustering

We propose using a hierarchical algorithm to cluster product titles based on a distance matrix. Centroid-based algorithms (e.g., k-means, fuzzy c-means) require the number of clusters to be known in advance. Distribution-based (e.g., EM-clustering, DBCLASD) and density-based algorithms (e.g., DBSCAN, OPTICS) are not appropriate for clustering sparse text-feature × object matrices as is the case with short texts such as product titles. We adapt agglomerative hierarchical algorithms due to their lower computational complexity compared to divisive algorithms (Kaufman and Rousseeuw 2005). Agglomerative algorithms start with one cluster for each title and combine two clusters based on their distance. The distance between clusters is a function of the distance between pairs of titles. We use average linkage as function to compute between-cluster distances, because of its accuracy which is typically higher compared to other linkage functions such as single linkage or complete linkage (Kaufman and Rousseeuw 2005).

We follow existing literature and compute the distance between two *bi*-gram vectors $G_i$ and $G_{i'}$ for two product titles $i$ and $i'$ based on the cosine measure (Cagnina et al. 2014; Li et al. 2008; Papapetrou et al. 2011).

$$dist(i,i') = 1 - \frac{G_i \cdot G_{i'}}{\|G_i\|\|G_{i'}\|} \tag{1}$$

We use *bi*-grams because the possible number of larger *n*-grams is very high which consequently leads to very sparse *n*-gram vectors when $n > 2$. If we consider an alphabet consisting of, for example, 40 symbols (letters, numbers and punctuation marks), there are $40^2 = 1600$ possible *bi*-grams but already $40^3 = 64000$ possible *tri*-grams. The average number of titles that contain a particular *n*-gram hence substantially decreases when *n* increases. Ultimately, the distance between similar but not identical product titles increases which might reduce the clustering accuracy.

The cosine measure is scaled in [0, 1] and it is a measure that is independent of the Euclidean length of the *bi*-gram vectors. Adding further *bi*-grams with a value of 0 to the *bi*-gram vectors of two product titles does not affect the cosine measure between these two titles. This independence of the Euclidean length is important for the identification of robust clusters. *Bi*-gram vectors that differ in their length – as is the case for transaction 1 and 3 in Table 2 – are distant only due to different *bi*-grams, but not due to different lengths.

We get a distance of 0.27 between transaction 1 and 2, a distance of 0.18 between transaction 1 and 3 and a distance of 0.41 between transaction 2 and 3 of our example in Table 2. Spelling errors hence increase the distance between two titles but only moderately. Since the distance between two products is symmetric (i.e., *dist(i,i') = dist(i',i)*), we need to compute $\binom{m}{2}$ distance values to cluster *m* product titles.

The result of an agglomerative clustering for *m* product titles contains *m* to 1 clusters, depending on the maximal distance allowed between two clusters for agglomerating them. The existence of any of the possible clusters is, however, unknown because the true number of distinct products in a data set is unknown. Any definition of a maximally endorsed distance between two clusters is thus without evidence. We therefore use multiscale bootstrap resampling, a technique that has been developed for approximately unbiased tests for the exponential family with unknown expectation parameters (Shimodaira 2004), to estimate the accuracy of any of the possible cluster solutions.

This method conducts an agglomerative clustering for $d = |r|B$ samples of the title × *bi*-gram-matrix. *B* is the number of replications a title × *bi*-gram-matrix is generated with relative sequence length $r_k \in r$ of the bootstrap replicate. Each sample *k* is characterized by a title × *bi*-gram-matrix that consists of all product titles as rows and $r_k$ *bi*-grams randomly drawn with replacement as columns. In the case of $r_k = 0.5$ and 500 identified *bi*-grams, sample *k* is characterized by a matrix with all titles as rows and 250 *bi*-grams randomly drawn with replacement as columns. The frequency of the samples that generate a particular cluster is counted for each scaling factor $r_k \in r$ and used to calculate *p*-values (Shimodaira 2002, 2004). The more frequent a particular cluster occurs in the *B* replications for a particular size $r_k$ and the less this frequency changes when changing $r_k$, the higher is the *p*-value.

To calculate the *p*-values, we first need to count the frequency $f_i$ for a specific cluster *i* in the *B* replications for a particular sample size $r_k$. We get the bootstrap probability $prop_{i,r_k}$ by dividing $f_i$ by *B*. Next, we can estimate the following regression model

$$-\phi^{-1}(prop_{i,r_k}) = X\beta + \epsilon \tag{2}$$

where $\phi^{-1}(\bullet)$ is the inverse density function of a normal distribution and *X* is a two-element vector representing the probability scale for the occurrence of a particular cluster with $\sqrt{r_k}$ as the first and $1/\sqrt{r_k}$ as the second element. Vector *β* consists of the estimates for the two elements of *X* and is estimated by the weighted least squares method. The error term $\epsilon$ is assumed to be multivariate normally distributed with mean vector 0 and non-constant variance-covariance matrix $\sigma^2 W$. The weights *W* are calculated as the inverse of the variance of each sample *k* as follows

$$W^{-1} = Var(k) = \frac{B(1 - prop_{i,r_k})prop_{i,r_k}}{\phi\left(-\phi^1(prop_{i,r_k})\right)^2} \tag{3}$$

The *p*-value is now calculated as

$$p = 1 - \mathcal{N}(\beta_1, \beta_2) \tag{4}$$

where $\mathcal{N}(\bullet)$ refers to the standard normal distribution function.

The *p*-value is normalized in [0, 1] and expresses the probability of a cluster to be existent in different variations of the input data. In other words, the *p*-value expresses the probability that the null hypothesis that a cluster exists in the data cannot be rejected. We are interested in those clusters for which the alternative hypothesis (a cluster is not existent in the data) would be highly unlikely (i.e., clusters that have a *p*-value which surpasses a pre-specified *α*-value).

## 2.5 Clustering Accuracy

We collect data from eBay auctions to illustrate the capabilities of the proposed method AICT and examine its performance. Online auctions have become increasingly important with eBay being the largest online auction platform. With a gross merchandise volume of more than US$ 67.76 billion in 2013, eBay assists millions of transactions a year. Recent research has intensively investigated the determinants of online auction success (Duan 2010; Gilkeson and Reynolds 2003; Shiu and Sun 2014), closing prices (Easley et al. 2010; Shiu and Sun 2014), bidding behavior (Cheema et al. 2012; Srinivasan and Wang 2010), consumer surplus (Bapna et al. 2008) or the effect of attention scarcity on auction outcomes (Hinz et al. 2016).

Auction platforms such as eBay provide its users functionality to sell almost everything – books as well as cars or tickets. Although a seller can define only a few auction parameters like the starting price, the auction length or the number of pictures, auctions are rather heterogeneous due to the products offered to potential buyers. Most auction platforms provide no mechanism for identifying the product offered in an auction. Auction data are hence typically subject to unobserved product heterogeneity.

We focus in the following on the domain of online auctions while our method is not limited to this area. Other data like those from price comparison sites or other platforms can suffer from this problem too, but we use the case of online auctions to illustrate our method and then ultimately to test it empirically.

### 2.5.1 Data

We collect data from eBay auctions for books, music CDs and digital cameras between 22nd and 29th of April 2014. We focus on products that were listed as top sellers between 2009 and 2013 at Amazon.com to ensure that multiple auctions exist for exactly the same product. "CD" is a sub-category of the top-level category "Music" and has no sub-categories. "Books" is a top-level category with 12 sub-categories such as "Children & Adults", "Cookbooks", "Fiction & Literature", "Nonfiction", and "Textbooks & Education". "Digital Camera" is a sub-category of the top-level category "Cameras & Photo" and has no sub-categories.

**Table 3: Descriptive Statistics**

| Variable | All Auctions | | Successful Auctions | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Bidders per Auction | 2.22 | 6.21 | 6.58 | 9.27 |
| End Price (in US$) | – | – | 125.17 | 370.88 |
| Start Price (in US$) | 53.03 | 294.24 | 64.57 | 245.00 |
| Duration (in days) | 5.79 | 1.96 | 5.77 | 2.01 |
| # Pictures per Auction | 7.79 | 9.31 | 9.20 | 10.93 |
| Seller Feedback Score | 17,698.15 | 100,186.50 | 35,392.22 | 165,799.70 |
| # Payment Methods | 1.07 | 0.46 | 1.06 | 0.40 |
| Title Length (in Letters) | 63.51 | 14.38 | 62.79 | 14.37 |
| # Ending During Week | 3,829 | | 1,332 | |
| # Ending During Weekend | 1,682 | | 523 | |
| # Accepting Returns | 2,770 | | 825 | |
| Observations | 5,511 | | 1,855 | |

Our data set includes information about the products offered at eBay (e.g., start price, closing price, and title) and information about the sellers (e.g., seller feedback score). All data were collected using eBay's Developer API. Our data set consists of 5,511 auctions (3,718 book auctions, 1,257 music CD auctions and 536 camera auctions) that were initiated by 3,595 different sellers. We observe 2.22 bids on average per auction and 1,855 auctions resulted in a sale. Table 3 provides a comprehensive summary of the variables used in our study.

### 2.5.2 Analysis

We prepare auction titles as described in Section "Text Preparation" and finally extract a title × *bi*-gram-matrix which constitutes the basis for computing distances between pairs of titles[3]. We then cluster the resulting title distance matrix using an agglomerative method with average linkage function. We use multiscale bootstrap resampling as previously introduced with scaling factors $r$ = {0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4} and generate $B$ = 10,000 replications for each scaling factor.

To assess the quality of the clustering method and identify auctions offering the same product, we also manually add a product identifier to each auction. We evaluate the quality of our clustering algorithm with two common measures: purity and adjusted Rand index (Manning et al. 2009). Purity is the number of correct auction to cluster assignments divided by the number of auctions and ranges from 0 (poor clustering) to 1 (perfect clustering). The purity index overestimates the accuracy if the number of clusters is large compared to the number of items to be clustered. Purity is 1 if each auction gets its own cluster, a typical case of overfitting. We hence compute the adjusted Rand index which calculates for all $\binom{m}{2}$ pairs of auctions one of four possible cluster assignments: i) two similar auctions are assigned to the same cluster (true positive, TP), ii) two dissimilar auctions are assigned to the same cluster (false positive, FP), iii) two dissimilar auctions are assigned to different clusters (true negative, TN) and iv) two similar auctions are assigned to different clusters (false negative, FN). The Rand index is computed as the percentage of decisions that are correct (i.e., the sum of true positives and true negatives divided by the number of auctions). The Rand index is scaled between 0 and 1, but its value when randomly generating clusters is not constant. Hubert and Arabie (1985) introduced the adjusted Rand index (ARI) which can also take negative values and has a value of 0 for a random cluster assignment. Values between 0 and 1 characterize a clustering that is better than a random segmentation. ARI is computed as follows:

$$ARI = \frac{\binom{m}{2}(m_{TP} + m_{TN}) - [(m_{TP} + n_{FP})(m_{TP} + m_{FN}) + (m_{FN} + m_{TN})(m_{FP} + m_{TN})]}{\binom{m}{2}^2 - [(m_{TP} + m_{FP})(m_{TP} + m_{FN}) + (m_{FN} + m_{TN})(m_{FP} + m_{TN})]} \tag{5}$$

[3] We use a very short stop word list containing only one entry for each year between 1999 and 2013, and the following words: album, cd, digipack, dvd, edition, emi, hardback, hardcover, limited, paperback.

A cluster is considered as robust if the null hypothesis "the cluster is existent in the data" cannot be rejected. The likelihood of a cluster to be supported by the data is expressed by the *p*-value computed based on the proposed multiscale bootstrap resampling method. We consider clusters with a *p*-value of at least 0.95 to be robust (i.e., clusters having a probability of not more than 5% to be not existent in the data) and thus allow an $\alpha$-error of 0.05.

### 2.5.3 Results

In total, we collected auction data for 525 different products, a number that we manually determined. We found 10.50 auctions per product on average (*SD* = 25.32). 65 distinct products in our data set (12.38%) were offered in only one auction each. Our proposed cluster algorithm identified 452 clusters with an average of 12.19 (*SD* = 27.47) auctions per cluster. Most clusters (80.31%) consist of auctions that offer exactly the same product. These highly accurate clusters cover 3,164 of our 5,511 observed auctions. All other clusters consist of auctions for at least two different products. Although different, these products are quite similar (e.g., "Sony Cyber-shot DSC-HX50V 20.4 MP Digital Camera Black" and "Sony Cyber-shot DSC-RX100 20.2 MP Digital Camera – Black"). With a purity of 0.81 and an adjusted Rand index of 0.68 we achieved an outstanding clustering accuracy. We hence can use the clusters as additional information in econometric models.

### 2.5.4 Robustness Checks

We conducted several additional tests in order to check the robustness of AICT parameters. Specifically, we tested different linkage methods, different $\alpha$-levels and different *n*-grams. Although average linkage – the linkage method we propose to use – typically performs better than single linkage and complete linkage, there is no evidence that average linkage is always the best option. Complete linkage, for example, will result in very homogeneous clusters in the early agglomeration stages (Kaufman and Rousseeuw 2005). This can improve the accuracy of identifying clusters of very homogeneous auction titles. We also tested different levels of $\alpha$ for the proposed multiscale bootstrap resampling method and different *n*-grams for computing the distances between the auction titles.

**Table 4: Clustering Accuracy across Different Linkage Methods**

| Metric | Average | Single | Complete |
|---|---|---|---|
| Clusters | 452 | 1 | 648 |
| Auctions per Cluster | 12.19 | 5,511 | 8.50 |
| Purity | 0.81 | 0.06 | 0.85 |
| Adjusted Rand Index | 0.68 | 0.00 | 0.51 |

Average linkage clustered auctions with highest accuracy as presented in Table 4. Complete linkage is better in terms of purity compared to average linkage which is due to a higher number of identified clusters. The adjusted Rand index indicates that average linkage is superior to complete linkage. The application of single linkage resulted in a one-cluster solution which is of course not accurate.

We propose using multiscale bootstrap resampling to identify those clusters that are highly accurate. Our previous results are based on an $\alpha$-level of 0.05 which is used as quasi-standard in most empirical studies. With a higher $\alpha$-level AICT will agglomerate more clusters which will finally lead to a solution with a lower number of clusters. Decreasing the $\alpha$-level will hence result in a solution with more clusters. We also computed the cluster accuracy for an $\alpha$-level of 0.10 and 0.01.

**Table 5: Clustering Accuracy across Different $\alpha$-Levels**

| Metric | $\alpha$ = 0.05 | $\alpha$ = 0.10 | $\alpha$ = 0.01 |
|---|---|---|---|
| Clusters | 452 | 2 | 695 |
| Auctions per Cluster | 12.19 | 2,755.50 | 7.93 |
| Purity | 0.81 | 0.06 | 0.83 |
| Adjusted Rand Index | 0.68 | 0.00 | 0.58 |

Table 5 indicates that an $\alpha$-level of 0.10 leads to a solution with only two clusters and is ultimately very inaccurate. With an $\alpha$-level of 0.01, we can identify clusters also highly accurate. However, such a low α-level produces more clusters which finally reduces the adjusted Rand index compared to an $\alpha$-level of 0.05.

We proposed to use distances of *bi*-gram vectors as input for clustering. *Uni*-grams transport not much information because each *uni*-gram is rather likely to occur also in short texts such as product titles. In our data set, we found 40 different *uni*-grams. With an average title length of 63.51, each *uni*-gram occurs on average 1.59 times in each title which makes *uni*-grams not very helpful in separating distinct products based on their titles. *Tri*-grams transport much more information than *bi*-grams because it is rather unlikely that a particular *tri*-gram occurs in a particular title. We found 5,249 distinct *tri*-grams in our data set. Each *tri*-gram hence occurs 0.01 times on average in each title. Each *bi*-grams occurs on average 0.07 times in each title[4]. *Tri*-grams are more sensitive to spelling errors and rephrased titles which results in a larger distance between similar titles.

The number of *n*-grams determines the size of the distance matrix and hence the $\alpha$-level that should be chosen. Larger distance matrices (i.e., more observations) require smaller $\alpha$-levels (Mingfeng et al. 2013). We hence computed the clustering accuracy for *uni*-grams and *tri*-grams with different $\alpha$-levels.

[4] We found 942 different *bi*-grams in our data set.

Table 6 shows that AICT performs best with *bi*-gram vectors as input for the clustering. *Uni*-gram vectors are not very distant to each other which leads to many misclassifications and hence a low clustering accuracy. We found an average cosine distance of 0.165 (*SD* = 0.082) between the *uni*-gram vectors. *Tri*-gram vectors in contrast are very different to each other – we found an average cosine distance of 0.928 (*SD* = 0.099). This high distance is because of the low probability that a particular *tri*-gram occurs in a particular auction title and due to the fact that spelling errors and rephrased titles do have a high impact on the distance of *tri*-gram vectors. Slightly different variants of the same title do not have such a high impact when using *bi*-gram vectors. We found an average distance of 0.826 (*SD* = 0.106) between the *bi*-gram vectors.

**Table 6: Clustering Accuracy across Different *n*-grams**

| Metric | 2-g $\alpha$ = 0.05 | 1-g $\alpha$ = 0.05 | 1-g $\alpha$ = 0.25 | 3-g $\alpha$ = 0.05 | 3-g $\alpha$ = 0.001 |
|---|---|---|---|---|---|
| Clusters | 452 | 708 | 438 | 2 | 836 |
| Auctions per Cluster | 12.19 | 7.78 | 12.58 | 2755.5 | 6.59 |
| Purity | 0.81 | 0.173 | 0.143 | 0.07 | 0.72 |
| Adjusted Rand Index | 0.68 | 0.00 | 0.01 | 0.00 | 0.33 |

In summary, AICT consists of three parameters: i) the distance measure used to identify similar products, ii) the $\alpha$-level, and iii) the type of *n*-grams. Our robustness checks provide implications for calibrating AICT with respect to these parameters. First, average linkage is the preferred method for AICT. Single linkage is not recommendable and complete linkage tends to extract too many clusters. Second, an $\alpha$-level of higher than 0.05 seems to be not meaningful whereas a lower $\alpha$-level produces also very accurate results but tends to also extract too many clusters. Third, the distances between the auction titles should be computed based on *bi*-grams, because *bi*-gram vectors are not as sensitive to spelling errors than *tri*-gram vectors and *bi*-gram vectors of different auction titles and hence different products are more likely to be rather distant to each other than *uni*-gram vectors. We compare AICT to other clustering methods in the next section.

### 2.5.5 Comparison to Other Clustering Methods

The previous section demonstrated an outstanding clustering accuracy for AICT. Our proposed method is based on agglomerative clustering with average linkage. The difference to a simple agglomerative clustering is the multiscale bootstrap resampling, a method that ensures to identify only robust clusters. We computed the clustering accuracy also for a simple agglomerative clustering in order to demonstrate the effect of the multiscale bootstrap resampling method. We found a height of $h = 10$ to produce a similar number of clusters (467) as we have products (525) in our data set.

We furthermore evaluated the accuracy of a distribution-based (EM-clustering; Dempster et al. 1977) and a density-based (DBSCAN; Ester et al. 1996) clustering method, in order to better interpret the accuracy values presented in the previous section. Distribution-based and density-based clustering methods have been found to be very accurate in many applications (Aliguliyev 2009; Chehreghani et al. 2009; Mumtaz et al. 2007; Sarkar et al. 2008). Expectation maximization (EM) and density computation are both not very robust for very sparse data matrices as is the case with *bi*-gram matrices for short texts such as auction titles.

The results in Table 7 reflect the problem of distribution-based (EM-clustering) and density-based (DBSCAN) clustering methods in our case. The accuracy of both is below that of AICT. DBSCAN performed much better than EM-clustering but found only 161 clusters in our data set that consisted of 525 actual clusters.

**Table 7: Clustering Accuracy across Different Clustering Methods**

| Metric | AICT (with $\alpha$ = 0.05) | Agglomerative Clustering (with $h$ = 10) | EM-Clustering | DBSCAN |
|---|---|---|---|---|
| Clusters | 452 | 467 | 9 | 161 |
| Auctions per Cluster | 12.19 | 11.80 | 612.33 | 34.23 |
| Purity | 0.81 | 0.42 | 0.11 | 0.31 |
| Adjusted Rand Index | 0.68 | 0.33 | 0.02 | 0.62 |

Table 7 also shows that our proposed method – AICT – outperformed a simple agglomerative clustering. The difference between both methods is the process of extracting clusters. A simple agglomerative clustering forms clusters based on those objects whose similarity is within a defined threshold (height $h$ in a dendrogram). AICT forms clusters based on multiple simple agglomerative clustering iterations. Clusters that occur in many resampling iterations independent of a similarity threshold are more likely to be really existent and are hence extracted as final clusters in AICT.

AICT shows the highest accuracy compared to a simple agglomerative clustering, EM-clustering and DBSCAN. The results of our robustness check also indicate that AICT might be inferior to other clustering methods if single linkage or a too large $\alpha$-level is chosen. We therefore recommend using the above mentioned average linkage and an $\alpha$-level of 5%.

## 2.6 Enhancing Econometric Analysis

We conduct regression analyses with closing price of our collected eBay auctions (c.f. described in the previous section) as dependent variable. We consider all successful auctions to investigate the closing price of auctions.

### 2.6.1 Models

As baseline, we compute a pure fixed-effects model without integrating the identified clusters (see Model 1 in Table 8). The clusters identified with AICT allow us to i) model cluster-specific intercepts, ii) model cluster-specific slopes, and iii) create new variables such as the number of overlapping auctions. Model 2 adds an idiosyncratic random-intercept effect $\psi_i$ at the cluster level to Model 1. We compute clusters based on all auctions for Model 2. Several of our variables characterize sellers' strategies (e.g., duration, payment methods) and we do not expect product-specific differences across these variables. Starting price, however, is a variable that represents a seller's strategy and a product's value. Model 3 thus extends Model 2 with a cluster-specific random slope for starting price. We finally count the number of overlapping auctions (i.e., auctions – successful as well as unsuccessful – offering the same product as auction *i* at a time span that overlaps with the time span of auction *i*) based on the results of the clusters identified with AICT.

We use the starting price, the duration, the number of pictures, the seller's feedback, the number of payment methods, whether the auction ends at weekend and the ending time as independent variables $X_i$ for auction *i*. These variables have been found to impact an auction's outcome in recent research (Hou and Blodgett 2010; Shiu and Sun 2014; Tan et al. 2010).

**Table 8: Regression Model Specification**

| Model | Specification | Description |
|---|---|---|
| Model 1 | $y_i = \alpha + \beta X_i + \epsilon_i$ | Baseline model |
| Model 2 | $y_i = \alpha + \beta X_i + (1\|\psi_i) + \epsilon_i$ | Baseline + cluster-specific random-intercept |
| Model 3 | $y_i = \alpha + \beta X_i + (1 + StartingPrice_i\|\psi_i) + \epsilon_i$ | Baseline + cluster-specific random-intercept and cluster-specific random-slope for starting price |
| Model 4 | $y_i = \alpha + \beta X_i + \gamma OverlappingAuctions_i + (1 + StartingPrice_i\|\psi_i) + \epsilon_i$ | Baseline + cluster-specific random-intercept, cluster-specific random-slope for starting price and overlapping auctions as additional variable |

According to Lucking-Reiley et al. (2007) we did not include the number of participating bidders in our models because it is endogenously related with the number of bids.

### 2.6.2 Analysis

Each of the models presented in Table 8 is estimated with a restricted maximum likelihood (REML) estimator. We used the natural logarithm of starting price and closing price in our regressions. We found support ($p < 0.05$) to split auctions into two categories according to ending hour: those closed between 11 p.m. and 2 a.m. and those closed between 2 a.m. and 11 p.m.[5] We hence created a

[5] We used conditional inference trees (see Hothorn et al. (2006) for a detailed description of this method) with 10,000 Monte Carlo replications to test for significant splits of ending time.

dummy variable for ending time that is 0 for auctions ending in the daytime (i.e., between 2 a.m. and 11 p.m.) and 1 otherwise. We compared our regression models with Akaike's information criterion (AIC) and the Bayesian information criterion (BIC).

Researchers as well as practitioners often are interested in predicting online transaction outcomes such as closing prices of online auctions. We estimated prediction accuracy of our 4 regression models with 10-fold cross validations. In each replication, we divided our sample in 10 parts and use 9 of them for estimating the regression models and 1 for prediction (hold-out). We ran 10,000 such replications and calculated the average MPE (mean percentage error) as prediction accuracy measure. The MPE is the average of percentage errors by which the predictions differ from observed values and is used as a measure of the bias of a prediction.

### 2.6.3 Results

Table 9 demonstrates that the identified clusters help to significantly improve the goodness of fit of our econometric models. Model 1 has the worst AIC, BIC and $R^2$ and predicts closing prices for new auctions with a very high error. Adding a cluster-specific random intercept significantly improves all fitness measures and hence leads to less erroneous interpretations and predictions as Model 1. We could further improve the econometric analysis by adding a cluster-specific random slope effect for starting price (Model 3) and an additional variable representing the number of overlapping auctions (Model 4).

**Table 9: Regression Results for Closing Price as Dependent Variable**

| Variable | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | Est. | SE | Est. | SE | Est. | SE | Est. | SE |
| Intercept | 1.380*** | 0.148 | 1.417*** | 0.119 | 1.196*** | 0.108 | 1.295*** | 0.106 |
| ln(Start Price) | 0.455*** | 0.016 | 0.257*** | 0.012 | 0.473*** | 0.025 | 0.440*** | 0.024 |
| Duration | -0.011 | 0.017 | 0.006 | 0.012 | 0.014 | 0.010 | -0.006 | 0.010 |
| # Pictures | 0.065*** | 0.003 | 0.024*** | 0.002 | 0.018*** | 0.002 | 0.017*** | 0.002 |
| Seller Feedback/10,000 | 0.021*** | 0.002 | 0.013*** | 0.002 | 0.019*** | 0.002 | 0.018*** | 0.002 |
| # Payment Methods | 0.203* | 0.082 | 0.041 | 0.058 | 0.029 | 0.048 | 0.040 | 0.046 |
| Weekend | -0.023 | 0.072 | -0.048 | 0.049 | -0.045 | 0.040 | -0.029 | 0.039 |
| Accept Returns | -0.226*** | 0.066 | 0.029 | 0.047 | -0.028 | 0.040 | -0.034 | 0.038 |
| Ending Time | -0.179* | 0.077 | -0.037 | 0.052 | -0.048 | 0.044 | -0.042 | 0.042 |
| # Overlapping Auctions | – | – | – | – | – | – | 0.009*** | 0.001 |
| Log-Likelihood | -3,236.05 | | -2,660.27 | | -2,388.01 | | -2,338.72 | |
| AIC | 6,492.10 | | 5,342.54 | | 4,802.02 | | 4,705.45 | |
| BIC | 6,547.36 | | 5,403.32 | | 4,873.85 | | 4,782.81 | |
| $R^2$ | 0.53 | | 0.84 | | 0.90 | | 0.91 | |
| MPE (in %) | 122.86 | | 89.52 | | 48.63 | | 46.02 | |

Table 9 indicates that so far unobserved product heterogeneity significantly affects regression estimates and prediction accuracy. Based on Model 1, we expect the number of payment methods, whether returns are accepted and ending time to have a significant influence on an auction's closing price. This is, however, not indicated by Model 4, the best model in our analysis. Model 1 furthermore overestimates closing prices (as indicated by a positive MPE) by approximately 123% whereas the overestimation reduces to approximately 46% with Model 4.

The number of overlapping auctions is interestingly positively correlated with the auctions' closing price. We found support that more expensive products are offered significantly more frequently in our data set ($p < 0.001$). The number of overlapping auctions is thus rather a proxy for a product's attractiveness and not a proxy for competition.

### 2.6.4 Comparison to Other Clustering Methods

**Table 10: Regression Results for Model 4 and Different Clustering Methods**

| Variable | AICT | | Agglomerative Clustering | | EM-Clustering | | DBSCAN | |
|---|---|---|---|---|---|---|---|---|
| | Est. | SE | Est. | SE | Est. | SE | Est. | SE |
| Intercept | 1.295*** | 0.106 | 2.178*** | 0.156 | 2.053*** | 0.410 | 1.136*** | 0.190 |
| ln(Start Price) | 0.440*** | 0.024 | 0.351*** | 0.024 | 0.357*** | 0.036 | 0.489*** | 0.022 |
| Duration | -0.006 | 0.010 | -0.035** | 0.011 | 0.022 | 0.015 | -0.008 | 0.017 |
| # Pictures | 0.017*** | 0.002 | 0.018*** | 0.002 | 0.039*** | 0.003 | 0.057*** | 0.003 |
| Seller Feedback/10,000 | 0.018*** | 0.002 | 0.021*** | 0.001 | 0.012*** | 0.002 | 0.020*** | 0.002 |
| # Payment Methods | 0.040 | 0.046 | -0.003 | 0.048 | 0.087 | 0.073 | 0.178* | 0.078 |
| Weekend | -0.029 | 0.039 | 0.002 | 0.040 | 0.008 | 0.063 | -0.033 | 0.069 |
| Accept Returns | -0.034 | 0.038 | 0.029 | 0.038 | -0.339*** | 0.058 | -0.193** | 0.065 |
| Ending Time | -0.042 | 0.042 | 0.028 | 0.043 | -0.077 | 0.068 | -0.173* | 0.074 |
| # Overlapping Auctions | 0.009*** | 0.001 | 1.203*** | 0.067 | 0.636*** | 0.106 | -0.099 | 0.114 |
| Log-Likelihood | -2,338.72 | | -2,377.45 | | -3,046.75 | | -3,223.17 | |
| AIC | 4,705.45 | | 4,782.89 | | 6,121.50 | | 6,474.34 | |
| BIC | 4,782.81 | | 4,860.25 | | 6,198.86 | | 6,551.70 | |
| $R^2$ | 0.91 | | 0.90 | | 0.64 | | 0.59 | |
| MPE (in %) | 46.02 | | 47.44 | | 138.03 | | 116.40 | |

We computed Model 4 also with clusters generated with a simple agglomerative clustering, EM-clustering and DBSCAN in order to show the impact of different clustering methods. Table 10 shows that Model 4 has the lowest AIC, BIC and MPE as well as the highest $R^2$ if product clusters are generated with AICT. Although the cluster accuracy of DBSCAN is moderately high, the generated clusters do not enhance econometric models. Model 1 (without clustering) is even better than Model 4 with clusters identified with DBSCAN in terms of AIC and BIC. Model 4 with DBSCAN furthermore indicates that the number of overlapping auctions does not have an effect on an auction's closing price. Accepting returns has been found to have no effect on closing prices in Model 4 with AICT. We, however, found a significant effect in Model 4 with EM-clustering and DBSCAN.

The applied clustering methods hence lead to substantially different cluster solutions that determine the fitness of econometric models and which effects are statistically significant. Poor cluster solutions might not improve econometric models much and can also decrease the fitness of econometric models. Clustering products should help to reduce unobserved product heterogeneity in econometric analyses. EM-clustering and DBSCAN do not help much to reduce unobserved product heterogeneity which would finally result in an improvement of econometric models. AICT and a simple agglomerative clustering both substantially improved our analysis of closing prices. AICT showed the highest accuracy and is hence well suited to overcome the problem of unobserved product heterogeneity in econometric analyses.

## 2.7 Conclusions

This paper highlights the problem of unobserved product heterogeneity in many online data sets. For example, online auctions for identical products can typically not be matched without tremendous manual effort. The same is true for product variants on most online stores. If products are virtually identical, we can assume that they attract the same type of consumers, have the same probability to be purchased or reviewed. In other words, the process that generates economically relevant data such as the number of purchases, the number of reviews, the average product rating or the closing price in an auction is similar if not identical across virtually identical products.

We propose the Ambiguous Identifier Clustering Technique (AICT) that clusters online offers based on their titles to identify offers for (quasi) identical products and use this information to control for product-specific heterogeneity. We apply AICT to data from eBay auctions and demonstrate that i) AICT can identify auctions for identical products with a high accuracy, ii) clustering auctions helps to reduce unobserved product heterogeneity and improves the goodness of the econometric models for explaining an auction's closing price, iii) AICT is better suited for reducing product heterogeneity than other methods such as agglomerative clustering, EM-clustering and DBSCAN, and iv) we can define further determinants of closing price based on product clusters. With multiscale bootstrap resampling, AICT identifies only those product clusters having a high probability to exist in the data.

Researchers and practitioners hence benefit from AICT if they face the problem of missing unique or ambiguous product identifiers. This is, for example, the case if online auctions for identical products (Easley et al. 2010) or entries in price comparison engines should be matched (Kocas 2002). This matching can be furthermore used to form clusters of online offer data or to generate new variables such as the number of overlapping online auctions.

The implications for managers and consumers are twofold. First, managers can use AICT to analyze the performance of their products in the electronic secondary market (e.g., eBay auctions). Previous research has presented conditions under which firms should operate in such markets to improve their profit (Ghose et al. 2005). AICT helps to evaluate if these conditions are valid for a particular product or product category. And second, AICT can help to improve online-market transparency. Online consumers intensively contact price comparison engines to get information about products and their prices. However, prices listed by these engines are still rather nontransparent due to varying shipping costs, product variants and product supplements that are listed when searching for a particular product[6]. Our method could help to separate products from product bundles or supplements and ultimately support consumers to identify the best market price of a particular product. AICT could further help managers to easier observe their competitors and improve their own pricing strategies.

AICT identifies clusters of products having very similar titles, but it does not incorporate other information such as product quality. Identifying a product's quality based on product descriptions and product pictures might further improve econometric analysis as presented in this paper. Incorporating additional information in AICT hence provides an interesting avenue for future research.

[6] PriceGrabber.com, for example, lists more than 120 products when searching for "canon powershot sx50 hs". Some of these products are colored variants of the camera, some are bundles including the camera and some are supplements such as bags or tripods.

# 3 Measuring Consumers' Willingness to Pay with Utility-Based Recommendation Systems

**Title:** Measuring Consumers' Willingness to Pay with Utility-Based Recommendation Systems

**Authors:** Scholz, Michael; University of Passau; Germany
Dorner, Verena; University of Passau; Germany
Franz, Markus; TU Darmstadt; Germany
Hinz, Oliver; TU Darmstadt; Germany

**Published:** Decision Support Systems (2015), 72, 60-71

**Abstract**

Our paper addresses two gaps in research on recommendation systems: first, leveraging them to predict consumers' willingness to pay; second, estimating non-linear utility functions – which are generally held to provide better approximations of consumers' preference structures than linear functions – at a reasonable level of cognitive consumer effort. We develop an approach to simultaneously estimate exponential utility functions and willingness to pay at a low level of cognitive consumer effort. The empirical evaluation of our new recommendation system's utility and willingness to pay estimates with the estimates of a system based on linear utility functions indicates that exponential utility functions are better suited for predicting optimal recommendation ranks for products. Linear utility functions perform better in estimating consumers' willingness to pay. Based on our experimental data set, we show how retailers can use these willingness to pay estimates for profit-maximizing pricing decisions.

**Keywords:** Willingness to pay, utility-based recommendation system, utility function, e-commerce

## 3.1 Introduction

Consumers' willingness to pay (WTP) and their utilities for different products are indispensable inputs for many prediction and optimization models that support core business decisions and processes. They help decision makers define efficient pricing strategies (e.g. Schlereth and Skiera 2012; Wu et al. 2014), estimate market share (e.g. Jedidi et al. 1996), determine the optimal pace for product updates (e.g. Druehl et al. 2009), realign service operations (e.g. Coltman et al. 2010), or identify optimal product assortments (e.g. Rusmevichientong et al. 2010).

These decision models will, of course, yield reliable results only if valid estimates for consumer preferences and WTP are available, which is difficult for two reasons. First, the cost of estimating current utility values and WTP is high. Eliciting this information empirically is expensive because valid estimates for WTP can often be obtained only at the time of purchase and under the prevailing marketing mix conditions (Wertenbroch and Skiera 2002). This problem is exacerbated by the fact that consumer preferences change over time. Hence, preference elicitation must be repeated at regular intervals to monitor the estimates' validity. Second, consumers are typically not willing to expend a lot of effort and time (repeatedly, no less) for specifying their preferences (De Bruyn et al. 2008), which reduces the reliability of their inputs and, in consequence, reduces recommendation accuracy (Huang 2011).

These two issues can be solved with utility-based recommendation systems. Recommendation systems are very frequently employed as consumer decision support systems in online retailing, and most consumers use them regularly. That consumers are intrinsically motivated to use them, and to use them repeatedly, makes it possible to recognize structural changes in consumer preferences and WTP immediately, and to obtain utility estimates for new products almost at once. Despite these advantages, recommendation systems data are seldom suggested as inputs for marketing and management decision models (Denguir-Rekik et al. 2009) or as inputs for WTP estimation.

We contribute to recent research in operations research, information systems and marketing by proposing a utility-based recommendation system which measures consumer preferences and WTP 1) reliably and 2) at low costs for companies and 3) low cognitive effort for consumers. Specifically, we develop a new low-effort approach, based on Butler et al.'s (2001) utility exchange approach. This approach proposes that utility in one attribute can be exchanged for utility in another. We propose to exchange utility in price for the utility of entire products to estimate consumers' WTP.

We extend utility-based recommendation systems to estimate consumers' individual WTP for zero-switch utility functions (Abbas and Bell 2012): linear and exponential utility functions. In addition, our research sheds light on the question which of the two most popular single-attribute utility (SAU) functions from operations and marketing research – exponential (Harvey 1981; Van Ittersum and Pennings 2012) or linear functions (Green et al. 2001; Scholz et al. 2010) – are better suited for utility-based recommendation systems.

The remainder of our paper is organized as follows. The next section briefly introduces the utility and WTP model we are proposing. Section 3.3 shows how this model can be integrated into recommendation systems. Section 3.4 describes our design and setting for testing this model in a laboratory experiment. Section 3.5 presents the results of this experiment. Section 3.6 provides a closer look at how this information can be used to optimize pricing decisions and discusses economic implications. Section 3.7 concludes the paper with an outlook on future research.

## 3.2 Modeling Utility-Based Recommendation Systems

Utility-based recommendation systems are a rather novel feature of online retailing. Usually, recommendations are generated with collaborative filtering or content-based techniques (Xiao and Benbasat 2007). Content-based techniques recommend items similar to those a consumer has bought in the past. Collaborative filtering techniques recommend items to a consumer based on purchase decisions by other consumers who have similar tastes and preferences. But though ubiquitous, both techniques frequently produce low quality recommendations, which is due to three major issues (Ansari et al. 2000). First and most important is the cold-start problem (Kim et al. 2011). Content-based and collaborative filtering techniques cannot provide recommendations unless multiple-item purchasing profiles for a number of consumers, or at least for the consumer currently using the system, are available. Second, preference estimates based on purchasing profiles are inaccurate when, as is often the case, these profiles contain products purchased as gifts for or on behalf of other consumers. Third, purchasing profiles are historical data, revealing past but not necessarily current preferences.

None of these issues arises in utility-based recommendation systems. Utility-based recommendation systems compute consumers' individual utilities for all products of a given category (Huang 2011; Scholz and Dorner 2012). Individual multiple-attribute utility (MAU) functions are constructed based on explicit preference statements provided by the consumers (i.e. ratings of attributes or products). This preference information can then be used to obtain estimates for consumers' individual WTP (Gensler et al. 2012). However, the drawback of utility-based recommendation systems is the fact that consumers must actively provide input before a recommendation is possible.

Designing a utility-based recommendation system for estimating consumers' WTP poses three major challenges. First, recommendation systems ought to be flexible with regard to the shape of the estimated utility function. Utility functions for different attributes and consumers may be linear, convex or concave (Van Ittersum and Pennings 2012), and the shape of utility functions must be defined prior to estimating and generating recommendations. In other words, only if all attribute levels' utility values are known (in theory, consumers need to specify utility values for each attribute level for each attribute), or at least predictable, can the system generate recommendations. Estimating utility values for only a few attribute levels and manually selecting an appropriate utility function based on these utility values – as done in many marketing studies (e.g. DeSarbo et al. 1995; Green et al. 2001) – is not feasible for a recommendation system. Flexible recommendation systems thus require higher consumer effort, whose level depends on the methods used for measuring attribute weights (Section 3.2.1), SAU functions (Section 3.2.2) and WTP (Section 3.2.4). This raises the second challenge for utility-based recommendation system design: system usage ought to be easy and require little effort. Consumers are typically unwilling to invest much time and cognitive effort in eliciting preferences or utility functions (De Bruyn et al. 2008). High-effort systems have a negative influence both on consumers' willingness to use these systems and on the reliability of consumers' inputs. And third, consumers' WTP must be computed for all conceivable products.

In the next sections, we show how these challenges can be met by adapting the utility model and system interaction design, and propose a new model for a utility-based recommendation system.

### 3.2.1 Utility Estimation

Utility estimation methods have been applied to a wide range of issues including vendor selection (Yu et al. 2012), the evaluation of knowledge portal development tools (Kreng and Wu 2007), the prediction of market shares (Gensch and Soofi 1992), WTP estimation (Gensler et al. 2012) and product recommendation (Huang 2011). But despite their importance and the large number of different approaches to estimating utilities, no model has been proven superior so far (e.g. Corner and Buchanan 1997; Moore 2004; Pöyhönen and Hämäläinen 2001). Table 11 provides an overview of utility estimation methods that have been put forward as core methods for utility-based recommendation systems.

**Table 11: Utility Estimation Methods in Utility-Based Recommendation Systems**

| Utility Estimation | Input | Effort | Accuracy | References |
|---|---|---|---|---|
| DR | Attribute Weights | Low | High | Bottomley et al. (2000); Cao and Li (2007); Theetranont et al. (2007) |
| SMARTER | Attribute Weights | High | High | Huang (2011) |
| RBFN | Attribute Weights | Low | Low | Huang (2011) |
| RBCA | Product Ratings | High | High | Scholz and Dorner (2012) |
| CBCA | Product Choices | High | High | Pfeiffer and Scholz (2013) |

The majority of commonly used utility estimation methods have severe shortcomings with respect to at least one criterion, accuracy or consumer effort. Methods such as SMARTER (Huang 2011) or rating-based conjoint analysis (RBCA; Scholz and Dorner 2012) generate highly accurate estimates, but they are cognitively very expensive (Pfeiffer and Scholz 2013). Radial basis function networks (RBFN) are easier to use, but less accurate than SMARTER (Huang 2011). Using product ratings for estimating individual utility functions poses much the same problems as using content-based and collaborative filtering systems: consumers rate only a fraction of available products, and these ratings become obsolete over time (Ansari et al. 2000).

One particularly simple method, direct rating (DR), has been shown to generate surprisingly accurate attribute weights (Bottomley et al. 2000), outperforming more complex approaches like AHP and requiring much less consumer effort (Häubl and Trifts 2000; Pöyhönen and Hämäläinen 2001). Attribute weights are, however, only a part of utility estimates and need to be combined with single-attribute utility functions for computing overall product utilities. Recent research often assumes linear single-attribute utility functions with a utility of 0 for the worst and a utility of 1 for the best level of a particular attribute (Cao and Li 2007; Theetranont et al. 2007). These linear functions can be estimated without incurring user effort and are easily combined with attribute weights elicited with direct rating, direct ranking or AHP (see Section 3.2.2).

In the following sections, we show how reliable WTP and utility values can be estimated based on consumer inputs elicited with DR, which are then combined in multi-attribute utility functions.

### 3.2.2 Multi-Attribute Utility Model

Multi-attribute utility theory (MAUT) assumes that products are bundles of attributes and that consumers evaluate products by evaluating their attributes. Each attribute $i$ that affects the purchase decision is described by a weighted single-attribute utility (SAU) function $w_i u_i(x_i)$, $x_i$ being the level of attribute $i$ and $w_i$ the weight for attribute $i$. Linear SAU functions are easier to estimate than non-linear SAU functions, but the latter are held to be a better approximation of consumers' preferences (Van Ittersum and Pennings 2012) and to generate more accurate estimates.

For linear SAU functions, estimating the scaling parameters $a_i$ and $b_i$

$$u_i(x_i) = a_i + b_i x_i \tag{6}$$

does not pose a large problem. We assign a utility value of 0 to the worst level $x_i^{worst}$ of attribute $i$ and a utility value of 1 to the best level $x_i^{best}$, assuming that $x_i$ is normalized in [0; 1] (Butler et al. 2008). The scaling parameters are now given as $a_i$ = 0 and $b_i$ = 1.

For modeling non-linear SAU functions, we choose the more flexible exponential shape (Butler et al. 2001; Harvey 1981; Van Ittersum and Pennings 2012).

$$u_i(x_i) = a_i - b_i e^{c_i x_i} \tag{7}$$

If empirical SAU functions are non-linear (Van Ittersum and Pennings 2012), exponential SAU functions provide more accurate utility estimates. By introducing the scaling constant $c_i$, SAU functions can now be modeled as linear, convex or concave. Despite its great flexibility, however, Equation (7) is an unpopular choice for practical applications of utility-based recommendation systems (Butler et al. 2008; Scholz and Dorner 2012). Estimating $c_i$ requires additional consumer input and thus effort, which increases the likelihood of consumers terminating the recommendation process prematurely and of obtaining less reliable preference and WTP estimates.

We propose a new effort-minimizing approach for estimating $c_i$ which requires consumers merely to specify their utility for the average attribute level $x_i^{average}$. We suggest that the additional effort will be overcompensated by better recommendations due to more accurate utility estimates.

Given that $x_i^{worst}$ = 0, $x_i^{best}$ = 1, $u_i(x_i^{worst})$ = 0, and $u_i(x_i^{best})$ = 1, we can then estimate $c_i$ (Equations (8) and (9)).

$$c_i = ln\left(\left(\frac{1-\sqrt{1-4u_i(x_i^{average})(1-u_i(x_i^{average}))}}{2u_i(x_i^{average})}\right)^2\right) \qquad \text{if} \quad u_i(x_i^{average}) > 0.5$$

$$c_i = ln\left(\left(\frac{1+\sqrt{1-4u_i(x_i^{average})(1-u_i(x_i^{average}))}}{2u_i(x_i^{average})}\right)^2\right) \qquad \text{if} \quad u_i(x_i^{average}) \leq 0.5$$

$$c_i = 0 \quad \text{if} \quad u_i(x_i^{average}) = 0.5 \tag{8}$$

Parameters $a_i$ and $b_i$ are now given by

$$a_i = \frac{1}{1 - e^{c_i}} \tag{9}$$

$$b_i = a_i.$$

Depending on the level of $u_i(x_i^{average})$, the exponential SAU function is either convex ($u_i(x_i^{average})$ < 0.5), concave ($u_i(x_i^{average})$ > 0.5) or linear ($u_i(x_i^{average})$ = 0.5). Compared to the specification effort for linear SAU functions, exponential SAU functions thus require only one additional step per function (i.e. per attribute) in which the consumer states her utility for $x_i^{average}$.

To compute the overall product utility $U(X_k)$, SAU functions are aggregated in a MAU function (Huang 2011). The additive MAU function is

$$U(X_k) = \sum_{i=1}^{n} w_i u_i(x_i) \tag{10}$$

where 0 ≤ $w_i$ ≤ 1, and $\sum_{i=1}^{n} w_i$ = 1.

Specifying additive functions requires the lowest consumer effort, but they rest on two strict assumptions: mutual utility independence for each pair of attributes and additive independence among all attributes. Attributes are mutually utility independent if and only if each subset of attribute levels $x$ = ($x_1$, $x_2$, …, $x_n$) is independent of its complementary subset in terms of its utility. Attributes are additively independent if the interaction between two or more attributes has no effect on alternatives' utilities. Clearly, these assumptions hold in few decision scenarios. But additive models have been shown to be robust even if additive independence does not hold (Butler et al. 1997; Dawes 1979). Combining $I$ exponential SAU functions in an additive MAU function requires consumer input on $I$ parameters for $w_i$ and $I$ parameters for $u_i\left(x_i^{average}\right)$; combining linear SAU functions additively requires input on $I$ parameters. In the following sections, we suggest a new approach for estimating WTP with a MAUT-based approach.

### 3.2.3 Willingness to Pay Estimation

WTP estimation methods suitable for practical application must meet two criteria: one, that they yield valid measurements for products and product categories, and two, that they require very little consumer input. Empirical studies usually apply WTP estimation methods that elicit WTP i) directly, ii) based on auction bids for a product or iii) from utility functions (e.g. conjoint analysis).

For direct WTP elicitation, consumers are simply asked to state their WTP, for instance by way of an open question. Direct WTP measurements are surprisingly accurate, considering the low level of consumer effort required (Miller et al. 2011). However, this only holds true for one or few products. In cases where WTP for a greater number of products or even an entire product category needs to be elicited, direct WTP elicitation is quite taxing for the consumer. Also, consumers may not like to openly state their WTP if they feel that companies use this information for price discrimination.

In auctions like the Vickrey auction, first price auction, or the BDM mechanism, the dominant bidding strategy is to exactly bid the WTP. It is therefore not surprising that auctions measure consumers' WTP with high accuracy (Barrot et al. 2010). Drawbacks of auctions are, first, that auction fever (Jones 2011), market competition (Chan et al. 2007) and other factors can bias WTP measurement. Second, WTP estimates based on auctions are limited to the products used in the auction. It is not possible to infer the WTP for other products in the same category. As with direct WTP measurement, one data point must be elicited for each (new) product.

Methods which estimate WTP from utility functions do not share this drawback. WTP for all products in a category for which a set of utility functions is valid can be estimated without further consumer input for each product. Some utility-based methods, like direct rating or self-explication approaches (e.g. Netzer and Srinivasan 2011), require direct input on SAU functions. Others, like rating-based or choice-based conjoint analysis, measure SAU functions by decomposing consumer evaluations of entire products (e.g. Gensler et al. 2012). Direct utility elicitation methods are cognitively less demanding than compensatory approaches (De Bruyn et al. 2008; Pfeiffer and Scholz 2013).

Yet another data source for WTP estimation is market data: scanner data, for instance, are often used for estimating demand curves (Kamakura and Russell 1993; Leeflang and Wittink 1992). Their major advantage is that real purchases are used. Their major drawback is their aggregation level, which makes them unsuitable for estimating individual WTP. Perhaps even more importantly, data for estimating demand curves must contain a sufficiently high level of price variation which, in practice, it may not always be feasible to induce.

**Table 12: Comparison of WTP Estimation Methods**

| Method | Individual Measurement | Revealed Preferences | WTP Prediction for Similar Products | Repeated Measurement | Consumer Effort | References |
|---|---|---|---|---|---|---|
| Direct WTP Elicitation | Yes | No | No | No* | Low | Backhaus et al. (2005); Miller et al. (2011); Voelckner (2006) |
| Auctions | Yes | Yes | No | Yes | Low | Barrot et al. (2010); Chan et al. (2007); Voelckner (2006) |
| Direct utility Elicitation | Yes | No | Yes | No* | Low | Louviere and Islam (2008); Park et al. (2008) |
| Conjoint Analyses | Yes | No | Yes | No* | High | Jedidi and Zhang (2002); Gensler et al. (2012); Voelckner (2006) |
| Market Data | No | Yes | No | Yes | Low | Leeflang and Wittink (1992); Kamakura and Russell (1993) |
| UBRS | Yes | Yes | Yes | Yes | Low | This study |

*Note: * Repeated measurements are possible, but usually expensive.*

Comparing methods for WTP measurement (Table 12) shows that most methods are designed for single-point measurements. Repeated measurements with conjoint experiments, auctions, or surveys lead to high effort on part of the consumer and high costs on part of the executing company.

We propose to solve this problem by integrating direct utility elicitation (see Section 3.2) in recommendation systems. Direct utility elicitation methods require little consumer effort (Table 12), but, as stated by (Park et al. 2008), they need to be adapted for WTP measurement. We introduce an extension to direct utility elicitation, which is based on the utility exchange approach (Butler et al. 2001), in the next section.

### 3.2.4 Willingness to Pay Model

A consumer's WTP for a specific product $X_k$ is defined as the price at which she is indifferent between buying and not buying $X_k$ (Moorthy et al. 1997). Consumer $j$'s utility function for product $X_k$ is given by $U_j(X_k, y_j)$, where $y_j$ refers to the composite product consisting of all products other than $X_k$ which consumer $j$ purchases. She uses her budget $m_j$ for purchasing both the composite product $y$ and zero or one units of product $X_k$. Consumer $j$'s budget equation is

$$m_j = p(X_k) + p(y_j) \tag{11}$$

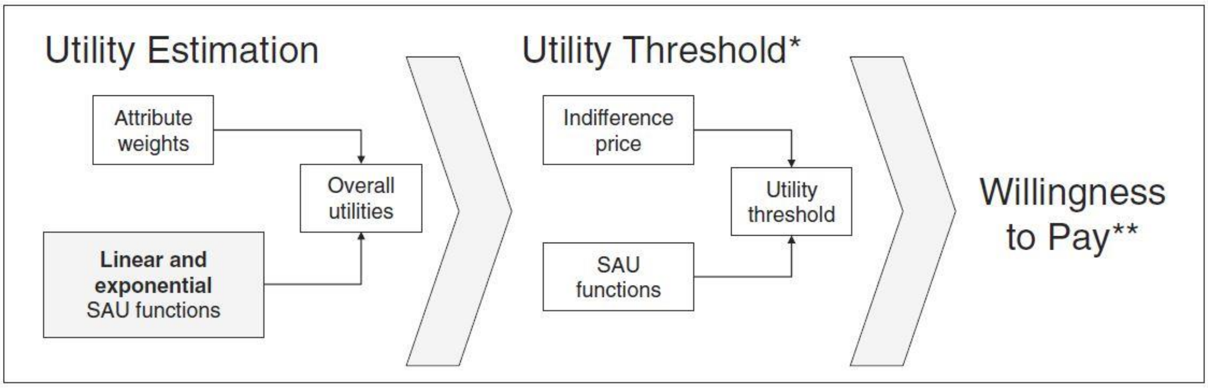
where $p(X_k)$ is the price of product $X_k$ and $p(y_j)$ is the price of the composite product $y_j$. Her utility function is $U_j(X_k, (m_j - p(X_k))/p(y_j))$ if $X_k$ is attractive enough to be purchased. This is only the case if the utility of $X_k$ exceeds the so-called utility threshold $\tau_j$ (Scholz and Dorner 2012). All products whose utilities lie below the threshold $\tau_j$ are not attractive enough to consumer $j$ for her to seriously consider purchasing them. In that case, her utility function equals $U_j(0, m_j/p(y_j))$ and her WTP equals[7]

$$U_j\left(X_k, \frac{m_j - WTP_j(X_k)}{p(y_j)}\right) - \tau_j \equiv 0. \tag{12}$$

Once we know consumer $j$'s utility threshold $\tau_j$, we can predict her purchase decision for any product $X_k$. Prediction accuracy thus depends on obtaining reliable estimates for consumers' utility thresholds $\tau_j$. We propose to identify $\tau_j$ and $WTP_j(X_k)$ for product $X_k$ by combining MAUT with Butler et al.'s (2001) utility exchange approach (Figure 3).

At a given level of WTP for product $X_k$, the utility threshold $\tau_j$ expresses the utility of product $X_k$ with a price of $WTP_j(X_k)$. Consumer $j$ is indifferent between purchasing and not purchasing the product. We now need to find, for all other attractive products, the prices which would render their utilities equal to $\tau_j$. In order to do that, we use the utility exchange approach as proposed by Butler et al. (2001). This approach is based on the idea of even swaps (Hammond et al. 1998; Mustajoki and Hämäläinen 2007).

[7] See Jedidi and Zhang (2002) for a similar definition of WTP.




*Note: * adapted from Butler et al. (2001); ** Jedidi and Zhang (2002); Keeney and Raiffa (1993)*

**Figure 3: Estimating WTP from Recommendation Data**

Consider a consumer whose utility threshold is $\tau_j$ = 9 and for whom $U(X_k)$ = 10. If we were to downgrade one of the attributes of $X_k$ just so much that the consumer's utility for $X_{k'}$ decreased by one utility unit to 9 units, she would still perceive the product to be attractive. Now assume that the attribute in question is the price. A reduction in utility corresponds to an increase in price, and the new increased price equals the consumer's WTP for $X_{k'}$. Thus one utility unit has been "exchanged" for price units, at the consumer's individual "exchange rate".

In Equations (13) and (14), price is assumed to be attribute $i$ = 1 and the underlying MAU function to be additive. In the case of exponential SAU functions, WTP for product $X_{k'}$ is given by

$$WTP_j(X_{k'}) = ln\left(\frac{a_1 - \left(\frac{\left(\tau_j - \sum_{i=2}^{n} w_i u_i(x_i)\right)}{w_1}\right)}{b_1}\right)/c_1. \tag{13}$$

For linear SAU functions, WTP for product $X_k$ can be computed as:

$$WTP_j(X_{k'}) = \frac{\frac{\tau_j - \sum_{i=2}^{n} w_i u_i(x_i)}{w_1} - a_1}{b_1}. \tag{14}$$

The next section describes the design of a utility-based recommendation system implementing our approach (Figure 3), in particular how attribute weights $w_i$, average attribute level utility $u_i\left(x_i^{average}\right)$ and WTP for any product $X_k$ are elicited and estimated.

## 3.3 Recommendation System Design

In designing our recommendation system, we must balance consumer effort and recommendation accuracy. The level of consumer effort depends on the methods used for measuring attribute weights, SAU functions and WTP. We use DR for eliciting attribute weights $w_i$, a method both simple and accurate (Section 3.2.1; Pöyhönen and Hämäläinen 2001).

For estimating single attribute utilities, we use linear and exponential SAU functions (Section 3.2.2). Exponential SAU functions can generate more accurate recommendations than linear functions (Section 3.2) but require additional information about the average attribute levels' utilities (Equation (8)). We elicit this information by displaying the lower and upper bounds, $u_i(x_i^{best})$ = 1 and $u_i(x_i^{worst})$ = 0, of the attribute level intervals to the consumer and asking her to state $u_i\left(x_i^{average}\right)$ (Figure 4; Netzer and Srinivasan 2011; Scholz et al. 2010).


**Figure 4: Elicitation of Parameter $c_i$ of Exponential SAU Functions**

Finally, we need to elicit consumer $j$'s utility threshold $\tau_j$ to estimate WTP. Butler et al. (2001) suggest asking her: "If products $A$ and $B$ are identical on all criteria except price and product $A$ costs $P$, what is the lowest price of $B$ that could make you feel that $B$ is significantly better than $A$?".

Unfortunately, this approach has shown low empirical validity (Scholz and Dorner 2012). For one, thinking in utility units and thinking in price units appear to be two cognitively different tasks. For another, rational consumers will feel that $B$ is significantly better (i.e. they will prefer it over $A$) even if the price difference is extremely small. Extremely small price differences, however, lead to an overestimation of the utility threshold $\tau_j$ (Scholz and Dorner 2012). WTP elicitation can be a sensitive issue when stated WTP is used to predict WTP for other products. We avoid this issue by asking consumers to state their individual WTP for a hypothetical product $X_k$. Because our approach requires WTP for $X_k$ to be greater than zero, we define $X_k$ as a modified version of the best expected product. Specifically, we take the best expected product and decrease the level of its least important attribute (only if $w_i$ > 0) to $x_i^{worst}$ to generate the hypothetical product $X_k$, which ensures that the probability for $WTP_j(X_k)$ > 0 is high. After consumer $j$ stated her $WTP_j(X_k)$, we can compute her utility threshold $\tau_j$. Applying Equations (13) or (14) yields the WTP estimates for all real products.

Consumers typically do not know exactly how much they are willing to pay for a product — they seldom possess perfect information about product quality (March 1978). It is easier for consumers to specify WTP as a range than a price point, and results are generally more valid (Schlereth et al. 2012; Wang et al. 2007). WTP is therefore elicited as a range over three price points (Wang et al. 2007):

1. Floor WTP: The highest price at which a consumer will (still) definitely purchase the product, i.e. 100% purchase probability.
2. Indifference WTP: The price at which a consumer is indifferent between purchasing and not purchasing the product, i.e. 50% purchase probability.
3. Ceiling WTP: The lowest price at which a consumer will definitely not purchase the product (anymore), i.e. 0% purchase probability.

WTP is elicited directly with open questions because they are of low cognitive complexity and measure consumers' WTP as validly as more complex methods like choice-based conjoint analysis (Miller et al. 2011).

## 3.4 Empirical Investigation

We conducted two laboratory experiments with between-subject designs. The main experiment was used to compare the performance of linear and exponential SAU functions, specifically with regard to the accuracy of utility function and WTP estimates. We implemented one recommendation system for each treatment (linear and exponential). Both recommendation systems (treatments) operated on a product data base of 162 camera models by 16 manufacturers, which reflected the actual market situation at the time of the experiment quite well. Each camera was described by eight attributes: photo resolution, optical zoom, camera size, display resolution, video resolution, range of settings options, ISO sensitivity and price.

We chose a search good because estimating product utilities requires attribute levels to be operationalized objectively and reliably (Butler et al. 2008), which is impossible for experience goods. Their attribute levels can be determined, subjectively, only after purchase and use (Mudambi and Schuff 2010; Nelson 1970). Digital cameras are a very popular category of search goods, which we could assume all our participants to be reasonably familiar with (Wang and Benbasat 2007). To ensure a minimal level of product expertise, we provided all participants with information on each of the eight camera attributes.

Since one of our goals was to measure participants' WTP, we did not display product pictures or prices to participants when measuring WTP, thus excluding two potential sources of bias. Product pictures can affect consumers' WTP (Dewally and Ederington 2006), and prices can serve as an anchor for consumers' WTP (Bohm et al. 1997).

We conducted a supplementary experiment for checking three important assumptions inherent in our approach and empirical setting, namely that SAU functions were monotonous, that participants in our subject pool were indeed sufficiently familiar with the product category to state their preferences, and that there existed no systematic gender related differences in expertise between participants in our subject pool, as have sometimes been found for technical products (Meeds 2004). Setting and instructions remained unchanged from the main experiment, but the SAU functions in this experiment were "free" in the sense that no prior shape (linear or exponential or otherwise) was assumed, and additional questions on participant familiarity with digital cameras were added to the post-experimental survey (see Section 3.4.1).

The following subsections describe the experimental procedure and results.

### 3.4.1 Experimental Procedure

All recommendation systems (treatments) implement the screening and evaluation stages of the purchasing process as described by Payne (1976) and Hauser and Wernerfelt (1990) (Figure 5).

<table>
<tr><th></th><th></th><th>Task A</th><th>Task B</th><th>Task C</th><th>Task D</th></tr>
<tr><td></td><td></td><td>Screening</td><td>Evaluation</td><td>Utility exchange</td><td>Recommendation set evaluation</td></tr>
<tr><td rowspan="2">Main experiment</td><td>Linear SAU functions</td><td rowspan="3">Define aspiration levels</td><td>Define attribute weights …</td><td rowspan="3">Specify WTP for modified version of best recommended product</td><td rowspan="3">Rate top ten recommended products and specify WTPs</td></tr>
<tr><td>Exponential SAU functions</td><td>… and rate average attribute levels</td></tr>
<tr><td></td><td>Free SAU functions</td><td>… and rate 5 levels for each attribute</td></tr>
</table>

**Figure 5: Research Procedure**

In the screening task (Task A in Figure 5), designed to remove unattractive cameras from the recommendation set, participants specified aspiration levels for all attributes. In the evaluation task (Task B in Figure 5), participants indicated attribute weights $w_i$ by directly allocating between one and eleven points to each attribute. Participants in the exponential treatment additionally stated for each attribute how attractive they would consider a camera equipped with an average level ($x_i^{average}$) of that particular attribute.[8] Participants in the supplementary experiment were asked the same questions for 5 levels for each attribute so that we could test the shape of the SAU functions and estimate

[8] This information was used to estimate the exponential SAU functions as defined in Equations (8) and (9).

more flexible SAU functions.[9] The information from tasks A and B was used to compute overall product utilities.

In the third task (Task C in Figure 5), participants specified their WTP for a modified version (Section 3.3) of the best recommended product. This information was then used to estimate participants' utility thresholds $\tau_j$ (Section 3.2.4).

In the fourth task (Task D in Figure 5), recommendation set evaluation, we displayed the top 10 recommended products in descending order of their utilities. Participants rated these products on an eleven-point scale and specified their WTP for each product in three open questions (Section 3.3).We used task D for validating both utility and WTP accuracy.[10]

After both experiments, participants answered a questionnaire on basic demographic information; their perceptions of the system (perceived ease of use, perceived satisfaction with the system, perceived usefulness, perceived reuse intention and perceived task difficulty); and their e-shop usage. Thus we were able to control for potential effects of participant and system characteristics on utility and WTP estimation accuracy.

After the supplementary experiment, participants answered an additional expertise questionnaire with two comprehension questions for each attribute. We examined both "theoretical" knowledge, i.e. how attribute functionality is defined, and "applied" knowledge, i.e. how to use different functions in a real setting.

### 3.4.2 Pretest

We tested the treatments with 13 undergraduate students in think aloud protocols (Gena and Weibelzahl 2007), which are particularly well suited for exploring user-system interactions and user decision making (Isaacs and Senge 1992). After completing the experiment, participants answered a small questionnaire on treatment comprehensibility and perceived effort of system use. We conducted four rounds of pretests with three or four participants each, adjusting the system according to the feedback from each round until no further suggestions for improvement were made.

[9] The effort required for this approach is much greater (see Section 3.5.1), which makes it impractical to use in real-life recommendation systems. We used it only to check our assumptions (Section 3.4).

[10] While task C – eliciting WTP for a hypothetical product – is part of our proposed approach, task D – asking for WTP for real products – is only part of the experimental evaluation of our approach. In a real-life setting, it is doubtful whether consumers would openly specify their WTP for real products (see Section 3.3). The experimental setting, however, is both anonymous and does not culminate in a real purchase. Thus participants have no reason to deviate from their true WTP.

### 3.4.3 Samples

77 undergraduate and graduate students of a German university took part in the supplementary experiment. 55 were female and 22 were male. In this experiment, we specifically tested for gender-related differences in camera expertise. More precisely, we compared, for each attribute, the percentages of male and female participants who were able to answer the two expertise questions correctly. Participants had to select the correct answer out of four given answers. There were no significant differences between male and female participants' comprehension of camera attributes with the exception of ISO sensitivity. Although a greater percentage of female than male participants knew what ISO sensitivity is ($F$ = 96.4% and $M$ = 81.8%, $p$ = 0.034), male participants had a better idea of how to apply this knowledge in practice ($F$ = 52.7% and $M$ = 77.2%, $p$ = 0.049). For all other attributes, at least 78% of participants (male and female respectively) were able to answer both questions correctly; indeed for most attributes, over 90% could do so. These results are much higher than the 25% criterion (randomly choosing one out of the given four answers). We conclude that in our sample, the level of product expertise was sufficiently high for attribute-level preference elicitation, and that differences in expertise could not be explained with gender.

In the main experiment, 93 students from the same university took part. The samples for both experiments were drawn randomly from the same subject pool (without replacement), and we have no indication that they differed substantially. The sample was evenly distributed across treatments with 43 participants in the linear treatment and 50 in the exponential treatment. Participants were aged between 19 and 32 years. 61 participants were female, 32 were male.

Ordered logit regressions did not show any significant differences with respect to gender ($p$ = 0.60), e-shop usage frequency ($p$ = 0.26), perceived ease of use ($p$ = 0.45), perceived usefulness ($p$ = 0.84), end user satisfaction ($p$ = 0.33) or reuse intentions ($p$ = 0.49) between treatments. We did find a significant difference in age between treatment groups ($p$ = 0.02), but no effect of age on utility accuracy ($p > 0.3$)[11] or WTP accuracy ($p$ = 0.66).[12] Results indicate that differences in accuracy depend on the different SAU functions only.

## 3.5 Analysis and Results

As suggested by (Xiao and Benbasat 2007), we used search time to approximate the level of cognitive effort during the use of the recommendation system (Section 3.5.1) and to determine the additional effort of specifying exponential utility functions. To ensure that WTP estimates were based on relia-

[11] We used first-choice hit rate and the correlation between predicted and actual ranks of a product (Section 3.5.2) as the dependent variables.

[12] We used correlation between predicted and actual WTP as dependent variable (Section 3.5.3).

ble and valid input, we examined the accuracy of utility estimates (Section 3.5.2) before comparing the accuracy of WTP estimates (Section 3.5.3) of the two experimental groups in the main experiment.

### 3.5.1 Effort

Search time was measured as time lapsed between participants' first interaction (defining aspiration levels) and last interaction (rating recommended products) with the recommendation agent (tasks A–D in Figure 5). Task difficulty was measured with 4 items ranging from 1 (very easy) to 7 (very difficult). All items referred to the measurement of the utility parameters.

It took our participants on average 804.721 s (*SD* = 325.153) to complete the search task with the linear system and 889.640 s (*SD* = 280.438) with the exponential system. In other words, participants spent significantly more time (gamma regression; $p < 0.1$) on specifying exponential functions. Despite this difference in time, however, perceptions of task difficulty did not differ between treatments (ordered logit regression; $p = 0.200$). All participants considered the task to be quite easy on a seven-point scale ranging from1 (very easy) to 7 (very hard) (linear: 2.634 (*SD* = 1.009); exponential: 2.430 (*SD* = 1.080)).

The supplementary experiment with “free” function shapes, allowing for instance U-shaped SAU functions, was by far the most strenuous for participants. It took them on average 1038.038 s (*SD* = 279.370) to complete the search task. In other words, participants in the supplementary experiment spent 16% more time on that task than participants in the exponential treatment and nearly 30% more than participants in the linear treatment.

### 3.5.2 Utility Accuracy

Before assessing the utility accuracy for linear and exponential systems, we checked whether treatments affected participants' preferences. We found no effect of preference elicitation method on the stated preferences.

Average attribute weights in the two treatments indicate that there was virtually no difference in overall attribute rankings between treatments. Moderately high standard deviations point to heterogeneous preferences within the two treatments (Table 13).

We found no significant differences for estimated utility thresholds $\tau_j$ (t-test, $p > 0.1$). Participants in the linear treatment had an average utility threshold of 28.345 (*SD* = 7.277), and participants in the exponential treatment a threshold of 29.315 (*SD* = 10.070).

Exponential SAU functions were predominantly concave ($u_i(x_i^{average})$ > 0.5; Table 13). Only 16% of participants in the exponential treatment specified the scaling parameter $c_i$ such that the function shape was equivalent to a linear SAU function ($u_i(x_i^{average})$ = 0.5) for most attributes. This is consistent with previous findings indicating that consumers generally do not perceive linear relationships between attribute levels and utilities (Van Ittersum and Pennings 2012) and underlines the importance of estimating exponential functions.

**Table 13: Elicited Utility Parameters (Minimum = 0, Maximum = 1)**

| Attribute | Linear | | Exponential | | | |
|---|---|---|---|---|---|---|
| | Weights | | Weights | | $u_i(x_i^{average})$ | |
| | Mean | SD | Mean | SD | Mean | SD |
| Photo Resolution | 0.789 | 0.193 | 0.736 | 0.229 | 0.670 | 0.158 |
| Zoom Factor | 0.611 | 0.267 | 0.632 | 0.157 | 0.662 | 0.178 |
| Size | 0.586 | 0.266 | 0.484 | 0.264 | 0.550 | 0.212 |
| Display Size | 0.448 | 0.278 | 0.432 | 0.225 | 0.638 | 0.147 |
| Video Resolution | 0.448 | 0.287 | 0.472 | 0.281 | 0.626 | 0.159 |
| Settings | 0.595 | 0.241 | 0.544 | 0.252 | 0.542 | 0.142 |
| Photosensitivity | 0.650 | 0.255 | 0.714 | 0.175 | 0.568 | 0.158 |
| Price | 0.714 | 0.222 | 0.720 | 0.257 | 0.582 | 0.199 |

We used our supplementary experiment to test the notion that SAU functions may be of (inverted) U-shape. For most attributes, the great majority of participants specified monotonously increasing SAU functions (i.e. utility is increasing from the worst to the best attribute level), except size and the range of settings options (Table 14).

**Table 14: Supplementary Experiment: Percentage of Participants' SAU Shapes**

| Attribute | Monotonously Increasing | U-Shaped | Other Shape |
|---|---|---|---|
| Photo Resolution | **81.58** | 13.16 | 5.26 |
| Zoom Factor | **71.05** | 23.68 | 5.27 |
| Size | 23.68 | **76.32** | 0.00 |
| Display Size | **76.32** | 21.05 | 2.63 |
| Video Resolution | **86.84** | 13.16 | 0.00 |
| Settings | 42.11 | **55.26** | 2.63 |
| Photosensitivity | **84.21** | 10.53 | 5.26 |
| Price | **52.63** | 36.84 | 10.53 |

*Note: Bold values indicate the shape the majority of consumers has specified.*

Apparently, participants perceived these attributes not as unidimensional but containing trade-offs between several factors; “size”, for instance, may imply a trade-off between “ease of transportation” and “fragility”. All remaining attributes were considered independently of other attributes by most participants. This suggests that by splitting multi-dimensional attributes into atomic attributes, the performance of our approach could be further improved due to better SAU estimation.

The accuracy of the utility estimates in both treatments was similar to that reported in other empirical studies on utility estimation and prediction of consumers' purchase decisions (Table 15). Note that the studies' settings were very diverse, which somewhat reduces comparability. For instance, task complexity ranged from 12 attributes and 42 levels (Netzer and Srinivasan 2011) to 5 attributes and 10 levels (Jedidi and Zhang 2002). Required effort varied widely, with rating-based conjoint methods requiring as many as 24 attribute-level ratings and 24 product-level pairwise comparisons and self-explicated approaches only 24 attribute-level ratings (Netzer and Srinivasan 2011). Choice-based conjoint methods typically required greater effort, for instance in (Gensler et al. 2012) with 12 choice sets containing 3 products each. Predictive accuracy was computed with differently sized and constructed choice sets, ranging from four (Gensler et al. 2012; Netzer and Srinivasan 2011) to twelve (Green et al. 1993), which brings random FCHR to between 25% and 8.33%, and ratings were given on different scales (e.g. 4-point scales in Green et al. (1993), 100-point scale in Jedidi and Zhang (2002)). Overall, our results for first-choice hit rate – 66.7% (linear) and 56% (exponential) – fall within the bounds of previously attained accuracy values (35.5% to 74%); rank correlations are slightly worse. Bearing in mind that the best-performing methods are variants of conjoint analysis and thus cognitively very challenging and not easily applicable for repeated measurement, our low-effort approach performed very well.

Surprisingly, first-choice hit rate was higher in the linear treatment and, on average, participants in the linear treatment rated the top 10 recommended products higher although the exponential system predicted product ranks beyond first place better. The exponential system produced slightly higher correlations between product ranks and participant ratings.[13] One possible explanation might be the fact that specifying exponential SAU functions is cognitively more complex and leads to higher error levels than does specifying linear functions. This supposition is supported by our system usage log data: participants in the exponential treatment spent on average 10% more time on specifying their preferences than participants in the linear treatment.

Based on the five attribute levels for which participants specified preferences in the supplementary experiment, we computed piecewise linear SAU functions to assess preference fit.[14] Although most

[13] A higher $R^2$ (linear = 0.49, exponential = 0.51, free = 0.28) and a lower RMSE (linear = 1.52, exponential = 1.23, free = 4.28) underline that the exponential system better predicted the sorting of the products.

[14] We also fitted linear, exponential, quadratic and cubic SAU functions based on the 5 attribute level evaluations, but found utility accuracy to be lower than that of the linear and exponential treatment in each case.

flexible, utility accuracy of the "free" specification of SAU functions was below the accuracy of both linear and exponential utility functions (Table 15).

**Table 15: Utility Accuracy Compared to Other Studies**

| Method | FCHR (%) | Rank Correlation (%) | Study |
|---|---|---|---|
| RBCA | 38.0 – 74.0 | 52.0 – 67.7 | De Bruyn et al. (2008); Green et al. (1993); Karniouchina et al. (2009); Moore (2004); Moore et al. (1998); Scholz and Dorner (2012) |
| Adaptive RBCA | 35.5 – 39.8 | 63.5 | Green et al. (1993); Netzer and Srinivasan (2011) |
| CBCA | 38.0 – 67.0 | – | Gensler et al. (2012); Karniouchina et al. (2009); Moore (2004); Moore et al. (1998) |
| Adaptive CBCA | 56.5 – 58.5 | – | Gensler et al. (2012) |
| Augmented CA | 54.7 – 62.3 | – | Jedidi and Zhang (2002) |
| Self-Explicated | 43.6 – 44.0 | 53.7 | Green et al. (1993); Netzer and Srinivasan (2011) |
| Adaptive Self-Explicated | 61.0 | – | Netzer and Srinivasan (2011) |
| DR + Linear SAU | 66.7 | 47.7 | |
| DR + Exponential SAU | 56.0 | 49.4 | This Study |
| DR + Free SAU | 36.1 | 41.0 | |

*Note: FCHR = First choice hit rate; CBCA= Choice-based conjoint analysis; RBCA= Ranking-based conjoint analysis; CA= Conjoint analysis; DR= Direct rating.*

### 3.5.3 Willingness-to-Pay Accuracy

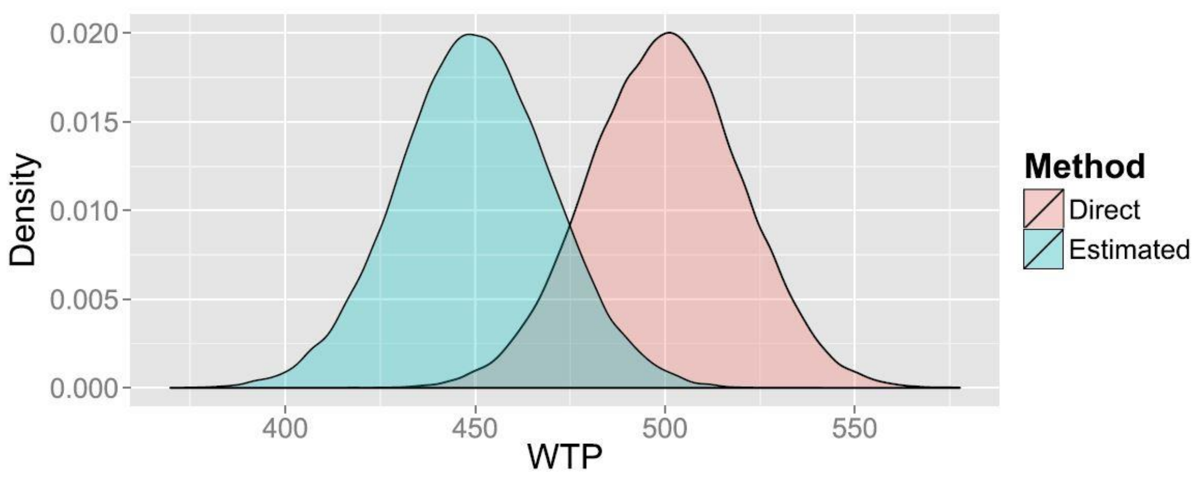



**Figure 6: WTP Density Functions**

For the top 10 products, participants specified their WTP as a range and our systems estimated each participant's WTP as a range. We assume that both WTP ranges follow a truncated normal distribution (Dost and Wilken 2012) with $N(WTP^{ind.}, \sigma, WTP^{floor}, WTP^{ceil})$. If a participant's WTP is estimated accurately, its range must be within the stated WTP range (see overlapping area in Figure 6).

We computed the percentage of the estimated WTP range within the stated WTP range as:

$$\begin{aligned} Accuracy = {} & F\left(WTP_{direct}^{ceil}, WTP_{est.}^{ind.}, \sigma_{est.}, WTP_{est.}^{floor}, WTP_{est.}^{ceil}\right) \\ & -F\left(WTP_{direct}^{floor}, WTP_{est.}^{ind.}, \sigma_{est.}, WTP_{est.}^{floor}, WTP_{est.}^{ceil}\right) \end{aligned} \quad (15)$$

with

$$\begin{aligned} & F\left(WTP_{direct}^{\pi}, WTP_{est.}^{ind.}, \sigma_{est.}, WTP_{est.}^{floor}, WTP_{est.}^{ceil}\right) \\ & = \frac{\phi\left(\frac{WTP_{direct}^{\pi} - WTP_{est.}^{ind.}}{\sigma_{est.}}\right) - \phi\left(\frac{WTP_{est.}^{floor} - WTP_{est.}^{ind.}}{\sigma_{est.}}\right)}{\phi\left(\frac{WTP_{est.}^{ceil} - WTP_{est.}^{ind.}}{\sigma_{est.}}\right) - \phi\left(\frac{WTP_{est.}^{floor} - WTP_{est.}^{ind.}}{\sigma_{est.}}\right)} \end{aligned} \quad (16)$$

where $\sigma_{est.}$ is $(WTP_{est.}^{ceil} - WTP_{est.}^{floor})/s$ and $\phi$ is the cumulative distribution function of $N(WTP^{ind.}, \sigma, WTP^{floor}, WTP^{ceil})$. We used 10 variations of $s$ (2, 4, 6, 8, 10, 12, 14, 16, 18, 20).

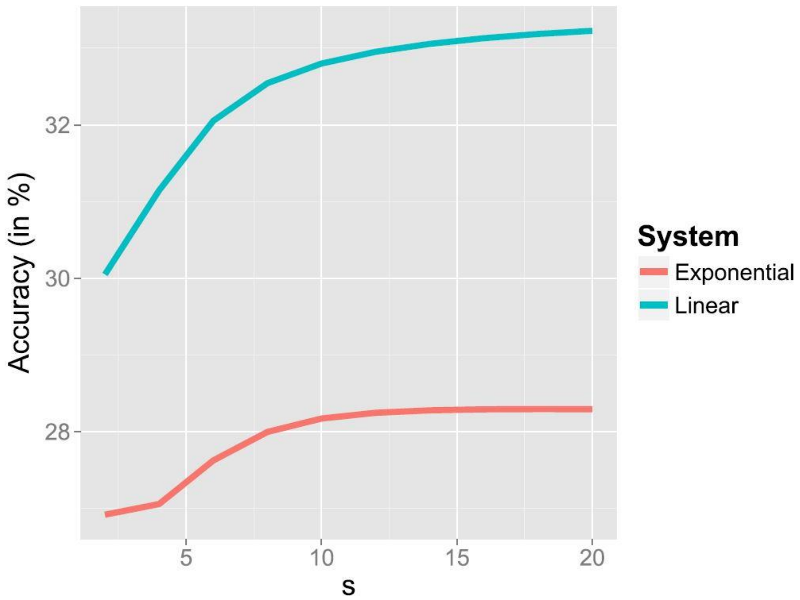



**Figure 7: WTP Accuracy**

WTP accuracy was much higher in the linear treatment than in the exponential treatment (Figure 7). This result is due to the higher first-choice hit rate in the linear treatment, which indicates that the product with the highest expected utility was identical with product perceived to be the best by a participant more frequently. We used variants of the products with the highest expected utilities to estimate a participant's WTP for all products (Section 3.3). WTP accuracy for the supplementary experiment was much worse than for either treatment in the main experiment at values between 13.04% and 15.52%.

We also found evidence that, on average, estimated WTP was lower than stated WTP in both treatments. For floor WTP, differences between estimated and stated indifference WTP were larger in the exponential treatment. For ceiling WTP, average differences were virtually identical in both treatments. In summary, stated WTP was underestimated in both treatments, and the exponential system was more prone to underestimation than the linear system. Identical utility thresholds $\tau_j$ in both treatments ($p$ = 0.595) and identical attribute weights $w_i$ (see Table 13) indicate that the utility parameters ($a_i$, $b_i$ and $c_i$) of the exponential SAU functions were subject to a higher estimation error than the utility parameters ($a_i$ and $b_i$) of the linear SAU functions.[15]

Logit regressions showed that WTP accuracy was not influenced by any differences in participants' preference structures. We also controlled for effects of demographic factors, e-shop usage and system perception on accuracy but found no significant differences between treatments ($p$ > 0.1).

Linear system WTP estimates outperformed exponential system WTP estimates in terms of correlation between estimated WTP and stated WTP. Indifference and ceiling WTP were estimated accurately by the linear system, but predictions for floor WTP were less accurate than those of the exponential system (Table 16). Again, the supplementary experiment results show that piecewise linear (i.e. free) SAU functions lead to less rather than more accurate WTP estimates.

**Table 16: WTP Correlation (in %)**

| WTP Point | Linear | Exponential | Free |
|---|---|---|---|
| Floor | 41.51 | **44.54** | 40.25 |
| Indifference | **50.38** | 47.33 | 44.59 |
| Ceiling | **67.48** | 44.94 | 41.51 |

*Note: Bold values indicate the system with the higher WTP correlation (i.e. more accurate WTP prediction).*

Compared to most other WTP estimation methods, WTP accuracy was outstanding with correlations of nearly 50% in both groups in the main experiment. Prior studies report correlations between 10% and 21% for choice-based conjoint analysis and correlations of 33% to 42% for adapted choice-based conjoint analysis (Gensler et al. 2012), and correlations of 15% to 43% for augmented conjoint analy-

[15] This can be derived from Equations (13) and (14).

sis (Jedidi and Zhang 2002). We conclude that our recommendation systems estimated WTP as accurately as or better than established methods.

## 3.6 Implications

Knowledge about a consumer's utility threshold and willingness to pay makes it possible to compute revenue-maximizing market prices and profit-maximizing individual product configurations. We discuss both practical implications in the following subsections.

### 3.6.1 Pricing

The WTP estimates of our proposed utility-based recommendation systems can be used to compute revenue-maximizing market prices. To compare the effect of both treatments on revenue, we followed a four-step procedure. In the first step, we extracted all products that had been rated in Task D (see Figure 5) by at least 10 participants (32 out of 162) to obtain robust estimates for the demand functions. In the step 2, we estimated the demand function $d(p)$ for each of these products. We used a logit model of the form $d(p) = (e^{\alpha+\beta p})/(1 + e^{\alpha+\beta p})$ (Dost and Wilken 2012; Miller et al. 2011; Wertenbroch and Skiera 2002), where $p$ denotes participants' stated indifference price and $d(p)$ is the ratio of participants willing to purchase the product at price $p$. In step 3, we predicted a revenue-optimal price $p^*_{prediction}$ based on the WTP estimates. We selected the revenue-maximizing value (i.e. price multiplied with number of participants willing to purchase at this price) for WTP as $p^*_{prediction}$. Instep 4, we computed revenue as $r(p) = d(p)p$ for real market prices $p_{real}$ and predicted revenue-maximizing prices $p^*_{prediction}$. Since $d(p)$ is the predicted probability that a consumer will purchase a product at price $p$, revenues were calculated for each consumer considering the product.

For each estimated demand function $d(p)$, we conducted a likelihood-ratio test, computed Nagelkerke's $R^2$ and checked whether the function was monotonously decreasing. All estimated demand functions were monotonously decreasing and fitted our data well with significant likelihood-ratio tests ($p < 0.05$). On average Nagelkerke's $R^2$ was 69.73% ($SD$ = 5.26%) indicating validly estimated demand functions.

We found that, on average, $p^*_{prediction}$ was significantly lower than the market price $p_{real}$ for both treatments. This is not surprising considering that both treatments underestimated stated WTP. Still, market prices calculated based on the WTP estimates of either system led to an increase in revenue in all but one case (Table 17).

Increase in revenue is highest when calculated based on linear system indifference WTP estimates. As a reference, we computed maximal revenue improvements based on the estimated demand functions $d(p)$ and found a maximal improvement of 18.35%. Although WTP accuracy was only moderate, WTP estimates were accurate enough for improving pricing strategies.

**Table 17: Average Revenue Improvement with WTP Estimates (in %)**

| Point in Estimated WTP-Range | Linear System | Exponential System |
|---|---|---|
| Floor WTP | **+ 1.87** | − 5.57 |
| Indifference WTP | **+ 6.05** | + 1.53 |
| Ceil WTP | + 1.61 | **+ 2.78** |

*Note: Bold values indicate the system with the highest revenue improvement.*

### 3.6.2 Special Offers

Establishing new market prices is often impossible due to high price transparency, especially in online markets. However, knowledge about a consumer's willingness to pay and utility threshold can be employed to create individually optimal and profit-optimal product configurations (e.g. digital cameras or racing bikes) or special offers (Table 18).

**Table 18: Special Offer Opportunities**

| Notebook | Processor in GHz | Memory in GB | HDD in GB | Procurement Costs in EUR |
|---|---|---|---|---|
| A | 3.2 | 4 | 500 | 500 |
| B | 2.8 | 8 | 500 | 550 |
| C | 2.8 | 4 | 750 | 520 |

Consider a consumer who is planning to purchase a new notebook and has recently used our utility-based recommendation system to search for attractive products. Based on the elicited utility functions for processor speed ($w_i$ = 6 and $u_i(x_i)$ = −1.25 + 0.625$x_i$), memory size ($w_i$ = 5 and $u_i(x_i)$ = −0.143 + 0.071$x_i$), HDD size ($w_i$ = 7 and $u_i(x_i)$ = −0.333 + 0.0013$x_i$) and price ($w_i$ = 8 and $u_i(x_i)$ = 1.667 − 0.0017$x_i$), we can compute each notebook's utility.

Let us assume that the retailer offers a notebook *X* with 2.8 GHz processor speed, 4 GB memory size and 500 GB HDD for a regular price of 600 EUR. We further assume that the retailer has purchased this notebook for 450 EUR. Because its utility ($U(X)$ = 11.1) lies below the consumer's utility threshold ($\tau_j$ = 12.0), she will not be tempted to buy it. If the retailer wishes to attract this particular consumer, they might consider giving her a special offer applying a price discrimination strategy. In this case, the retailer can realize a profit of 83 EUR, since the consumer has a willingness to pay of $WTP_X$ = 533. Alternatively, the retailer might offer her one of the three other notebooks (notebooks *A*, *B*, and *C* in Table 18). Assuming that prices for *A*, *B*, and *C* are identical, notebook *A* is most attractive from the retailer's point of view because its procurement costs are lowest. However, notebook *C* leads to the highest utility improvement for our consumer. In other words, our consumer has the highest willingness to pay (701 Euro) for product *C* ($WTP_A$ = 644 Euro and $WTP_B$ = 638 Euro). Selling notebook *C* at 701 Euro will maximize the retailer's profit (181 Euro).

## 3.7 Summary and Future Research

Our proposed recommendation system generates real-time data on consumer purchasing behavior, especially consumers' WTP, which are the basis for many models of market share estimation, pricing or product design. Because such data are (theoretically) easy to collect online, the increasing popularity of e-commerce has led to renewed interest in developing methods for estimating consumers' preferences and WTP in operations research (Abbas and Bell 2011; Gensler et al. 2012; Miller et al. 2011; Rusmevichientong et al. 2010). Specifically, a low-effort method for repeated measurements of consumers' WTP and attribute-level utilities is needed. Our approach is a promising step towards the development of such a method, extending utility-based recommendation systems.

The empirical evaluation shows that our proposed recommendation system predicts consumers' utility functions and, ultimately, their WTP with high accuracy. Considering that prior studies which reported similar levels of accuracy used more complex and cognitively exhausting measurement methods (e.g. choice-based or ranking-based conjoint analysis), our approach performed very well. Our results indicate that, contrary to prior suppositions, linear SAU functions are better suited for estimating WTP. Exponential SAU functions provide better approximations of consumers' preference structures and are therefore better suited for predicting product ranks. That overall utility accuracy for exponential functions is lower than for linear functions, although SAU functions were predominantly concave, suggests that more complex and therefore challenging methods lead to higher error levels in SAU function specification, which then accumulate in the overall utility function. This supposition is supported by the fact that complex conjoint analysis methods often do not attain much higher accuracy levels than simpler approaches although they too model utility at the attribute level. We suggest that for utility estimation, other approaches are as well suited as ours, but for combined individual utility and WTP estimation, our approach seems to perform above average.

Our research is subject to some limitations. First, our approach is designed for products with at least ordinal attributes for which it is possible to estimate reliable SAU functions, not for experience goods with nominal attributes such as “design” or “color”. Second, as shown in our supplementary experiment, SAU functions can be of (inverted) U-shape if consumers think about trade-offs between two attributes during the evaluation of one attribute (e.g. between “ease of transportation” and “fragility” when evaluating the attribute “size”). Exponential approximation of SAU functions was particularly robust when consumers did not consider within-attribute trade-offs during evaluation. Splitting multi-dimensional attributes into atomic attributes would likely improve accuracy and reduce effort. Third, the accuracy of utility estimates on the product level was only slightly better in the exponential system than in the linear system although preference structure (SAU functions) was approximated much better with exponential functions. It is possible that the 10-point scale we used to determine SAU function curvature was not precise enough, resulting in small discrepancies at the SAU function level (see Figure 8) which might then have added up to a larger discrepancy between estimated and actual, or perceived, utilities at the product level. Fourth, user preferences may be influenced by exogenous factors like the decision support system used for measuring utility functions and WTP (Adomavicius et al. 2013). Our proposed approach for estimating exponential utility functions is interactive, requiring user input for several parameters, and may affect user preference building. The extent to which and the circumstances under which decision support systems change users' preferences are as yet not fully understood and provide an avenue for future research.

The practical implications from our findings are threefold. First, it is relatively easy to use and generates good recommendations. These results suggest that online consumers could benefit from using it in their purchasing process, saving effort in the process of finding attractive products. Since multi-dimensional attributes make utility estimation more difficult and increase consumer effort without improving utility or WTP estimates, designers of utility-based decision support systems ought to make sure their systems are based on atomic attributes only. Second, online retailers could easily extend existing utility based recommendation systems, such as the Dell Computer Advisor, to include WTP estimation. As illustrated in Section 3.6, retailers could use this information as a basis for a number of business decisions, e.g. product pricing. Third, online retailers could generate new revenue streams by selling the recommendation data to product manufacturers, who can then determine individual profit-maximizing product configurations more easily. Open questions that remain in this area are the degree of consumer acceptance and the profitability of WTP-based pricing strategies. Both still need to be evaluated in field studies to shed more light on consumers' reactions and decision processes in real-life situations.

Research into collaborative recommendation systems could profit from our approach. Currently, one of the most commonly used collaborative recommendation algorithms is matrix factorization (e.g. Ge et al. 2014), which helps identify latent product features that contribute most to product utility. However, these features are generally not identical to product attributes, which makes it more difficult to use the information in business decisions and to compute WTP. Combining our approach, which produces data on the individual level, with collaborative data on the user group level could generate better insights into the composition of latent features and preference differences between consumers. Conversely, integrating collaborative recommendation system data in our approach may further increase recommendation quality and decrease the level of consumer effort during the purchasing process, especially if consumer preferences can be estimated to a satisfactory degree during the specification process and recommendations provided at an early stage.

## 3.8 Acknowledgment

The authors thank Christian Schlereth for his insightful comments. The authors also thank the editor and two anonymous reviewers for their helpful comments and suggestions.

## 3.9 Appendix A. Proof for Equations (8) and (9)

Exponential utility functions consist of three parameters. We hence need three points of such a function for estimation. We suggest using the best, the worst and the average attribute level and the utility values of these levels to estimate an exponential utility function. Assuming that both, the attribute levels as well as the utility values of an attribute, are scaled to [0; 1], the first two points are given by

$$P_1\left(x_i^{worst};\, u_i\left(x_i^{worst}\right)\right) = (0;0) \tag{17}$$

$$P_2\left(x_i^{best};\, u_i\left(x_i^{best}\right)\right) = (1;1) \tag{18}$$

The utility of the third point is specified by the consumer on a 11-point scale (see Figure 4) and rescaled to [0; 1]. The third point is hence given as:

$$P_3\left(x_i^{average};\, u_i\left(x_i^{average}\right)\right) = \left(0.5; u_i\left(x_i^{average}\right)\right). \tag{19}$$

Based on these three points, we can define the following three equations that can be solved with Gaussian elimination:

$$0 = a_i - b_i, \tag{20}$$

$$1 = a_i - b_i e^{c_i}, \tag{21}$$

$$u_i\left(x_i^{average}\right) = a_i - b_i e^{0.5c_i}. \tag{22}$$

Equation (20) indicates that $a_i = b_i$ whereas Equation (21) indicates that $a_i = \frac{1}{1-e^{c_i}}$. By substituting $a_i$ and $b_i$ in Equation (22) we get:

$$u_i\left(x_i^{average}\right) = \frac{1-e^{0.5c_i}}{1-e^{c_i}}. \tag{23}$$

Next, we can substitute $e^{c_i}$ by $z^2$ with $z \geq 0$ to get a simple quadratic equation:

$$0 = u_i(x_i^{average})z^2 - z - u_i(x_i^{average}) + 1. \tag{24}$$

Isolating $z$, gives:

$$z_{1/2} = \frac{1 \pm \sqrt{1 - 4u_i(x_i^{average})(1 - u_i(x_i^{average})}}{2u_i(x_i^{average})}. \tag{25}$$

If $u_i(x_i^{average})$ = 0.5 we have $z_1$ = $z_2$ = 1 and therefore $c_i$ = $\ln(z^2)$ = 0. If $u_i(x_i^{average} > 0.5)$ then $z_2$ is the only non-negative solution whereas only $z_1$ is a non-negative solution if $u_i(x_i^{average})$ < 0.5.

Based on the rating $r$ of the average attribute level $x_i^{average}$, an attribute utility function is defined as one of the functions given in Figure 8.

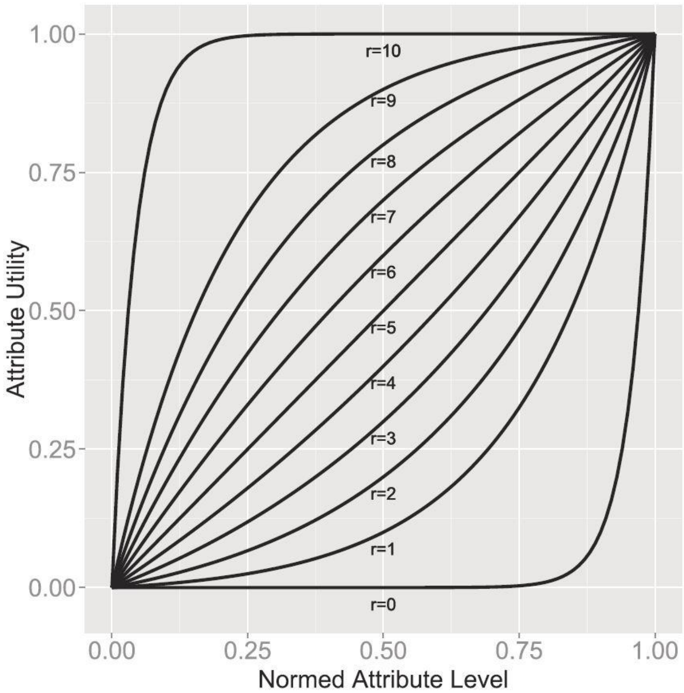



**Figure 8: Possible Attribute Utility Functions**

# 4 2D versus 3D Visualizations in Decision Support – The Impact of Decision Makers' Perceptions

**Title:** 2D versus 3D Visualizations in Decision Support – The Impact of Decision Makers' Perceptions

**Authors:** Franz, Markus; TU Darmstadt; Germany
Scholz, Michael; University of Passau; Germany
Hinz, Oliver; TU Darmstadt; Germany

**Published:** International Conference on Information Systems (2015), Fort Worth, USA

**Abstract**

Decision makers often do not easily understand the decision space, i.e. the available alternatives and relations among their attributes. Misunderstanding the relations between attributes might lead to a bad decision. Recent research in decision analysis has addressed this problem and proposed to support decision makers with visual information about the decision space. However, the question of which visualization method and format supports decision makers best is largely unanswered. We focus on coordinate systems as visualization method and investigate the impact of 2D and 3D formats on decision making processes. We show that 3D is not superior to 2D in terms of several decision making performance variables, such as time to make a decision. We, however, provide first evidence that 2D and 3D visualizations differ in the decision makers' perceptions and that these differences are moderated by the complexity of the decision space.

**Keywords:** Information visualization, decision support system, decision space, dimensionality, decision making performance, system perceptions

## 4.1 Introduction

Decision makers often face decision making scenarios that are more complex and challenging than decision analysts might expect at a first glance. Assume, for example, a decision maker who is searching for a rental apartment with a large living space. The decision maker might be aware that apartments with a larger living space will require a higher rental price. The decision maker has thus to deal with potentially conflicting attribute relations. A conflicting attribute relation exists, if it is necessary to exchange the outcome of one attribute to attain a higher outcome of another one. In this example, the decision maker's preference for a larger living space requires accepting losses in another potentially desired apartment attribute, i.e. a low rental price.

It is further likely that the decision maker simply does not know of all conflicting attribute relations that need to be accepted or considered in the search process. Larger apartments might, for example, be either primarily located in the highly frequented city center, which reduces the likelihood of finding an apartment simultaneously offering an also preferred parking lot. It is also possible that large apartments are only located far away from the city center, which requires traveling undesirable large distances to the decision maker's workplace on a daily basis. Which story is true depends on the individual city structure and is likely to be unknown to the decision maker. Some conflicting attribute relations may be more obvious (e.g., floor space vs. rental price) than others (e.g., floor space vs. parking lot or location). Simply searching for large apartments may lead to dissatisfying results. The decision maker might become aware of this situation, though, most likely after a time-consuming and effortful search and comparison process.

This problem is also known in the scientific literature. Previous studies indicate that knowing the available decision alternatives, related attribute combinations and conflicting attribute relations cannot be presumed, but may be beneficial in the decision making process (Butler et al. 2008; Hoeffler and Ariely 1999; Huber and Klein 1991; Keeney 2002). In the following we refer to the information about the available decision alternatives and the available attributes as the decision space.

Decision spaces are often complex in real decision making scenarios. In consumer decision making, consumers as a special kind of decision makers, for instance, consider up to eight rather than two attributes when comparing alternatives for a purchase (Jacoby et al. 1977; Moorthy et al. 1997; Olson and Jacoby 1972; Sheluga et al. 1979). This increases the number of attribute relations describing the decision space from 1 to 28 (= 8(8-1)/2). As these studies show, consumers do even consider less than eight attributes in many cases. Other decision making scenarios, such as managerial decision making, can easily comprise more than eight attributes and surpass the level of complexity prevalent in consumer decision making. Since we focus on consumer decision making in this study, higher levels of complexity are beyond the scope of this research.

Decision makers using a decision support system might benefit from getting information on the decision space, since this information helps to recognize and incorporate complex relations into their decision making process. Supporting the decision making process can i) reduce the decision making effort required for processing information on and the comparison of decision alternatives, ii) improve the quality of the decisions induced by the system's recommendations (e.g., in terms of choosing non-dominated alternatives or the decision makers' subjective evaluation of the recommendations) and iii) the decision makers' perceptions and subjective evaluations of the decision support provided by a decision support system (Xiao and Benbasat 2007).

Providing decision makers with additional information, however, does not come without threats. Due to the limited cognitive capacities to process information (Payne et al. 1993), supporting decision makers with additional information might lead to information overload and might thus reduce the quality of decisions (Lurie 2004). If information is, however, provided in a visual form like graphs or pictures, the decision makers' perceptual system can easily encode this information (Lohse 1997; Zhang and Whinston 1995). Decision makers are able to extend their information processing capacities and reduce the threat of information overload (Lohse 1997; Tegarden 1999; Zhang and Whinston 1995). It is hence important to support decision makers with information on the available alternatives and attribute relations, but it seems to be even more important to support them with the right visualization type.

Information visualization has gained popularity as it seems to be a plausible way to support decision makers efficiently with information on decision making scenarios (Lurie and Manson 2007; Turetken and Sharda 2001). Information visualization is defined as the transformation and presentation of information using a visual medium (Lurie and Manson 2007). It makes use of the human ability to transform visual cues like detected patterns or differences in shape or color of visual objects into knowledge (Kosslyn 1994). Previous research developed different visualization types to support decision making, such as tables, coordinate systems or bar graphs (e.g., Kumar and Benbasat 2004; Theetranont et al. 2007; Tractinsky and Meyer 1999). In this study, we focus on coordinate systems as visualization type that is easy to interpret, since decision makers are typically familiar with coordinate systems. We are especially interested in analyzing the effects of the visualization dimensionality (2D vs. 3D) of coordinate systems that provide information about decision spaces on decision maker's decision quality, decision effort and their perceptions.

A review of previous studies comparing 2D and 3D visualizations reveals a gap in the literature: there is mixed evidence on the question whether 2D or 3D visualizations provide superior decision support. The choice of the right visualization format is important, since inappropriate decision support may lead to a decrease in decision quality (Xiao and Benbasat 2007), which might be costly when it comes to management decisions. Shopping websites may further lose customers, since inappropriate decision support has been shown to reduce the intention of consumers to adopt decision support systems for future use (Xiao and Benbasat 2007). Further, the effects observed by prior research might not be complete as the decision support of 2D and 3D visualizations has primarily been compared and evaluated using measures regarding the observable decision making performance (i.e., time to make a decision, objective decision quality). Recent research, however, cites evidence that modifications to a decision making process are likely to affect a wide range of user perceptions regarding the decision support (Xiao and Benbasat 2007).

We contribute to recent research by investigating whether the 2D or the 3D visualization format provides better decision support in simple or complex consumer decision making scenarios. We evaluate the quality of the decision support in terms of i) observable decision making performance and ii) decision makers' perceptions of the decision support process. In a laboratory experiment, we ask the study participants to use one of four decision support systems providing visual information about the decision space to search for a digital camera for purchase. To observe the interplay of dimensionality and decision making complexity in case of consumer decision making, we compare both visualization formats in i) a simple purchase decision scenario (alternatives are described by four attributes) and ii) a complex purchase decision scenario (alternatives are described by eight attributes). Overall, we use four treatments: i) 2D/four attributes, ii) 3D/four attributes, iii) 2D/eight attributes and iv) 3D/eight attributes. We collect data on the decision making performance in terms of effort and quality and the participants' perceptions of each decision support system.

The remainder of this article is organized as follows: In the next section, we relate our work to prior studies comparing 2D and 3D visualizations and describe our contribution in more detail. We then explain the visualization method applied to our study. We thereafter describe our empirical investigation and the results of the laboratory experiment. Finally, we discuss the implications for practice and future research.

## 4.2 Related Work

In the following section, we briefly review previous studies that compare 2D to 3D visualizations. Existing studies fall into three categories: i) studies generally supporting the predominance of either 2D or 3D visualizations, ii) studies finding mixed evidence on the value of 2D versus 3D visualizations and iii) studies recommending the use of a specific format (either 2D or 3D) depending on specific conditions. Table 19 summarizes the findings of our literature review.

There are only a few studies that clearly prefer one visualization format. Pilon and Friedman (1998) compare the efficiency of the search for specific objects in 2D and 3D visualized environments, concluding that 2D visualizations enable more efficient searches than 3D visualizations. In contrast, the studies of Dull and Tegarden (1999) and Kumar and Benbasat (2004) recommend using 3D rather than 2D visualizations. Dull and Tegarden (1999) investigate the impact of 2D, 3D and rotatable 3D visualizations on decision makers' effort and decision quality. They find that 3D visualizations help making better decisions without decreasing the time required for decision making. Only the comparison of 2D to a rotatable 3D visualization leads to reduced decision making time. The study of Kumar and Benbasat (2004) focuses on the interaction of graph complexity (nine and 25 data points displayed) and the task type (pattern and trend recognition and extraction of concrete data values) on the comprehension effort, i.e., the time taken to comprehend specific information from 2D versus 3D visualizations. In their conclusion, Kumar and Benbasat (2004) clearly recommend using 3D over 2D visualizations, since 3D visualizations allow processing information quicker in all cases, independent of task type and graph complexity.

**Table 19: Prior Studies Comparing 2D and 3D Visualizations**

| Source | Decision Task | Visualization Type | Performance Measures | Results |
|---|---|---|---|---|
| Pilon and Friedman (1998) | Object retrieval | 2D and 3D objects | Effort | 2D is superior |
| Dull and Tegarden (1999) | Forecasting | 2D and 3D line graphs | Effort and quality | 3D is superior |
| Kumar and Benbasat (2004) | Graph comprehension | 2D and 3D line graphs | Effort | 3D is superior |
| Zhu and Chen (2005) | Knowledge retrieval | 2D and 3D geographic maps | Effort and quality | Mixed evidence |
| Kim et al. (2011) | Task fulfillment | 2D and 3D cell phone menus | Effort and user perceptions | Mixed evidence |
| Nah et al. (2011) | Task fulfillment | 2D and 3D virtual world environments | User perceptions | Mixed evidence |
| Van der Land et al. (2013) | Graph comprehension and group decision making | 2D and 3D virtual world environments | Effort, quality and comprehension | Mixed evidence |
| Lee et al. (1986) | Managerial decision making | 2D and 3D scatter grams and block diagrams | Effort and quality | Depending on the information format |
| Tractinsky and Meyer (1999) | Situational visualization choice | Table, 2D and 3D graph | User perceptions | Depending on visualization purpose |
| This study | Consumer decision making | 2D and 3D coordinate systems | Effort, quality and user perceptions | Depending on complexity |

The vast majority of studies, and especially studies in the recent past, provide rather mixed evidence and do not make general recommendations for the use of a specific visualization format. Zhu and

Chen (2005) shed light on the question whether the visualization format (2D versus 3D) impacts the effort and the quality of conveying spatial knowledge in a geographical information visualization system. They did not find a consistent performance difference between 2D and 3D visualizations in conveying declarative, configurational or procedural knowledge. The study of Kim et al. (2011) compares 2D and 3D cell phone menus with respect to task performance in terms of time required for task fulfillment and user perceptions in terms of perceived space use, fun of use and satisfaction. They find evidence that users need less time to find specific menu entries in large menus using 2D menus, but there is no difference in task performance for smaller menus. They further find mixed evidence with respect to user evaluations of 2D and 3D menus for different menu sizes. Nah et al. (2011) examine 2D and 3D virtual environments, and especially the impact of dimensionality on user telepresence, enjoyment, brand equity and the behavioral intention. They observe that users do enjoy 3D virtual worlds more than 2D, but 3D virtual worlds lead to lower brand equity. Finally, Van der Land et al. (2013) investigate the understanding of objects and the performance of group decision making in 2D and 3D virtual environments. They conclude that 3D visualizations are more effective in supporting the understanding of individuals and groups, but reduce the efficiency of group decision making.

Two further studies provide implications regarding the use of 2D and 3D visualizations depending on certain conditions. First, Lee et al. (1986) consider the influence of different information formats (continuous and discrete data) on the performance of decisions when decision makers are supported by 2D and 3D visualizations. The study provides evidence that 3D visualizations help improving the quality of decisions at a constant effort in case of information is continuous. 2D visualizations provide better decision support when visualizing discrete information. Tractinsky and Meyer (1999) focus on the interaction of the visualization purpose (decision support and impressing others) and desirability of the information content (e.g., negative information about oneself) with an individual's visualization choice for that specific purpose (2D bar graph, 2D bar graph augmented by perspective, 3D bar graph and tables). They find that 2D visualizations are preferred for decision making scenarios over 3D visualizations. 3D visualizations are rather preferred for the purpose of impressing others, especially when the content of the information provided was undesirable for the individual.

In summary, we can draw two conclusions from the review of prior work: First, prior studies primarily compared 2D to 3D visualizations focusing on decision making performance in terms of decision making effort and the quality of the decisions made. Second, previous research provides mixed evidence on the question whether 2D or 3D visualizations provide superior decision support and is, with except of Lee et al. (1986), lacking clear advice regarding the use of a specific format for decision making.

We address this research gap by investigating the decision support of 2D and 3D coordinate systems in simple and complex consumer decision making scenarios. More specifically, we evaluate the decision support of 2D and 3D coordinate systems in terms of decision making performance (comprising effort and quality of the decisions) as well as commonly applied user perception measures from information systems research. We use several different perception measures in order to better understand when and why a specific visualization format (2D or 3D) should be preferred. In the following section, we briefly describe how the visualizations applied in this study (2D and 3D coordinate systems), support decision makers in simple and complex consumer decision making scenarios by providing relevant information on their decision task.

### 4.3 Visualization Method

Coordinate systems can provide meaningful information about decision spaces to assist the process of decision making. As a simple example, assume a decision space for rental apartments where apartments are described by the two attributes living space and rental price. Apartments with a larger living space typically come at a higher rental price. We can visualize information about this decision space in a 2D coordinate system as plotted in Figure 9; i.e., by assigning each apartment attribute to one axis of the coordinate system and depicting rental apartments using their attribute combination as coordinates. Apartments (represented by the red dots in Figure 9) will be plotted the farther on the right, the higher the floor space and the farther in the upper region of the coordinate system, the higher the rental price of an apartment.

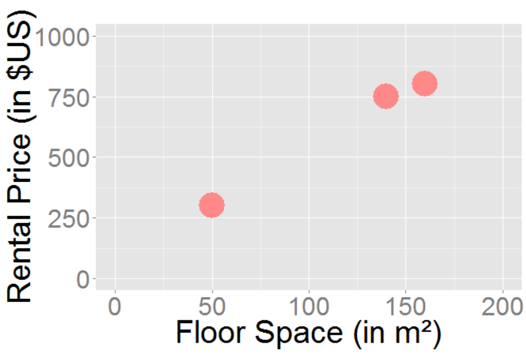



**Figure 9: Example of a 2D Coordinate System**

A decision maker using this visualization to support the search for a suitable rental apartment can retrieve essential information about the decision space, that is: i) the available decision alternatives (apartments), ii) its attribute level combinations and iii) relations among the apartment attributes. Attribute levels of the available apartments are represented by the coordinates of the apartments in the coordinate system. Apartments in the lower left corner of the visualization in Figure 9 are small and cheap whereas apartments in the upper right corner are large and expensive. The visualization hence enables a decision maker to easily compare the attribute combinations of multiple decision alternatives. A decision maker can further become aware of the relations among the attributes. The visualization clearly indicates that apartments with a higher floor space are available at a higher rental price. Floor space and rental price thus form a conflicting relation with respect to the given decision space. A decision maker can use the information inferred from the visualization and incorporate them into his/her decision making process.

In this study, we aim at comparing 2D and 3D coordinate systems in simple and complex consumer decision making scenarios. We define the complexity of a decision making scenario by the number of product attributes a decision maker considers when comparing alternatives. Our example is limited to only two attributes. Theetranont et al. (2007) proposed the application of a 3D coordinate system to visualize decision spaces consisting of three attributes. Several decision situations are described by more than three attributes and decision makers are willing and able to also consider more than three attributes. Consumers as a special instance of decision makers use up to eight attributes when making a complex purchase decision (Jacoby et al. 1977; Moorthy et al. 1997; Olson and Jacoby 1972; Sheluga et al. 1979). To support complex consumer decision making scenarios, 2D and 3D coordinate systems need to be prepared for providing information about higher dimensional decision spaces (i.e., more than 3 dimensions) that contain information about the relations among any required number of attributes and the decision alternatives. To enable displaying more than three dimensions in a coordinate system, a dimensional reduction of the decision space is required.

We suggest singular value decomposition (SVD) for reducing dimensions. SVD is the standard approach for linear dimension reduction (Zhang et al. 2007) and is, for example, used to compute a principal component analysis and to extract GAIA planes to visually plot decision spaces (Brans and Mareschal 1994)[16]. It allows reducing any dimensional spaces to two or three dimensions while maximizing the preserved variance of the data (Härdle and Simar 2003). For our study, the SVD reduces a high-dimensional decision space to a two or three dimensional visual representation, ensuring that the relations among the decision alternatives and up to eight attributes are interpretable in a meaningful manner.

In the following paragraphs, we briefly explain the process of dimension reduction and arriving at a visual representation of the decision space using SVD. We explain the visualization process using the

[16] The major difference between GAIA planes and our visualization method is the rescaling of the data points (i.e., alternatives and attributes). With the rescaling method used in our study (see Equations (27) and (28)), it is possible to not only interpret the distances between alternatives and the distances between attributes but also the distances between alternatives and attributes.

decision space for rental apartments as an example. The SVD process usually starts with a data matrix containing information about two variables. These two variables are here the decision alternatives and the attributes describing the alternatives which jointly constitute the decision space. Assume a decision space consisting of four rental apartments (A1, A2, A3, and A4) that are described by four attributes (rooms, furnishing, location, and price). The cells of the data matrix $x_{ij}$ in Table 20 contain normalized information about the decision space, i.e., the levels of the attributes $j$ of each alternative (apartment) $i$. The normalization yields a uniform operationalization of attribute levels so that low levels of $x_{ij}$ represent low or undesirable attribute levels (e.g., a low number of rooms or a high price) and vice versa.

**Table 20: Example of a Data Matrix**

| Alternative | Rooms | Furnishing | Location | Price |
|---|---|---|---|---|
| A1 | 5 | 7 | 7 | 4 |
| A2 | 4 | 5 | 8 | 9 |
| A3 | 7 | 3 | 9 | 6 |
| A4 | 8 | 2 | 5 | 5 |

The decision space consists of four dimensions (i.e., alternatives are described by four attributes). A visual representation of the relations among the rental apartments and the apartment attributes thus requires a reduction to two or three dimensions. The data matrix (Table 20) will be first transformed into a standardized matrix $Z$. $Z$ is then decomposed into the three following components (Härdle and Simar 2003): i) a diagonal matrix $\sum$ containing $K$ singular values, such that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$, which are used to extract the number of desired dimensions for the visual representation; ii) a $\Gamma(I \times K)$ matrix which contains information that is required to calculate the coordinates of the alternatives $i$ ($\Gamma$ contains the eigenvectors of $ZZ^T$); and iii) a matrix $\Delta(J \times K)$ containing information that is required to derive the coordinates of the attributes $j$ ($\Delta$ contains the eigenvectors of $Z^TZ$). Equation (26) describes the SVD of matrix $Z$.

$$Z = \Gamma \sum \Delta^T \tag{26}$$

The SVD uses the information described in Equation (26) to compute two separate visual parts: i) one part containing standard coordinates for each alternative $i$ which reflects the relations among the alternatives, and ii) one part containing standard coordinates for each attribute $j$ which represents the relations among the attributes, where $k$ represents the number of the visual dimensions of a coordinate. However, interpreting alternatives and attributes, i.e., deducing the attribute combination of an alternative from its position relative to the attribute points in the coordinate system, is not possible based on these standard coordinates. The standard coordinates are thus rescaled to allow a meaningful joint interpretation of both visual parts. If both variables are to be interpreted jointly, Greenacre (2007) suggests a symmetric rescaling. Equation (27) thus yields the final and rescaled coordinates for each alternative $i$, while Equation (28) yields the coordinates for each attribute $j$, where $k$ is the number of visual dimensions ($k$ = 1 is the first dimension, $k$ = 2 is the second dimension and $k$ = 3 is the third dimension).

$$r_{ik} = \frac{\gamma_{ik}\sigma_k}{\sqrt{p_{i\cdot}}} \tag{27}$$

$$c_{jk} = \frac{\delta_{jk}\sigma_k}{\sqrt{p_{\cdot j}}} \tag{28}$$

We extracted the first two singular values ($\sigma_1$ and $\sigma_2$) of our apartment example to generate a 2D visualization of the apartment decision space (see Figure 10).

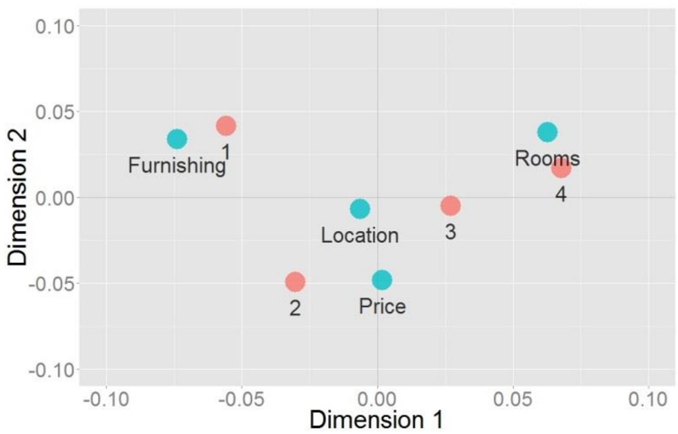



**Figure 10: Relationship between Attributes and Alternatives**

A decision maker can use the 2D coordinate system depicted in Figure 10 to retrieve essential information about the decision space, that is: i) the available decision alternatives (apartments), ii) its attribute combinations and iii) the relations among the apartment attributes. Decision makers can interpret the distance between alternatives (represented by the red dots in Figure 10) in the following way. The closer the points of two alternatives, the more similar are the corresponding apartments. The distance between alternatives and attributes (represented by the cyan dots in Figure 10) is interpretable in a similar manner. The closer a particular alternative to a particular attribute, the better the alternative fulfills that attribute (e.g., apartment 4 has the highest number of rooms and is hence plotted most closely to the attribute rooms). Finally, decision makers can infer conflicting relations among attributes by interpreting the distances between pairs of attributes. A greater distance indicates more conflict between two attributes (e.g., furnishing and rooms), and thus indicates a low probability of finding an alternative that fulfills both attributes equally.

Reducing the number of dimensions causes an information loss. We define this information loss as the amount of variance lost when reducing dimensions. Or in other words, the relations among the decision alternatives and attributes in the visualization may not perfectly coincide with the relations in the original decision space. Since each dimension of the decision space originally represents one attribute, the information loss is primarily affected by the interplay of the number of attributes that constitute the decision space and the visualization format (2D vs. 3D). The information loss increases with an increasing number of original dimensions and a decreasing number of extracted visual dimensions. This forms an interesting interaction between the complexity of a decision space (i.e., its dimensionality) and the dimensionality of the visualization generated to plot the decision space. The next section describes the laboratory experiment we conducted to investigate this interaction.

## 4.4 Empirical Investigation

The complexity of a decision making scenario and the dimensionality of a visualization as well as the interaction of both parameters are likely to affect the quality of the decision support which the particular visualization can provide. On the one hand, 2D coordinate systems have a limited ability to display information compared to 3D coordinate systems, since they have one fewer dimension to display relations among attributes and decision alternatives. The information loss will thus always be higher for 2D than for 3D coordinate systems. On the other hand, 2D coordinate systems are less complex to understand than 3D coordinate system especially if they are plotted on a 2D screen (Dull and Tegarden 1999). Whereas the complexity to understand a 2D or 3D coordinate system is independent from the dimensionality of the decision space, the information loss of a visualization does also depend on the dimensionality of the decision space (i.e., the number of attributes describing the decision space). We thus compare 2D and 3D coordinate systems with a varying dimensionality of the decision space in a laboratory experiment. The next sections describe the research methodology used for our empirical investigation, the treatments, measures and the experimental procedure.

### 4.4.1 Research Methodology

We conducted a laboratory experiment examining a consumer decision making situation with a 2x2 between-subjects design to investigate the interaction of visualization formats (i.e., the number of visual dimensions) and decision complexity (i.e., the number of dimensions in the decision space). Specifically, we tested two visualization formats (2D vs. 3D) and two dimensional levels of the decision space (4 vs. 8 attributes) to distinct between simple and complex consumer decision making scenarios. In order to investigate the interaction of visualization format and decision space complexity, we describe products with four attributes to create a simple situation in which the comparison of products is easy. Since our experiment uses a consumer decision making scenario, we derive complex situations from the level of complexity that is prevalent in consumer decision making scenarios. Recent studies indicate that consumers consider up to eight attributes when making a purchase decision (Jacoby et al. 1977; Moorthy et al. 1997; Olson and Jacoby 1972; Sheluga et al. 1979). We thus use eight attributes to create a complex situation.

During the experiment, the participants had to complete a certain search task: They were instructed to use a decision support system to find a digital camera matching their individual preferences. We decided to use a student sample for convenience reasons because there is no a priori reason why students should behave differently in such a setting than other participants and we are interested in the differences between the treatment groups and not in absolute performance and perception values.

### 4.4.2 Experimental Treatments

We developed four decision support systems as treatments for the laboratory experiment. To ensure that variations in the dependent variables are only referable to modifications in the amount of attributes considered (dimensionality of the decision space) and the visualization format, we constructed the systems as similar as possible.

All systems had access to the same product data base. The systems operated on a database with 131 cameras collected from Amazon.com. 79 of them were non-dominated (i.e., there is no camera that is better than the focal camera in at least one attribute and at least equally good in all other attributes). In the treatments with simple decision space, the four attributes photo resolution, optical zoom, camera size, and price were used to describe the cameras. Complex decision spaces additionally consisted of the attributes display resolution, video resolution, number of settings and photosensitivity.

We designed the decision support systems in conformity with decision making principles described in previous research. In line with the two stages of a purchase decision process (Gilbride and Allenby 2004; Hauser and Wernerfelt 1990; Payne 1976), we assume that consumers screen the alternatives using non-compensatory strategies in a first step, then employ compensatory decision rules for a more detailed evaluation in a second step. We base the compensatory (second) step on multi-attribute utility theory which is frequently applied to model and analyze multi-attribute decision scenarios (Keeney and Raiffa 1993; Dyer et al. 1992; Wallenius et al. 2008). The basic idea of this approach is that the overall value $v(x)$ of a decision alternative is the sum of the single-attribute values

$v_i(x_i)$ of the outcome $x_i$ of each attribute $i$, weighted by the importance $w_i$ of each attribute (see Equation (29)):

$$v(x) = \sum_{i=1}^{n} w_i v_i(x_i) \tag{29}$$

where $0 \leq w_i \leq 1$, and $\sum_{i=1}^{n} w_i = 1$. We used linear single attribute value functions $v_i(x_i)$ with a value of 0 for the worst attribute level and a value of 1 for the best attribute level. In a recent study, Scholz et al. (2015) cite evidence that linear functions appropriately describe the relationship between the camera attributes used in our experiment and the values participants assign them.

The systems all allowed weighing the attributes after the initial filtering process, using a nine-point scale. We used horizontal sliders initially set to 5 (middle), and participants could increase (decrease) an attribute's importance by moving the slider to the right (left). Participants thus stated attribute importance weights with direct rating, a method that has been found to be user friendly and highly accurate compared to other methods (Pöyhönen and Hämäläinen 2001). The underlying alternative evaluation model was an additive multi-attribute value model as defined in Equation (29).

All systems filtered camera brand, optical zoom and price in the first (non-compensatory) step. Participants then could state their attribute importance in the second step. Using the filtered subset of alternatives, each system generates a visual representation of the decision space with either two or three visual dimensions. Attributes are represented by circles (2D coordinate system) or bullets (3D coordinate system). A small distance between the attributes indicates a consonant relationship whereas a large distance indicates a conflicting relationship. Rectangles with numbers symbolize the alternatives. We provide representative screenshots in Appendix A.

The only differences between the four systems are then as follows: System 1 and system 2 both operate on a decision space described by four attributes (photo resolution, optical zoom, camera size, and price). They, however, differ in the visualization format: System 1 uses a 2D coordinate system whereas system 2 visualizes the decision space in a 3D coordinate system. Systems 3 and 4 differ from systems 1 and 2 in the number of attributes that describe the decision space. Cameras of systems 3 and 4 are described by eight attributes (photo resolution, optical zoom, camera size, price, display resolution, video resolution, number of settings, and light sensitivity). System 3 plots the decision space in a 2D coordinate system and system 4 in a 3D coordinate system.

Comparing system 1 (2D / 4 attributes) to 3 (2D / 8 attributes) and system 2 (3D / 4 attributes) to 4 (3D / 8 Attributes) enables us to isolate the effect of the decision space complexity on the dependent variables. Comparing system 1 (2D / 4 attributes) to 2 (3D / 4 attributes) and system 3 (2D / 8 attributes) to 4 (3D / 8 Attributes) enables us to analyze the impact of the visualization format (2D vs. 3D) on our dependent variables. Our dependent variables are described in the next section.

#### 4.4.3 Measures

We evaluate the decision support of 2D and 3D coordinate systems in simple and complex consumer decision making scenarios based on the framework by Xiao and Benbasat (2007). The measures fall into two categories: decision making performance and user perceptions.

Decision making performance refers to the ability of a decision support system to support a user's decision making process and encompasses decision effort and decision quality. Xiao and Benbasat (2007) define decision effort as the amount of effort incurred by a user in terms of processing information, comparing alternatives and making a final decision. We operationalize the decision effort of individual users by measuring the time needed from the start of the product search until a final decision is made.

Decision quality is evaluated either objectively or subjectively. Objective decision quality is based on the principle of coherence and typically measured in terms of users' choices of non-dominated[17] alternatives (Häubl and Trifts 2000; Payne et al. 1993).

Subjective decision quality, in contrast, is evaluated relatively to a user's preferences and is defined as the degree to which the alternatives considered for purchase match a user's preferences (Xiao and Benbasat 2007). To measure the subjective decision quality, our participants stated their probability of choosing each of the top 10 recommended alternatives. We compare the participants' evaluations of the top 10 recommended alternatives with the values the decision support system computed for each of the top ten alternatives based on Equation (29). This enables us to calculate three measures of subjective decision quality: i) the first-choice hit rate, which equals the fraction of cases where the system correctly predicts the user's most preferred alternative on rank 1 of the recommendation list (Johnson et al. 1989), ii) the user's rating of the alternative the system predicts on rank 1 and iii) the average probability of choosing each of the top 10 recommended alternatives. Applications of these decision quality measures can also be found in prior studies (Aksoy et al. 2011; Johnson et al. 1989; Diel and Zauberman 2005).

Modifications of a decision making process may also impact the users' perceptions, that is how users subjectively evaluate the support or interaction with a decision support system during their decision making process. We focus on major information system success indicators: perceived ease of use

[17] An alternative *X* is dominated by another alternative *Y* if *Y* is better than *X* in at least one attribute and not worse than *X* in the remaining attributes (Pfeiffer and Scholz 2013).

(PEOU), perceived usefulness (PU), end user's satisfaction (EUS), reuse intention (RI), and a net promoter score (NPS). PEOU and PU originate from the technology acceptance model (Davis 1989). PEOU is a user's believe that system usage is free of effort, while PU is a user's evaluation of how useful the system was in supporting the process of decision making. To determine their satisfaction, users trade off the costs and benefits of system use, which may result in either positive or negative evaluations (Xiao and Benbasat 2007). RI (Davis 1989; Igbaria 1997) covers the user's desire to keep using the decision support system in the future, after their initial use (Al-Natour et al. 2010; Venkatesh and Davis 2000). Finally, the NPS (Reichheld 2003) is an often-applied management metric that provides an alternative to traditional customer satisfaction measures. It can gauge customer loyalty and their propensity to engage in word-of-mouth activity.

We measured decision making performance and user perceptions by logging our participants' behavior and by a two-staged questionnaire. All the user perception constructs were measured with multiple items, to ensure a thorough assessment (Churchill 1979). We used seven-point Likert scales, ranging from 1 ("I fully disagree") to 7 ("I fully agree"). We took the PU and PEOU measures from Davis (1989), EUS from Au et al. (2008), and the RI measure from Al-Natour et al. (2010). For the NPS, we relied on the work by Reichheld (2003).

#### 4.4.4 Procedure

We invited undergraduate and graduate students from the University of Passau, who might be interested in camera offers and assigned them randomly to one of the four treatments. A short video explained the functionality of the decision support system and described the experimental task (see Appendix B). In the beginning of the experiment, the respondents were instructed that they were looking for a new digital camera. Every decision support system used the same database, graphical layout and navigational elements, so any differences in usage behavior, prediction accuracy, and user perceptions result from the experimental treatment – namely, the underlying decision space and the visualization dimensionality.

All systems logged detailed user behavior, and then the respondents completed a two-part questionnaire[18] after finishing their search. First, they indicated their visit probability for the top 10 recommended digital cameras. In all systems, the results appeared ordered by their estimated utility values in a separate list. Second, they completed a questionnaire consisting of items related to their demographics, psychographics, experience level, and usage perceptions.

[18] The questionnaire used in the experiment is available upon request.

## 4.5 Analysis and Results

The 112 student participants received € 7 (approximately US$ 10 at the time of the experiment) in compensation for their participation. We distributed them evenly across the different treatments (25 in the 2D & four attribute condition, 34 for 3D & four attributes, 27 for 2D & eight attributes, and 26 for 3D & eight attributes). We conducted the laboratory experiment at the University of Passau (Germany) with one instructor running all sessions.

There were slightly more female respondents (51.8%), and the average age was 22.5 years (*SD* = 2.94; *min* = 18, *max* = 31). All constructs were highly reliable with Cronbach's alphas of at least .89. The average time spent browsing for the camera (excluding the experimental instructions and a short learning phase) was about 10 minutes.

### 4.5.1 Decision Making Performance

In order to analyze differences in the decision making performance among the treatments, we used a Gamma regression with respect to search time, a Poisson regression for the number of dominated alternatives considered, a Logit regression for the first choice hit rate, and an Ordered-Logit regression for the rating of the best expected alternative and the mean probability of purchase. We observed that subjects needed significantly more time ($p < .001$) when they had information on eight instead of four attributes (see Table 21). This is unsurprising taking into account that more attribute weightings have to be specified. Further, we found moderate but almost always insignificant decision quality differences between the 2D and the 3D visualizations. Table 21 summarizes these results, which are in line with other empirical investigations that also found no differences between 2D and 3D visualizations (e.g., Zhu and Chen 2005). The advantage of lower information loss when using a 3D visualization seems thus negligible if alternatives are described by only a few attributes. Furthermore, 3D visualizations on a 2D panel (monitor) are not as easy to interpret (Dull and Tegarden 1999) and might hence prevent a better decision making performance.

**Table 21: Effects of Search Time and Decision Quality**

| Treatment | Average Search Time in Minutes (SD) | | # of Dominated Alternatives Considered (SD) | | First Choice Hit Rate | Rating of the Best Expected Alternative | Mean Probability of Purchase |
|---|---|---|---|---|---|---|---|
| 2D / 4 Attributes | 8.24 | (2.97) | 1.76 | (1.16) | 48.0% | 6.28 | 3.50 |
| 3D / 4 Attributes | 7.51 | (3.82) | 2.06 | (1.65) | 55.9% | 6.44 | 3.86 |
| 2D / 8 Attributes | 11.85*** | (4.83) | 1.59 | (1.08) | 51.8% | 6.07 | 4.09 |
| 3D / 8 Attributes | 10.09* | (2.82) | 1.65 | (1.16) | 57.7% | 6.15 | 3.83 |

*Note. * p<.1, ** p<.05, *** p<.01.*

Concerning decision quality, we found no significant differences in the consideration set sizes of our treatments (2D/4 Attributes: 5.32, 3D/4 Attributes: 5.94, 2D/8 Attributes: 5.85, 3D/8 Attributes: 5.50), the number of dominated alternatives, the first choice hit rate, the rating of the best expected alternative and the mean probability of purchase. Taking into account that it is much harder to correctly predict a decision maker's preference in situations where alternatives are more complex, it is surprising at first sight that the predictive validity in terms of first choice hit rate was higher when eight instead of four attributes were used. However, several alternatives might exist that especially differ in the hidden attributes. Products described with fewer attributes are more homogeneous making it more difficult for a decision support system to correctly rank the alternatives. Our results thus indicate no differences in either objective or subjective decision quality between the 2D and 3D visualizations.

### 4.5.2 User Perceptions

The effects of the systems on users' perceptions offer novel and interesting insights (Figure 11). If the decision space is simple, 2D visualization outperforms 3D visualization in terms of perceptions. As the complexity increases, perceptions of the 2D visualization decrease, whereas perceptions of the 3D visualization increase.

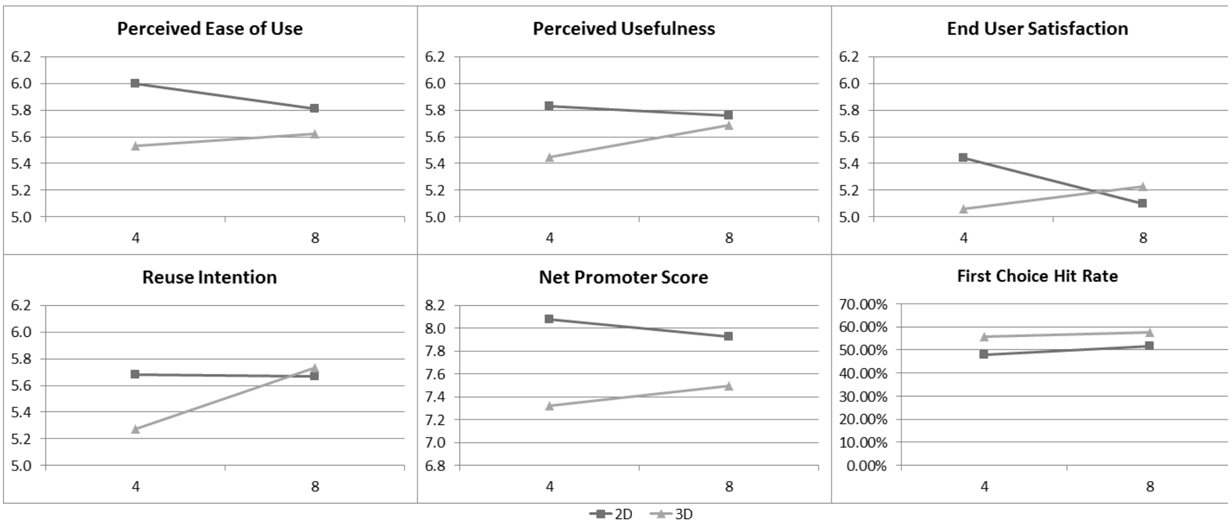



**Figure 11: Influence of Systems on User Perceptions and First Choice Hit Rate by Number of Attributes**

When the decision space consisted of four attributes, 2D visualization appeared significantly more useful ($p < .05$) than 3D visualization, including the weakly significant, higher EUS ($p < .1$) and RI ($p < .1$; Figure 11). The 2D visualization also attracted significantly more promoters ($p < .05$), though the number of detractors ($p > .4$) and the PEOU ($p > .3$) were not affected by the visualization method. Therefore, users preferred the 2D visualization when the number of attributes was reasonably small

– similar to the results reported by Tractinsky and Meyer (1999). For more complex products, this difference diminished in all user perception measures and became insignificant (*p* > .2). Specifically, perceptions of 2D visualization decreased with the number of attributes, but perceptions of 3D visualization increased with a growing number of attributes in the decision space.

## 4.6 Discussion

### 4.6.1 Implications

Decision makers often have to deal with cognitively challenging decision making scenarios since information on many product alternatives and attributes need to be processed to arrive at a final choice (Xiao and Benbasat 2007). This information is often complex because consumers have to take conflicting attribute relations into account, meaning that a preference for one attribute may cause losses with respect to other and eventually also desirable attributes. In complex consumer decision making scenarios, consumers take up to eight attributes into account and thus need to be aware of up to 28 attribute relations. Previous literature recognizes that knowing the decision space cannot be presumed and indicates that: i) knowing the decision space and getting familiar with the relations among attributes can help decision makers to construct more stable preferences which may ultimately lead to improved decision making performance (Butler et al. 2008; Keeney 2002; Huber and Klein 1991; Hoeffler and Ariely 1999) and ii) presenting visual information about decision spaces might be an appropriate way to provide consumers the required information while preventing them from information overload (Lurie and Manson 2007; Turetken and Sharda 2001).

Our literature review reveals an important gap in prior research regarding the question in which format information for decision support should be visualized. We find that: i) there is mixed evidence on the question whether 2D or 3D visualizations provide superior support for decision makers and ii) the majority of the studies rather evaluated decision support in terms of decision effort and quality thus neglecting decision makers' perceptions of the systems that assist during their decision making. In this study, we shed light on this research gap by investigating the decision support of 2D and 3D coordinate systems in simple and complex consumer decision making scenarios. We evaluate the decision support using both decision making performance and decision makers' perceptions.

To evaluate the interplay of decision making complexity and the decision support of the visualization format, we conducted a laboratory experiment on a consumer decision making task. We tested four decision support systems that provided decision makers with visual information: a 2D and a 3D coordinate system in a simple consumer decision making scenario (decision space consists of four attributes) and in a complex scenario (decision space consists of eight attributes). We demonstrated that a 2D visualization is perceived as superior to 3D when the decision objects are described by only a few attributes. We do not observe this difference in terms of objective performance measures such as the time needed to make a decision or the decision quality. Interestingly, our results thus indicate that decision making performance and user perceptions are not necessarily correlated and that an evaluation of visual decision support should consider both the users' perceptions and the users' behavior.

Our results offer implications for research and practice in terms of consumer decision makers and providers of decision support systems, such as online recommender systems. The comparison of the 2D and 3D visualizations revealed no significant differences in terms of decision quality. However, 2D visualizations prompted greater perceptions of ease of use and usefulness, as well as higher satisfaction, when only a few attributes were depicted. With eight attributes though, we found no significant effect of user perceptions between 2D and 3D visualization. We recommend using 2D visualizations only if the decision space is described by few attributes. These results also indicate that providers of decision support systems, such as online retailers providing recommender systems, should evaluate their systems based on decision makers' observable behavior (e.g., consideration set size, choice probability, time to make a decision) and decision makers' perceptions (e.g., perceived ease of use, end user satisfaction).

This study extends prior research by a novel finding. The complexity of a decision situation moderates the evaluation of a visualization format (2D vs. 3D), but only those evaluation measures that reflect decision makers' perceptions. Existing research has mainly focused on decision makers' observational behavior and found no clear support for any visualization format in most cases. Our results support this finding, but also demonstrate that the decision makers' perceptions differ from observational data and make a 2D visualization superior if the decision situation is rather simple.

### 4.6.2 Limitations and Future Research

This research is subject to some limitations that we summarize in the following. One limitation is the use of a convenience sample in our laboratory experiment. However, there is no a priori reason why students should behave differently in such a setting than a representative sample. Each participant might also have eliminated a different number of alternatives which leads to different visual projections of the decision space. However, the number of alternatives that remains in the decision space after the initial filtering step is not significantly different across the treatments on average (tested with ANOVA, F=0.604, $p$ = 0.614). We thus assume that the initial filtering has had approximately the same effect on all treatments. Furthermore, we provided rank numbers for the recommended decision alternatives when observing consumer choices. Providing predictive ratings tends to anchor users' evaluations of the recommendations (Cosley et al. 2003). Since we provided predicted ranks in all four treatments, the effect is hence equal for all treatments and we have no indication that this effect has changed the differences between the treatments in terms of the system quality or the users' perceptions. We believe these limitations do not limit the generalizability of our results, especially since we are interested in the differences between treatments instead of absolute values.

We also define complex decision making scenarios by the maximum number of product attributes considered when making a purchase decision. In managerial decision making scenarios, it is possible that even more than eight attributes are considered by a decision maker. The implications of our paper are restricted to the level of complexity that is found in consumer decision making scenarios. Higher levels of complexity are beyond the scope of this work but provide avenues for future research.

Finally, depicting high dimensional decision spaces in 2D or 3D visual representations requires reducing the dimensionality, which might lead to information loss and thus erroneous interpretations. This is especially critical in case of higher information loss. We calculated the information loss as percentage of the variance that is not covered by the two or three dimensions that were used for the visualization. We found an information loss of 20.6% on average across our treatments and a maximal information loss of 35.0% indicating that the visualizations presented the decision space rather accurate. It should, however, be mentioned that low dimensional visualizations of high dimensional decision spaces may help decision makers to get information on the available alternatives, their attribute combinations and potentially conflicting attribute relations. When it comes to the final choice among decision alternatives, the visual representations of the decision space used in this study cannot replace a comparison of the real attribute values of the decision alternatives. Our decision support systems thus listed all cameras depicted in the visual representation of the decision space in a separate panel. Each camera is here described with its levels of all four/eight attributes.

## 4.7 Conclusion

The results of our study suggest an interesting trade-off: 3D visualizations preserve more of the original information about the decision space than 2D visualizations. However, users often express some initial reluctance toward 3D visualizations, apparently because of the complex interpretations it demands, including information occlusion and ambiguous depth judgments (Kumar and Benbasat 2004). From a managerial point of view, our results suggest the use of 2D visualization when the number of attributes is rather small, which implies only moderate information loss. If the number of considered attributes increases, 3D visualization becomes more attractive, because the information loss is now more relevant. The lower information loss of 3D visualizations slightly improves decision making performance, and ultimately enhances user perceptions. People seem to prefer 2D visualizations because it is easier to understand, but the advantages of 3D visualizations can compensate for this initial reluctance with better results, which ultimately gets reflected in terms of user perceptions. Figure 12 illustrates this classification of visualization methods in the trade-off between information loss and user perceptions.

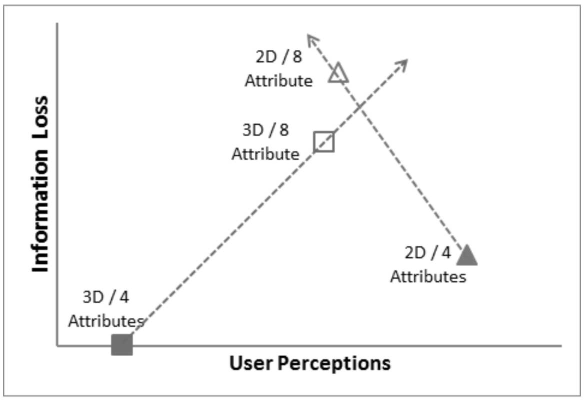
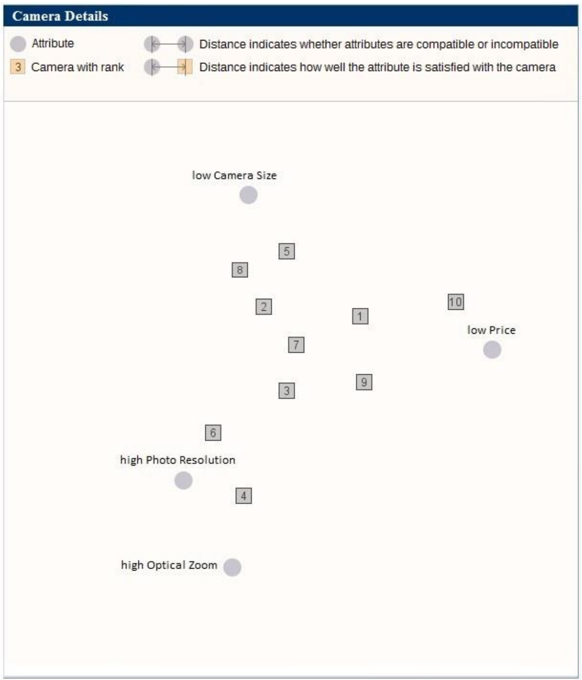


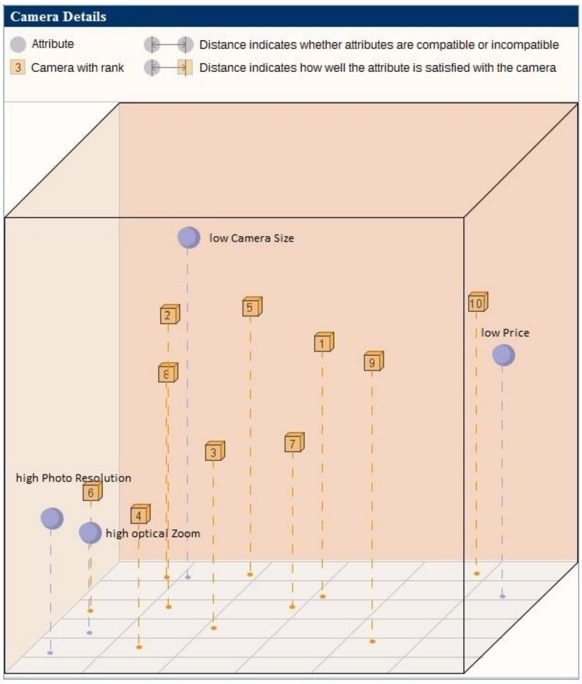

**Figure 12: Classification of Visualizations in the Trade-Off between Information Loss and User Perceptions**

## 4.8 Appendix

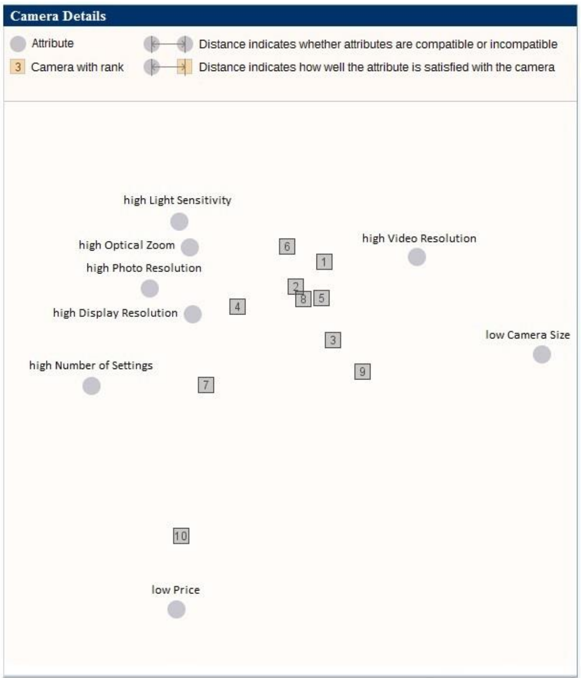
### 4.8.1 Appendix A: Treatments



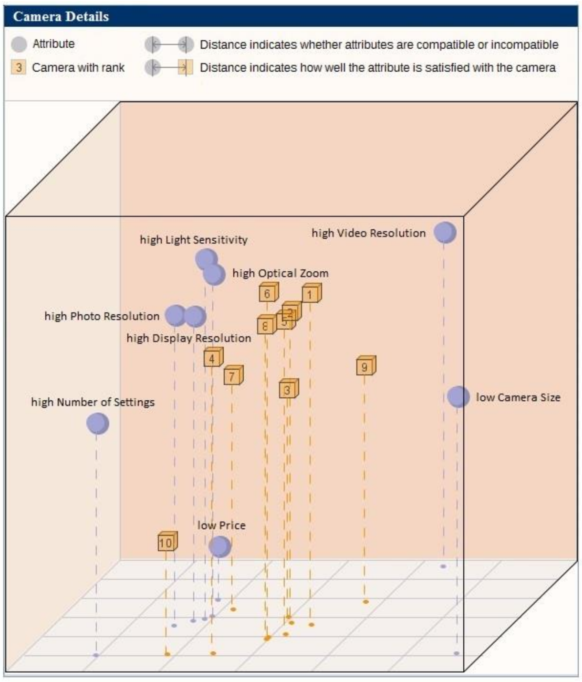

System 1: 2D / 4 Attributes




System 2: 3D / 4 Attributes




System 3: 2D / 8 Attributes




System 4: 3D / 8 Attributes

**Figure 13: Screenshots of the Systems**

### 4.8.2 Appendix B: Experimental Instructions

The following paragraphs contain the experimental instructions for system 1 (2D / 4 attributes). The modifications of the instructions for systems 2 to 4, i.e. when the visualization contained information on eight instead of four attributes or the visualization format was 3D instead of 2D, can be found in the brackets.

“Welcome and thank you for your attendance. Today, each of you will use a recommendation system to search for a digital camera matching your individual preferences. Now, we will briefly explain how the recommendation system works. The search for a digital camera will proceed in two steps. In the first step, digital cameras that do not meet your preferences can be filtered out upfront on the initial page. According to the default settings, the recommendation system searches for suitable digital cameras among all available brands. You can also restrict the search on certain brands by selecting them from the list on the left. Additionally, you can exclude digital cameras from the search results that have, for example, an undesirable zoom factor or an undesirable price. To do so, just set an upper and/or lower limit for the corresponding attributes or leave these fields empty when you do not want to set a restriction. By clicking the "Forward" button, you will be forwarded to a second search page.

On this next page, a list will display the digital cameras that meet your search criteria on the right side of the interface. The digital cameras are described by the attributes that could be restricted in the previous step. You can sort the digital cameras according to your personal preferences. Use the sliding controllers on the left to specify the importance of each of the four (eight) attributes "Photo Resolution", "Optical Zoom", "Camera Size" and "Price" (for 8 attributes additional: "Display Resolution", "Video Resolution", "Number of Settings" and "Light Sensitivity") that describe the digital cameras.

For a better understanding of the remaining digital cameras, a visualization containing information about four (eight) camera attributes and the first ten search results is displayed in the middle of the interface. Digital cameras are displayed as rectangles (3D: cubes). The particular camera attributes are depicted as circles (3D: bullets), where the size of a circle (3D: bullet) represents the importance that is adjusted for that particular attribute. Further, the proximity between attribute circles (3D: bullets) can be interpreted. If two circles (3D: bullets) are depicted afar, it is unlikely that a digital camera contains both attributes at high levels. In this example the circle (3D: bullet) "Camera Size" is depicted afar from "Optical Zoom". So, digital cameras will be either small or have a high optical zoom. It is not very likely to find a small digital camera simultaneously offering a high optical zoom. The proximity between camera attributes and distinct camera rectangles (3D: cubes) can also be interpreted. A digital camera contains attributes at high levels whose corresponding circles (3D: bullets) are depicted in close proximity and low levels of attributes that are depicted afar. The camera rectangles (3D: cubes) are labeled by numbers referring to their corresponding rank in the results list on the right side. Pointing with the mouse on an attribute circle (3D: bullet) dyes attribute bullets in green that are close and thus conceivable in combination and distant and inconceivable attributes in red. Further, a small popup appears beside the selected attribute circle (3D: bullet) where conceivable and inconceivable attributes are also listed. If you point with the mouse on a camera rectangle (3D: cube), a popup appears beside the corresponding camera where all available camera attributes are listed.

Finally, if you want to inspect more than the ten top-rated digital cameras, a click on the "Forward" button at the bottom of the interface shows the next ten search results. If you further want to reset your search specifications on the first page, please use the link at the top and not your browser's backwards button. As soon as you found a suitable digital camera and finished your search, please specify the likelihood that you would purchase each of the first ten digital cameras on the list. Use the sliding controller below to state your purchasing probability for the corresponding digital camera. And finally, please click on the button labeled "Survey" displayed at the bottom of the interface. You may now switch on the monitors and start your search for a digital camera.”

## References


Abbas, A. / Bell, D. (2012). One-switch conditions for multiattribute utility functions, Operations Research, 60(5), 1199-1212.

Abbas, A. / Bell, D. (2011). One-switch independence for multiattribute utility functions, Operations Research, 59(3), 764-771.

Abbasi, A. / Chen, H. (2008). CyberGate: A design framework and system for text analysis of computer-mediated communication, MIS Quarterly, 32(4), 811-837.

Adomavicius, G. / Bockstedt, J.C. / Curley, S.P. / Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects, Information Systems Research, 24(4), 956-975.

Aksoy, L. / Cooil, B. / Lurie, N.H. (2011). Decision quality measures in recommendation agents research, Journal of Interactive Marketing, 25(2), 110-122.

Al-Natour, S. / Benbasat, I. / Cenfetelli, R. (2010). Trustworthy virtual advisors and enjoyable interactions: Designing for expressiveness and transparency, 18th European Conference on Information Systems.

Aliguliyev, R.M. (2009). Performance evaluation of density-based clustering methods, Information Sciences, 179(20), 3583-3602.

Alter, S. (2004). A work system view of DSS in its fourth decade, Decision Support Systems, 38(3), 319-327.

Amazon (2016). http://www.amazon.de, latest access: May 31, 2016.

Ananthanarayanan, R. / Chenthamarakshan, V. / Deshpande, P. M. / Krishnapuram, R. (2008). Rule based synonyms for entity extraction from noisy text, 2nd ACM Workshop on Analytics for Noisy Unstructured Text Data, 31-38.

Ansari, A. / Essegaier, S. / Kohli, R. (2000). Internet recommendation systems, Journal of Marketing Research, 37(3), 363-375.

Arasu, A. / Ganti, V. / Kaushik, R. (2006). Efficient exact set-similarity joins, 32nd International Conference on Very Large Data Bases, 918-929.

Arasu, A. / Kaushik, R. (2009). A Grammar-based entity representation framework for data cleaning, Proceedings of the 35th ACM SIGMOD International Conference on Management of Data, 233-244.

Au, N. / Ngai, E.W. / Cheng, T.C.E. (2008). Extending the understanding of end user information systems satisfaction formation: An equitable needs fulfillment model approach, MIS Quarterly, 32(1), 43-66.

Backhaus, K. / Wilken, R. / Voeth, M. / Sichtmann, C. (2005). An empirical comparison of methods to measure willingness to pay by examining the hypothetical bias, International Journal of Market Research, 47(5), 543-562.

Bajari, P. / Hortaçsu, A. (2003). The winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions, RAND Journal of Economics, 34(2), 329-355.

Bapna, R. / Chang, S.A. / Goes, P. / Gupta, A. (2009). Overlapping online auctions: Empirical characterization of bidder strategies and auction prices, MIS Quarterly, 33(4), 763-783.

Bapna, R. / Goes, P. / Gupta, A. (2003). Replicating online Yankee auctions to analyze auctioneers' and bidders' strategies, Information Systems Research, 14(3), 244-268.

Bapna, R. / Goes, P. / Gupta, A. / Jin Y. (2004). User heterogeneity and its impact on electronic auction market design: An empirical exploration, MIS Quarterly, 28(1), 21-43.

Bapna, R. / Jank, W. / Shmueli, G. (2008). Consumer surplus in online auctions, Information Systems Research, 19(4), 400-416.

Barrot, C. / Albers, S. / Skiera, B. / Schäfers, B. (2010). Vickrey vs. ebay: Why second-price sealed-bid auctions lead to more realistic price-demand functions, International Journal of Electronic Commerce, 14(4), 7-38.

Becker, J.-M. / Rai, A. / Ringle, C.M. / Völckner, F. (2013). Discovering unobserved heterogeneity in structural equation models to avert validity threats, MIS Quarterly, 37(3), 665-694.

Benbasat, I. / DeSanctis, G. / Nault, B.R. (1991). Empirical research in managerial support systems: A review and assessment, in Recent Developments in Decision Support Systems, Springer, 383-437.

Bhattacharya, I. / Getoor, L. (2007). Collective entity resolution in relational data, ACM Transactions on Knowledge Discovery from Data, 1(1), 1-35.

Bohm, P. / Linden, J. / Sonnengard, J. (1997). Eliciting reservation prices: Becker–Degroot–Marschak mechanisms, The Economic Journal, 107(443), 1079-1089.

Bottomley, P. / Doyle, J. / Green, R. (2000). Testing the reliability of weight elicitation methods: Direct rating versus point allocation, Journal of Marketing Research, 37(4), 508-513.

Brans, J.-P. / Mareschal, B. (1994). The PROMCALC & GAIA decision support system for multicriteria decision aid, Decision Support Systems, 12, 297-310.

Butler, J. / Dyer, J.S. / Jia, J. / Tomak, K. (2008). Enabling e-transactions with multi-attribute preference models, European Journal of Operational Research, 186(2), 748-765.

Butler, J. / Morrice, D. / Mullarkey, P. (2001). A multiple attribute utility theory approach to ranking and selection, Management Science, 47(6), 800-816.

Butler, J. / Jia, J. / Dyer, J. (1997). Simulation techniques for the sensitivity analysis of multicriteria decision models, European Journal of Operational Research, 103(3), 531-545.

Cagnina, L. / Errecalde, M. / Ingaramo, D. / Rosso, P. (2014). An efficient particle swarm optimization approach to cluster short texts, Information Sciences, 265, 36-49.

Cao, Y. / Li, Y. (2007). An intelligent fuzzy-based recommendation system for consumer electronic products, Expert Systems with Applications, 33(1), 230-240.

Card, S.K. / Mackinlay, J.D. / Schneiderman, B. (1999). Readings in information visualization: Using vision to think, Morgan Kaufmann.

Chan, T.Y. / Kadiyali, V. / Park, Y.-H. (2007). Willingness to pay and competition in online auctions, Journal of Marketing Research, 44(2), 324-333.

Cheema, A. / Chakravarti, D. / Sinha, A.R. (2012). Bidding behavior in descending and ascending auctions, Marketing Science, 31(5), 779-800.

Chehreghani, M.H. / Abolhassani, H. / Chehreghani, M.H. (2009). Density link-based methods for clustering web pages, Decision Support Systems, 47(4), 374-382.

Chewning, E.G. / Harrell, A.M. (1990). The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task, Accounting, Organizations and Society, 15(6), 527-542.

Choi, S.H. / Cho, Y.H. (2004). An utility range-based similar product recommendation algorithm for collaborative companies, Expert Systems with Applications, 27(4), 549-557.

Churchill, G.A. (1979). A paradigm for developing better measures of marketing constructs, Journal of Marketing Research, 16(1), 64-73.

Cohen, W.W. (2000). Data integration using similarity joins and a word-based information representation language, ACM Transactions on Information Systems, 18(3), 288-321.

Coltman, T. / Gattorna, J. / Whiting, S. (2010). Realigning service operations strategy at DHL Express, Interfaces, 40(3), 175-183.

Corner, J. / Buchanan, J. (1997). Capturing decision maker preference: experimental comparison of decision analysis and MCDM techniques, European Journal of Operational Research, 98(1), 85-97.

Cosley, D. / Lam, S.K. / Albert, I. / Konstan, J. / Riedl, J. (2003). Is seeing believing? How recommender systems influence users‘ opinions, SIGCHI Conference on Human Factors in Computing Systems, 585-592.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Quarterly, 13(3), 319-340.

Dawes, R. (1979). The robust beauty of improper linear models in decision making, American Psychologist, 34(7), 571-582.

De Bruyn, A. / Liechty, J. / Huizingh, E. / Lilien, G. (2008). Offering online recommendations with minimum customer input through conjoint-based decision aids, Marketing Science, 27(3), 443-460.

Dempster, A.P. / Laird, N.M. / Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, Series B, 39(1), 1-38.

Denguir-Rekik, A. / Montmain, J. / Mauris, G. (2009). A possibilistic-valued multi-criteria decision-making support for marketing activities in e-commerce: feedback based diagnosis system, European Journal of Operational Research, 195(3), 876-888.

DeSarbo, W. / Ramaswamy, V. / Cohen, S. (1995). Market segmentation with choice-based conjoint analysis, Marketing Letters, 6(2), 137-147.

Dewally, M. / Ederington, L. (2006). Reputation, certification, warranties, and information as remedies for seller–buyer information asymmetries: Lessons from the online comic book market, The Journal of Business, 79(2), 693-729.

Dey, D. (2003). Record matching in data warehouses: A decision model for data consolidation, Operations Research, 51(2), 240-254.

Dey, D. / Sarkar, S. / De, P. (2002). A distance-based approach to entity reconciliation in heterogeneous databases, IEEE Transactions on Knowledge and Data Engineering, 14(3), 567-582.

Diehl, K. / Zauberman, G. (2005). Searching ordered sets: Evaluations from sequences under search, Journal of Consumer Research, 31(4), 824-832.

Dilla, W. / Janvrin, D.J. / Raschke, R. (2010). Interactive data visualization: New directions for accounting information systems research, Journal of Information Systems, 24(2), 1-37.

Dost, F. / Wilken, R. (2012). Measuring willingness to pay as a range, revisited: When should we care? International Journal of Research in Marketing, 29(2), 148-166.

Druehl, C. / Schmidt, G. / Souza, G. (2009). The optimal pace of product updates, European Journal of Operational Research, 192(2), 621-633.

Duan, W. (2010). Analyzing the impact of intermediaries in electronic markets: An empirical investigation of online consumer-to-consumer (C2C) auctions, Electronic Markets, 20(2), 85-93.

Dull, R.B. / Tegarden, D.P. (1999). A comparison of three visual representations of complex multidimensional accounting information, Journal of Information Systems, 13(2), 117-131.

Dyer, J.S. / Fishburn, P.C. / Steuer, R.E. / Wallenius, J. / Zionts, S. (1992). Multiple criteria decision making, multiattribute utility theory: The next ten years, Management Science, 38(5), 645-654.

Easley, R.F. / Wood, C.A. / Barkataki, S. (2010). Bidding patterns, experience, and avoiding the winner's curse in online auctions, Journal of Management Information Systems, 27(3), 241-268.

Einav, L. / Kuchler, T. / Levin, J. / Sundaresan, N. (2015). Assessing sale strategies in online markets using matched listings, American Economic Journal: Microeconomics, 7(2), 215-247.

Elfenbein, D. / Fisman, R. / McManus, B. (2012). Charity as a substitute for reputation: Evidence from an online marketplace, Review of Economic Studies, 79(4), 1441-1468.

Ester, M. / Kriegel, H.-P. / Sander, J. / Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, 2nd International Conference on Knowledge Discovery and Data Mining, 96(34), 226-231.

Fellegi, I.P. / Sunter, A.B. (1969). A theory for record linkage, Journal of the American Statistical Association, 64(328), 1183-1210.

Forman, C. / Ghose, A. / Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets, Information Systems Research, 19(3), 291-313.

Frischmann, T. / Hinz, O. / Skiera, B. (2012). Retailers' use of shipping cost strategies: Free shipping or partitioned prices?, International Journal of Electronic Commerce, 16(3), 65-87.

Ge, Y. / Xiong, H. / Tuzhilin, A. / Liu, Q. (2014). Cost-aware collaborative filtering for travel tour recommendations, ACM Transactions on Information Systems, 32(1), 4.

Gena, C. / Weibelzahl, S. (2007). Usability Engineering for the Adaptive Web, in The adaptive web, Springer, 720-762.

Gensch, D. / Soofi, E. (1992). A minimum discrimination information estimation of multiattribute market share models, Marketing Science, 11(1), 54-63.

Gensler, S. / Hinz, O. / Skiera, B. / Theysohn, S. (2012). Willingness-to-pay estimation with choice-based conjoint analysis: Addressing extreme response behavior with individually adapted designs, European Journal of Operational Research, 219(2), 368-378.

Ghose, A. / Telang, R. / Krishnan, R. (2005). Effect of electronic secondary markets on the supply chain, Journal of Management Information Systems, 22(2), 91-120.

Gilbride, T. / Allenby, G. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules, Marketing Science, 23(3), 391-406.

Gilkeson, J.H. / Reynolds, K. (2003). Determinants of Internet auction success and closing price: An exploratory study, Psychology & Marketing, 20(6), 537-566.

Goes, P. / Tu, Y. / Tung, A. (2013). Seller heterogeneity in electronic marketplaces: A study of new and experienced sellers in eBay, Decision Support Systems, 56, 247-258.

Green, P. / Krieger, A. / Agarwal, M. (1993). A cross validation test of four models for quantifying multiattribute preferences, Marketing Letters, 4(4), 391-406.

Green, P. / Krieger, A. / Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects, Interfaces, 31(3), 56-73.

Greenacre, M. (2007). Correspondence analysis and practice, Taylor & Francis.

Guan, S. / Ngoo, C.S. / Zhu, F. (2003). Handy broker: An intelligent product-brokering agent for m-commerce applications with user preference tracking, Electronic Commerce Research and Applications, 1(3), 314-330.

Hammond, J.S. / Keeney, R.L. / Raiffa, H. (1998). Even swaps: A rational method for making trade-offs, Harvard Business Review, 76(2), 137-150.

Härdle, W. / Simar, L. (2003). Applied multivariate statistical analysis, Springer.

Harvey, C. (1981). Conditions on risk attitude for a single attribute, Management Science, 27(2), 190-203.

Häubl, G. / Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids, Marketing Science, 19(1), 4-21.

Hauser, J. / Wernerfelt, B. (1990). An evaluation cost model of consideration sets, Journal of Consumer Research, 16(4), 393-408.

Hinz, O. / Eckert, J. (2010). The impact of search and recommendation systems on sales in electronic commerce, Business & Information Systems Engineering, 2(2), 67-77.

Hinz, O. / Hill, S. / Kim, J.-Y. (2016). TV's dirty little secret: The negative effect of popular TV on online auction sales, MIS Quarterly, forthcoming.

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.

Hosack, B. / Hall, D. / Paradice, D. / Courtney, J.F. (2012). A look toward the future: Decision support systems research is alive and well, Journal of the Association for Information Systems, 13(5), 315.

Hothorn, T. / Hornik, K. / Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework, Journal of Computational and Graphical Statistics, 15(3), 651-674.

Hou, J. / Blodgett, J. (2010). Market structure and quality uncertainty: A theoretical framework for online auction research, Electronic Markets, 20(1), 21-32.

Huang, S. (2011). Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods, Electronic Commerce Research and Applications, 10(4), 398-407.

Hubert, L. / Arabie, P. (1985). Comparing partitions, Journal of Classification, 2(1), 193-218.

Huber, J. / Klein, N.M. (1991). Adapting cutoffs to the choice environment: The effects of attribute correlation and reliability, Journal of Consumer Research, 18(3), 346-357.

Igbaria, M. (1997). The consequences of information technology acceptance on subsequent individual performance, Information & Management, 32(3), 113-121.

Isaacs, W. / Senge, P. (1992). Overcoming limits to learning in computer-based learning environments, European Journal of Operational Research, 59(1), 183-196.

Jacoby, J. / Szybillo, G.J. / Busato-Schach, J. (1977). Information acquisition behavior in brand choice situations, Journal of Consumer Research, 3(4), 209-216.

Jaro, M.A. (1989). Advances in record-linkage methodology as applied to matching the 1985 Census of Tampa, Florida, Journal of the American Statistical Association, 84(406), 414-420.

Jedidi, K. / Kohli, R. / DeSarbo, W. (1996). Consideration sets in conjoint analysis, Journal of Marketing Research, 33(3), 364-372.

Jedidi, K. / Zhang, Z.J. (2002). Augmenting conjoint analysis to estimate consumer reservation price, Management Science, 48(10), 1350-1368.

Johnson, E.J. / Meyer, R.J. / Ghose, S. (1989). When choice models fail: Compensatory models in negatively correlated environments, Journal of Marketing Research, 26(3), 255-270.

Jones, M.T. (2011). Bidding fever in eBay auctions of Amazon.com gift certificates, Economics Letters, 113(1), 5-7.

Kamakura, W.A. / Russell, G.J. (1993). Measuring brand value with scanner data, International Journal of Research in Marketing, 10(1), 9-22.

Karniouchina, E.V. / Moore, W.L. / van der Rhee, B. / Verma, R. (2009). Issues in the use of ratings-based versus choice-based conjoint analysis in operations management research, European Journal of Operational Research, 197(1), 340-348.

Kaufman, L. / Rousseeuw, P.J. (2005). Finding groups in data: An introduction to cluster analysis, Wiley & Sons.

Keeney, R.L. (2002). Common mistakes in making value trade-offs, Operations Research, 50(6), 935-945.

Keeney, R.L. / Raiffa, H. (1993). Decisions with multiple objectives: Preferences and value trade-offs, Cambridge University Press.

Kelton, A.S. / Pennington, R.R. / Tuttle, B.M. (2010). The effects of information presentation format on judgment and decision making: A review of the information systems research, Journal of Information Systems, 24(2), 79-105.

Kim, K. / Proctor, R.W. / Salvendy, G. (2011). Comparison of 3D and 2D menus for cell phones, Computers in Human Behavior, 27(5), 2056-2066.

Kim, H.-N. / El-Saddik, A. / Jo, G.-S. (2011). Collaborative error-reflected models for cold-start recommender systems, Decision Support Systems, 51(3), 519-531.

Kocas, C. (2002). Evolution of prices in electronic markets under diffusion of price-comparison shopping, Journal of Management Information Systems, 19(3), 99-119.

Kosslyn, S.M. (1994). Elements of graph design, H. Freeman.

Kreng, V.B. / Wu, C.-Y. (2007). Evaluation of knowledge portal development tools using a fuzzy AHP approach: The case of Taiwanese stone industry, European Journal of Operational Research, 176(3), 1795-1810.

Kumar, N. / Benbasat, I. (2004). The effect of relationship encoding, task type, and complexity on information presentation: An empirical evaluation of 2D and 3D line graphs, MIS Quarterly, 28(2), 255-281.

Larsen, M.D. / Rubin, D.B. (2001). Iterative automated record linkage using models mixture, Journal of the American Statistical Association, 96(453), 32-41.

Lee, W.P. (2004). Towards agent-based decision making in the electronic marketplace: Interactive recommendation and automated negotiation, Expert Systems with Applications, 27(4), 665-679.

Lee, J.M. / MacLachian, J. / Wallace, W.A. (1986). The effects of 3D imagery on managerial data interpretation, MIS Quarterly, 10(3), 257-269.

Leeflang, P.S.H. / Wittink, D.R. (1992). Diagnosing competitive reactions using (aggregated) scanner data, International Journal of Research in Marketing, 9(1), 39-57.

Li, X. / Morie, P. / Roth, D. (2005). Semantic integration in text: From ambiguous names to identifiable entities, AI Magazine, 26(1), 45-58.

Li, Y. / Luo, C. / Chung, S.M. (2008). Text clustering with feature selection by using statistical data, IEEE Transactions on Knowledge and Data Engineering, 20(5), 641-652.

Lim, E.-P. / Srivastava, J. / Probhakar, S. / Richardson, J. (1993). Entity identification in database integration, 9th IEEE International Conference on Data Engineering, 294-301.

Liu, D.R. / Shih, Y.Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value, Information & Management, 42(3), 387-400.

Liu, Y. / Sutanto, J. (2012). Buyers' purchasing time and herd behavior on deal-of-the-day group-buying websites, Electronic Markets, 22(2), 83-93.

Lohse, G.L. (1997). The role of working memory on graphical information processing, Behavior and Information Technology, 16(6), 297-308.

Louviere, J.J. / Islam, T. (2008). A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling, Journal of Business Research, 61(9), 903-911.

Lucking-Reiley, D. (1999). Using field experiments to test equivalence between auction formats: Magic on the Internet, American Economic Review, 89(5), 1063-1080.

Lucking-Reiley, D. / Bryan, D. / Prasad, N. / Reeves, D. (2007). Pennies from eBay: The determinants of price in online auctions, The Journal of Industrial Economics, 55(2), 223-233.

Lurie, N.H. (2004). Decision making in information rich environments: The role of information structure, Journal of Consumer Research, 30(4), 473-486.

Lurie, N.H. / Manson, C.H. (2007). Visual representation: Implications for decision making, Journal of Marketing, 71(1), 160-177.

Mallach, E.G. (2000). Decision support and data warehouse systems, McGraw-Hill Higher Education.

Manning, C.D. / Raghavan, P. / Schuetze, H. (2009). Introduction to information retrieval, Cambride University Press.

Manouselis, N. / Costopoulou, C. (2007). Experimental analysis of design choices in multiattribute utility collaborative filtering, International Journal of Pattern Recognition and Artificial Intelligence, 21(2), 311-331.

March, J. (1978). Rationality, ambiguity, and the engineering of choice, Bell Journal of Economics, 9(2), 587-608.

Meeds, R. (2004). Cognitive and attitudinal effects of technical advertising copy: The roles of gender, self-assessed and objective consumer knowledge, International Journal of Advertising, 23(3), 177-192.

Miller, K. / Hofstetter, R. / Krohmer, H. / Zhang, Z. (2011). How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches, Journal of Marketing Research, 48(1), 172-184.

Mingfeng, L. / Lucas Jr., H.C. / Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem, Information Systems Research, 24(4), 906-917.

Moore, W. (2004). A cross-validity comparison of ratings-based and choice-based conjoint analysis models, International Journal of Research in Marketing, 21(3), 299-312.

Moore, W. / Gray-Lee, J. / Louviere, J. (1998). A cross-validity comparison of ratings-based and choice-based conjoint analysis models, Marketing Letters, 9(2), 195-208.

Moorthy, S. / Rathchford, D. / Takudar, D. (1997). Consumer information search revisited: Theory and empirical analysis, Journal of Consumer Research, 23(4), 263-277.

Mudambi, S.M. / Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com, MIS Quarterly, 34(1), 185-200.

Mumtaz, A. / Coviello, E. / Lanckriet, G.R.G. / Chan, A.B. (2007). Clustering dynamic textures with the hierarchical EM algorithm for modeling video, IEEE Transactions on Pattern Analysis & Machine Intelligence, 35(7), 1606-1621.

Mustajoki, J. / Hämäläinen, R. (2007). Smart-swaps — A decision support system for multicriteria decision analysis with the even swaps method, Decision Support Systems, 44(1), 313-325.

Nah, F.F.H. / Eschenbrenner, B. / DeWester, D. (2011). Enhancing brand equity through flow and telepresence: A comparison of 2D and 3D virtual worlds, MIS Quarterly, 35(3), 731-347.

Nelson, P. (1970). Information and consumer behavior, Journal of Political Economy, 78(2), 311-329.

Netzer, O. / Srinivasan, V. (2011). Adaptive self-explication of multi-attribute preferences, Journal of Marketing Research, 48(1), 140-156.

O'Reilly, C.A. (1980). Individuals and information overload in organizations: Is more necessarily better?, Academy of Management Journal, 23(4), 684-696.

Olson, J.C. / Jacoby, J. (1972). Cue utilization in the quality perception process, Third Annual Conference of the Association for Consumer Research, 167-179.

Ostrovsky, M. / Schwarz, M. (2009). Reserve prices in Internet advertising auctions: A field experiment, Working Paper, http://faculty-gsb.stanford.edu/ostrovsky/papers/rp.pdf.

Papapetrou, O. / Siberski, W. / Fuhr, N. (2011). Decentralized probabilistic text clustering, IEEE Transactions on Knowledge and Data Engineering, 24(10), 1848-1861.

Park, Y.-H. / Ding, M. / Rao, V.R. (2008). Eliciting preference for complex products: A web-based upgrading method, Journal of Marketing Research, 45(5), 562-574.

Park, D.H. / Kim, H.K. / Choi, I.Y. / Kim, J.K. (2012). A literature review and classification of recommender systems research, Expert Systems with Applications, 39(11), 10059-10072.

Payne, J.W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis, Organizational Behavior and Human Performance, 16(2), 366-387.

Payne, J.W. (1982). Contingent decision behavior, Psychological Bulletin, 92(2), 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(3), 534-522.

Payne, J.W. / Bettman, J.R. / Johnson, E.J. (1993). The adaptive decision maker, Cambridge University Press.

Pfeiffer, J. / Scholz, M. (2013). A low-effort recommendation system with high accuracy: A new approach with ranked pareto fronts, Business & Information Systems Engineering, 5(6), 397-408.

Pilon, D.J. / Friedman, A. (1998). Grouping and detecting vertices in 2-D, 3-D, and quasi-3-D objects, Canadian Journal of Experimental Psychology, 52(3), 114-127.

Pöyhönen, M. / Hämäläinen, R.P. (2001). On the convergence of multiattribute weighting methods, European Journal of Operational Research, 129(3), 569-585.

Reicheld, F.F. (2003). The one number you need to grow, Havard Business Review, 81(12), 46-54.

Rusmevichientong, P. / Shen, Z.-J. / Shmoys, D. (2010). Dynamic assortment optimization with a multinomial logit choice model and capacity constraint, Operations Research, 58(6), 1666-1680.

Sadinle, M. / Fienberg, S.E. (2013). A generalized Fellegi-Sunter framework for multiple record linkage with application to homicide record systems, Journal of the American Statistical Association, 108(502), 385-397.

Sarkar, A. / Chakraborty, A. / Chaudhuri, A. (2008). A method of finding predictor genes for a particular disease using a clustering algorithm, Communications in Statistics – Simulation and Computation, 37(1), 203-211.

Schickel-Zuber, V. / Faltings, B. (2005). Heterogeneous attribute utility model: A new approach for modeling user profiles for recommendation systems, WebKDD Workshop, No. LIA-CONF-2006-003, 80-91.

Schlereth, C. / Eckert, C. / Skiera, B. (2012). Using discrete choice experiments to estimate willingness-to-pay intervals, Marketing Letters, 23(3), 761-776.

Schlereth, C. / Skiera, B. (2012). Measurement of consumer preferences for bucket pricing plans with different service attributes, International Journal of Research in Marketing, 29(2), 167-180.

Schmitt, C. / Dengler, D. / Bauer, M. (2002). The maut-machine: An adaptive recommender system, ABIS Workshop.

Scholz, M. / Dorner, V. (2012). Estimating optimal recommendation set sizes for individual consumers, 33rd International Conference on Information Systems.

Scholz, M. / Dorner, V. / Franz, M. / Hinz, O. (2015). Measuring consumers' willingness-to-pay with utility-based recommendation systems, Decision Support Systems, 72, 60-71.

Scholz, S. / 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.

Schultz, D.E. / Block, M.P. (2015). US online shopping: Facts, fiction, hopes and dreams, Journal of Retailing and Consumer Services, 23, 99-106.

Sheluga, D.A. / Jaccard, J. / Jacoby, J. (1979). Preference, search, and choice: An integrative approach, Journal of Consumer Research, 6(2), 166-176.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection, Systematic Biology, 51(3), 492-508.

Shimodaira, H. (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling, The Annals of Statistics, 32(6), 2616-2641.

Shiu, J.-L. / Sun, C.-H.D. (2014). The determinants of price in online auctions: More evidence from unbalanced panel data, Journal of Applied Statistics, 41(2), 382-392.

Simon, H.A. (1955). A behavioral model of rational choice, The Quarterly Journal of Economics, 69(1), 99-118.

Srinivasan, K. / Wang, X. (2010). Bidders' experience and learning in online auctions: Issues and implications, Marketing Science, 29(6), 988-993.

Stern, B. / Stafford, M.R. (2006). Individual and social determinants of winning bids in online auctions, Journal of Consumer Behavior, 5(1), 43-55.

Stolze, M. / Stroebel, M. (2003). Dealing with learning in ecommerce product navigation and decision support: The teaching salesman problem, Second Interdisciplinary World Congress on Mass Customization and Personalization.

Tan, C.-H. / Teo, H.-H. / Xu, H. (2010). Online auction: The effects of transaction probability and listing price on a seller's decision-making behavior, Electronic Markets, 20(1), 67-79.

Tegarden, D.P. (1999). Business information visualization, Communications of the AIS, 1(1), 4.

Tejada, S. / Knoblock, C.A. / Minton, S. (2001). Learning object identification rules for information integration, Information Systems, 26(8), 607-633.

Tewari, G. / Youll, J. / Maes, P. (2003). Personalized location-based brokering using an agent-based intermediary architecture, Decision Support Systems, 34(2), 127-137.

Theetranont, C. / Haddawy, P. / Krairit, D. (2007). Integrating visualization and multiattribute utility theory for online product selection, International Journal of Information Technology & Decision Making, 6(4), 723-750.

Todd, P. / Benbasat, I. (1994). The influence of decision aids on choice strategies: An experimental analysis of the role of cognitive effort. Organizational Behavior and Human Decision Processes, 60(1), 36-74.

Tractinsky, N. / Meyer, J. (1999). Chartjunk or goldgraph? Effects of presentation objectives and content desirability on information presentation, MIS Quarterly, 23(3), 397-420.

Turban, E. / Sharda, R. / Delen, D. (2010). Decision support and business intelligence systems, Pearson Education India.

Turetken, O. / Sharda, R. (2001). Visualization support for managing information overload in the web environment, 22nd International Conference on Information Systems.

Van der Land, S. / Schouten, A.P. / Feldberg, F. / van den Hooff, V. / Huysman, M. (2013). Lost in space? Cognitive fit and cognitive load in 3D virtual environments, Computers in Human Behavior, 29(3), 1054-1064.

Van Ittersum, K. / Pennings, J. (2012). Attribute-value functions as global interpretations of attribute importance, Organizational Behavior and Human Decision Processes, 119(1), 89-102.

Venkatesh, V. / Davis, F.D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies, Management Science, 46(2), 186-204.

Voelckner, F. (2006). An empirical comparison of methods for measuring consumers' willingness to pay, Marketing Letters, 17(2), 137-149.

Wallenius, J. / Dyer, J.S. / Fishburn, P.C. / Steuer, R.E. / Zionts, S. / Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead, Management Science, 54(7), 1336-1349.

Wang, W. / Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs, Journal of Management Information Systems, 23(4), 217-246.

Wang, T. / Venkatesh, R. / Chatterjee, R. (2007). Reservation price as a range: An incentive-compatible measurement approach, Journal of Marketing Research, 44(2), 200-213.

Wertenbroch, K. / Skiera, B. (2002). Measuring consumers’ willingness to pay at the point of purchase, Journal of Marketing Research, 39(2), 228-241.

Wu, J. / Li, L. / Xu, L.D. (2014). A randomized pricing decision support system in electronic commerce, Decision Support Systems, 58, 43-52.

Xiao, B. / Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact, MIS Quarterly, 31(1), 137-209.

Yu, M.-C. / Goh, M. / Lin, H.-C. (2012). Fuzzy multi-objective vendor selection under lean procurement, European Journal of Operational Research, 219(2), 305-311.

Zachary, W. (1986). A cognitively based functional taxonomy of decision support techniques, Human-Computer Interaction, 2(1), 25-63.

Zhang, L. / Marron, J. / Shen, H. / Zhu, Z. (2007). Singular value decomposition and its visualization, Journal of Computational & Graphical Statistics, 16(4), 833-854.

Zhang, P. / Whinston, A.B. (1995). Business information visualization for decision-making support: A research strategy, First Americas Conference on Information Systems.

Zhao, H. / Ram, S. (2008). Entity matching across heterogeneous data sources: An approach based on constrained cascade generalization, Data & Knowledge Engineering, 66(3), 368-381.

Zheng, R. / Li, J. / Chen, H. / Huang, Z. (2006). A framework for authorship identification of online messages: Writing-style features and classification techniques, Journal of the American Society for Information Science and Technology, 57(3), 378-393.

Zhou, W. / Hinz, O. (2015). Determining profit-optimizing return policies – A two-step approach on data from taobao.com, Electronic Markets, forthcoming.

Zhu, B. / Chen, H. (2005). Using 3D interfaces to facilitate the spatial knowledge retrieval: A geo-referenced knowledge repository, Decision Support Systems, 40(2), 167-182.

## Declaration of Honor

I declare upon my word of honor that the dissertation submitted herewith is my own work. All sources and aids used have been listed. All references or quotations in any form and their use have been clearly identified. The dissertation has not been submitted for examination purposes to any institution before.

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Sämtliche aus fremden Quellen direkt und indirekt übernommene Gedanken sind als solche kenntlich gemacht. Die Dissertation wurde bisher keiner anderen Prüfungsbehörde vorgelegt und noch nicht veröffentlicht.

Markus Franz

Darmstadt, 27.06.2016