A configuration-based recommender system for supporting e-commerce decisions

Michael Scholz[a,*], Verena Dorner[b], Guido Schryen[c], Alexander Benlian[d]

[a]*Assistant Professorship for Information Systems with a focus on Electronic Commerce, Faculty of Business Administration and Economics, University of Passau, Germany*
[b]*Chair for Information and Market Design, Institute of Information Systems and Marketing, Karlsruhe Institute for Technology, Germany*
[c]*Professorship for Management Information Systems, Department of Management Information Systems, University of Regensburg, Germany*
[d]*Chair of Information Systems and Electronic Services, Technical University Darmstadt, Germany*

Abstract

Multi-attribute value theory (MAVT)-based recommender systems have been proposed for dealing with issues of existing recommender systems, such as the cold-start problem and changing preferences. However, as we argue in this paper, existing MAVT-based methods for measuring attribute importance weights do not fit the shopping tasks for which recommender systems are typically used. These methods assume well-trained decision makers who are willing to invest time and cognitive effort, and who are familiar with the attributes describing the available alternatives and the ranges of these attribute levels. Yet, recommender systems are most often used by consumers who are usually not familiar with the available attributes and ranges and who wish to save time and effort. Against this background, we develop a new method, based on a product configuration process, which is tailored to the characteristics of these particular decision makers. We empirically compare our method to SWING, ranking-based conjoint analysis and TRADEOFF in a between-subjects laboratory experiment with 153 participants. Results indicate that our proposed method performs better than TRADEOFF and CONJOINT and at least as well as SWING in terms of recommendation accuracy, better than SWING and TRADEOFF and at least as well as CONJOINT in terms of cognitive load, and that participants were faster with our method than with any other method. We conclude that our method is a promising option to help support consumers' decision processes in e-commerce shopping tasks.

Keywords: E-Commerce, Recommender System, Attribute Weights, Configuration System,

1. Introduction

By providing consumers with access to great amounts of product information, e-commerce has been driving research on consumer decision support systems. Recommender systems in particular have proven valuable in helping consumers make faster and better choices among large numbers of decision alternatives by suggesting alternatives that ought to be considerable for a particular consumer (Dellaert and Häubl, 2012). The intuition behind the two most common approaches, collaborative filtering and content-based recommender systems (e.g., Yue et al., 2014; Adomavicius and Tuzhilin, 2005), is using past information about consumers' purchase decisions to predict future decisions. This can lead to low-quality recommendations when relevant data are missing (i.e., for new customers and new products), and when consumer preferences change over time (Scholz et al., 2015; Ansari et al., 2000). These problems are proposed to be solved by a third approach, multi-attribute value theory (MAVT)-based recommender systems (Pu et al., 2011). Its core idea is predicting decisions based on consumer-specific value functions and attribute importance weights estimated at the time of purchase.

Several methods for measuring attribute importance weights have been developed in recent decades (von Winterfeldt and Edwards, 1986; Edwards and Barron, 1994; Mustajoki et al., 2005). How well they are suited for application in MAVT-based recommender systems in e-commerce contexts largely depends on how many of the following characteristics they exhibit. For one, they need to present consumers with information about the valid ranges of attribute levels. In order to estimate value functions and attribute importance weights, MAVT-based recommender systems require consumer input. But consumers often have little knowledge about the alternatives available in a given purchase decision and are typically not aware of all the levels available for a particular attribute (Xu and Wyer, 2010; Bettman et al., 1993), which adversely affects the reliability of

*Corresponding author. Innstr. 43, 94032 Passau, Germany, Tel: +49 851 509 2416

Email addresses: michael.scholz@uni-passau.de (Michael Scholz), verena.dorner@kit.edu (Verena Dorner), guido.schryen@wiwi.uni-regensburg.de (Guido Schryen), benlian@ise.tu-darmstadt.de (Alexander Benlian)

weight specifications. Hence the first characteristic, i.e. information about valid attribute level ranges is necessary. The second characteristic is closely related to the first one: consumer input ought to be elicited based on evaluation of real alternatives. Evaluations of real alternatives have been found to be more reliable and accurate predictors for consumers' attribute weights than evaluations of hypothetical alternatives (Ding et al., 2005; Ding, 2007)[1]. Consumers are usually not well-trained decision analysts and might form wrong expectations faced with hypothetical options, i.e. they expect them to be available in the market, which may also adversely affect the reliability of weight specifications. In addition, evaluating real alternatives matches consumers' expectations of a typical purchase decision process better (Hauser, 2014), and disconfirming these expectations is likely to give rise to negative perceptions of the recommender system. The third and final characteristic of an ideal MAVT-based recommender system is that consumers need to spend as little time and cognitive effort as possible in giving their input. Positive perceptions of the recommender systems by consumers influence their willingness to use the system and are thus important from a managerial perspective. None of the existing methods for measuring attribute importance weights exhibit all three characteristics of an ideal MAVT-based recommender system for e-commerce contexts; existing methods are based on the evaluation of hypothetical alternatives and/or become too cognitively demanding quickly.

We address the three characteristics by using an attribute-based product configurator (Valenzuela et al., 2009) as the recommender system's interface and developing an attribute weight measurement method that takes into account the actions of decision makers throughout the configuration of a desired product. This type of product configurator is used on many companies' websites, such as Audi, Citroen, Dell, Ducati, eShakti, Ford, Lenovo, Mercedes Benz, Modern Tailor, MyMuesli, Shoes of Prey, Stevens Bikes, or Volvo. As an illustrative example, let us consider a decision maker who uses our recommender system to search for a new digital camera. Our configuration-based system displays all available attributes and ranges, but lets the decision maker select only those levels of an attribute that are available given the selected levels of other attributes.

[1] Although closely related to the first characteristic, they are not identical. Estimation methods such as conjoint analyses may inform decision makers about the attribute ranges of real alternatives, but force decision makers to judge hypothetical alternatives that are systematically designed based on the attribute ranges of real alternatives.

Price is then calculated according to the selected attribute level combination. The decision maker must weigh the configurable attributes and the price in order to make a reliable decision about which levels to select. For instance, if the decision maker selects maximum zoom factor of 30×, it might not be possible for her to also choose a camera of very small size. Hence, if the decision maker wants to purchase a small camera with a high optical zoom, she must thoroughly deliberate whether to accord a higher attribute weight to optical zoom or to camera size. If the size of the camera is more important for her, she may choose a lower level for optical zoom in order to be able to select a smaller size. Our proposed system uses the sequence in which attribute level selections are made by a decision maker for estimating attribute weights. The attribute weight estimation procedure we propose in this paper is based on two principles. First, the better the selected level of a particular attribute, the higher is this attribute's weight. For example, changing the optical zoom selection from 30× to 18× is interpreted as decreasing the attribute weight for optical zoom. Second, the fewer changes to the selection of a high attribute level throughout the recommendation process and the later in the configuration process they are made, the higher is the weight for that attribute compared to an attribute with a commensurate level. For instance, zoom will be accorded a higher weight than photo resolution if the decision maker initially selects the best levels for both attributes and then downgrades photo resolution twice (in order to be able to choose a better level for another attribute) while keeping the selection for optical zoom unchanged.

Our proposed configuration-based recommender system meets all three characteristics of an ideal MAVT-based recommender system in e-commerce contexts as specified above. First, it does not let decision makers select levels that are not available due to the selection of other attribute's levels. This meets the first characteristic – presenting information about available attribute ranges in order to obtain reliable estimates of attribute weights. Second, our system does not require that decision makers evaluate hypothetical products, which meets the second characteristic. Third, our system also incorporates behavioral principles from prospect theory (Tversky and Kahneman, 1992), specifically the value function, and is able to factor consumer behavior during configuration into attribute weight measurement. We believe that a configuration process is less time-consuming and cognitively less demanding than the evaluation of hypothetical alternatives – especially for decision makers who are not trained in using attribute weight elicitation methods. This addresses

the third characteristic.

We assess the performance of our configuration-based recommender system in a laboratory experiment in which we compare the performance of our method to three established attribute weight measurement methods. We chose three methods which, just like our proposed method, take into account that attribute weight formation is dependent on the range of available attribute levels (Van Ittersum et al., 2007): SWING, TRADEOFF, and ranking-based conjoint analysis. Finally, we discuss and empirically evaluate possible adaptations of our proposed recommender system, specifically accounting for reference point and anchoring effects (Tversky and Kahneman, 1974).

We contribute to recent research on MAVT-based recommender systems and the ongoing research on attribute weight elicitation methods by providing a novel attribute weight elicitation method that is i) tailored to support e-commerce purchase decisions, ii) provides information about the available attribute level ranges in an easily comprehensible and natural manner, and iii) can be easily integrated into retailing websites, many of which already use product configuration systems. From a managerial point of view, our approach helps improving consumer decision support and sales processes.

The paper is organized as follows. We briefly introduce multi-attribute value theory as the theoretical foundation of MAVT-based recommender systems in Section 2. Section 3 presents existing research on MAVT-based recommender systems. We introduce our novel attribute weighting method in Section 4. Section 5 presents an empirical comparison of our proposed method to other methods that aim at measuring attribute importance weights. Section 6 concludes the paper with a discussion of possible adaptations of our method (see supplementary material for details), practical and research implications as well as suggestions for future research.

2. Multi-attribute value theory

Multiple criteria decision analysis (MCDA) is employed in many disciplines, including management science, operations research, psychology, and marketing. One of the most frequently applied theories to MCDA problems is multi-attribute value theory (MAVT) (Fishburn, 1967; Keeney and Raiffa, 1976; Wallenius et al., 2008). MAVT is based on the assumption that, in a

decision situation, a real value function V exists which represents the preferences of the decision maker such that the more preferable an alternative is, the larger its numerical value. This function computes the value of each decision alternative by aggregating its performance in all attributes i (e.g., price, color, the model of a car). Its general form is represented by the equation

$$V = f(v_1(x_1), \ldots, v_n(x_n), w_1, \ldots w_n) \tag{1}$$

f is the multiple-attribute value function. x_i represents a particular level of attribute i. v_i is a single-attribute value function that assigns a real value to x_i and reflects the (subjective) preference of a particular decision maker. w_i is the weight for the single-attribute value. Single-attribute value functions are usually normalized, with the value of the worst level x_i° of attribute i set to 0 and the value of the best level x_i^* set to 1. The normalized single-attribute value functions v_i are then multiplied with attribute weights w_i (Fischer, 1995; Pöyhönen and Hämäläinen, 2001). The simplest form of Equation 1 is additive:

$$V = \sum_{i=1}^{I} w_i v_i(x_i) \tag{2}$$

Previous research has shown that additive models are robust as long as attribute weights are specified reliably (Butler et al., 1997; Dawes, 1979).

Attribute weights w_i are scaled from 0 to 1. The following constraints hold for two attributes i and i' if the value functions return values in $[0, 1]$ and the attribute weights are normed in $[0, 1]$ (Keeney and Raiffa, 1976):

$$\begin{aligned} 0 &= w_i v_i(x_i^\circ) + w_{i'} v_i(x_{i'}^\circ) \\ 0 \leq w_i &= w_i v_i(x_i^*) + w_{i'} v_i(x_{i'}^\circ) \leq 1 \\ 0 \leq w_{i'} &= w_i v_i(x_i^\circ) + w_{i'} v_i(x_{i'}^*) \leq 1 \\ \sum_{i=1}^{I} w_i &= 1 \end{aligned} \tag{3}$$

The largest weight is given to the attribute which contributes most to the overall value V. In other words, the weight of i represents the impact of attribute i on value V when the level of

attribute i is changed from x_i° to x_i^*. The weight of attribute i relative to the weight of attribute i' represents the impact of i on V compared to the impact of i' on V, assuming that the levels of all attributes are commensurable ($v_i(x_i) = v_{i'}(x_{i'})\ \forall i \neq i'$).

Consider, for example, a decision between alternatives described with two attributes A and B whose attribute values are normalized in $[0, 1]$, i.e., each attribute's best level has a value of 1. If the best level of A improves the overall value of an alternative more than the best level of B, A has a larger attribute weight than B. Hence, a decision maker who wants to decide between alternatives $X_1 = \{x_A^\circ, x_B^*\}$ and $X_2 = \{x_A^*, x_B^\circ\}$ would choose alternative X_2.

There are numerous approaches for eliciting attribute weights, some of which have been used in MAVT-based recommender systems. The following section briefly discusses relevant attribute weighting methods with respect to the three characteristics (presenting information about available attribute ranges, consumer input should be based on the evaluation of really existing alternatives, and demanding as little time and cognitive effort as possible) we put forward in Section 1.

3. Measuring attribute weights in MAVT-based recommender systems

Most recommender systems implement content-based or collaborative filtering techniques (Yue et al., 2014; Adomavicius and Tuzhilin, 2005). Content-based techniques recommend products similar to those a consumer has rated highly in the past. Collaborative filtering techniques recommend products to a consumer based on product ratings by other consumers who have similar tastes and preferences. Both techniques frequently produce low-quality recommendations, which is due to two major issues[2] (Ansari et al., 2000). First and most important is the cold-start problem (Kim et al., 2011). Traditional content-based and collaborative filtering techniques cannot provide recommendations unless multiple product ratings from a number of consumers, or at least from the consumer currently using the system, are available. Neither can they provide recommendations for new or seldom rated products. Although many approaches have been proposed in recent research to cope with the cold-start problem, there is no solution that can predict the value of new

[2] Another potential source of inaccuracy is, of course, consumers purchasing products as gifts or on behalf of other consumers (Ansari et al., 2000).

products for an existing consumer and the value of existing products for a new consumer without using additional data, such as explicit ratings (Zigoris and Zhang, 2006; Kim et al., 2011), product taxonomies (Weng et al., 2008), customer reviews (Levi et al., 2012), or social media data (Forsati et al., 2014; Yu et al., 2014; Zhao et al., 2016). Since additional data are not available in all contexts and for all consumers and products, the cold-start problem is still a challenge for content-based and collaborative filtering techniques. Second, prior product ratings are historical data which reveal past but not necessarily current preferences (Pfeiffer and Scholz, 2013). A change of preferences of those consumers that demand a recommendation from a collaborative filtering or a content-based system likely reduces recommendation accuracy[3]. These two issues prevent improvements in the recommendation quality of content-based and collaborative-filtering recommender systems even if these systems use efficient methods, such as matrix factorization (Forsati et al., 2014). Neither of these issues arises in MAVT-based recommender systems.

MAVT-based recommender systems estimate consumer-specific values for all products of a given category at the time of purchase (Huang, 2011; Pu et al., 2011; Scholz et al., 2015), based on individual value functions and attribute weights. The first characteristic of ideal MAVT-based recommender systems, is therefore reliable estimation of attribute weights. This is not an easy characteristic to implement in e-commerce contexts due to the nature of the typical decision maker in e-commerce: they often have little knowledge about the alternatives available in a purchase decision and are typically not aware of the levels available for a particular attribute (Xu and Wyer, 2010; Bettman et al., 1993). This adversely affects the reliability of weight specifications, and requires careful choice of an attribute weight elicitation method. Specifically, the first characteristic implies that decision makers need to be presented with information about differences between attribute levels and available attribute level ranges.

Existing attribute weight elicitation methods have been found to actually measure three differ-

[3]If consumer A's preferences change over time, a collaborative filtering system will identify consumers that are similar to A based on A's historical and maybe obsolete preferences. The preferences of these other consumers were similar in the past to consumer A's past preferences. But these past preferences are not good predictors of consumer A's actual preferences and purchases.

ent dimensions of attribute importance (Van Ittersum et al., 2007).[4] Since we require a method that takes differences between attribute levels and attribute level ranges into account, methods capturing "determinance" clearly appear the most suitable (Fischer, 1995). Among these methods are conjoint analyses, TRADEOFF (Keeney and Raiffa, 1976; Pöyhönen and Hämäläinen, 2001), SWING (von Winterfeldt and Edwards, 1986) and extensions (e.g., Mustajoki et al., 2005), and simple multi-attribute ranking method such as SMARTS (Edwards and Barron, 1994) and extensions (e.g., Mustajoki et al., 2005) that are based on SWING. For use in MAVT-based recommender systems, especially SWING (Huang, 2011), ranking-based (De Bruyn et al., 2008) and choice-based conjoint analysis (Pfeiffer and Scholz, 2013) have been put forward.

The first characteristic of ideal MAVT-based recommender systems (presenting information about available attribute ranges and attribute level differences) is met by several existing methods, such as TRADEOFF, SWING and conjoint analysis. These methods explicitly present information about the attribute ranges of real alternatives.

The second characteristic (consumer input should be based on the evaluation of really existing alternatives) is not met by existing methods: they generally rely on the evaluation of hypothetical alternatives.[5] This might be not an issue for decision makers who are well-trained in using MAVT-based methods, but consumers are usually unfamiliar with these methods. In addition, consumers are likely to expect – based on their experience with other online shopping situations – to be presented with real alternatives only. Evaluating obviously unrealistic alternatives can lead to false expectations about product availability in the market and thus to negative perceptions of the

[4]The three dimensions are determinance, salience and relevancy. Determinance reflects the importance of an attribute in specific choice situations. It is estimated based on decision makers' valuation of attribute level difference and hence depends strongly on the differences between attribute levels. The larger the difference of the valuation of the worst and the best level of an attribute, the more determinant this attribute becomes (Fischer, 1995). Salience reflects the ease with which a particular attribute comes to a decision maker's mind. Relevancy refers to the importance of attributes for a decision maker regardless of attribute level ranges (Van Ittersum et al., 2007).

[5]The alternatives for evaluation in a conjoint task are systematically generated based on the range of available attribute levels. Alternatives composed of the best and worst available levels in multiple attributes are particularly unlikely to exist in reality. SWING, for example, starts with an alternative in which all attributes are set to their worst levels.

recommender system.

The third characteristic of ideal MAVT-based recommender systems (demanding as little time and cognitive effort as possible) is not met by existing methods. Specifying attribute weights clearly can be very challenging with these methods, considering that in a purchase situation many attributes may be relevant to the consumer, and consumers are usually not well-trained in applying methods such as TRADEOFF, SWING or conjoint analyses. At the same time, prior research shows that a cognitively demanding task can help decision makers to come to more stable preferences (Hoeffler and Ariely, 1999), warning against oversimplification. Considering existing methods, however, we believe that there is scope to find a better way to balance individual time and cognitive effort. Taking TRADEOFF, each trade-off decision by itself is not very demanding – but it requires many such decisions to be made.[6] When attributes have many levels or are continuously scaled, the decision maker may find herself unable to make these trade-off decisions (Eisenführ et al., 2010). SWING, on the other hand, requires only few decisions – but these demand high cognitive capabilities on part of the decision maker. In empirical studies, both TRADEOFF and SWING showed statistically significant range sensitivity (Fischer, 1995). SWING exhibited high convergent validity (Borcherding et al., 1991) but low external validity (Borcherding et al., 1991), while TRADEOFF showed low internal consistency (Borcherding et al., 1991). Finally, conjoint analysis has been applied to a wide range of research problems beyond the scope of its original marketing applications (Wyner, 1992), due to its great flexibility in modeling interactions between attributes (Akaah and Korgaonkar, 1983). However, larger numbers of attributes lead to increasing numbers of attribute combinations to be evaluated in conjoint analyses. In such cases, responses are likely to become unreliable due to respondent fatigue (Wyner, 1992; Pfeiffer and Scholz, 2013).

To summarize, among existing methods for eliciting attribute weights, those that refer to attribute determinance seem most appropriate. None of them, however, exhibit all characteristics of ideal MAVT-based recommender systems in e-commerce contexts since they i) base the evaluation on hypothetical alternatives (characteristic 2 is not met) and/or ii) fast become too cognitively demanding (characteristic 3 is not met). As noted by prior research, lack of knowledge about attribute

[6] Additionally, it requires knowing the attribute value functions.

level ranges leads to unreliable weight specification (characteristic 1) and makes interpreting attribute weights impossible (see Mistake 8 in Keeney, 2002). This characteristic can be satisfied by methods eliciting attribute determinance by having the decision analyst, or in our case the decision support system, provide the relevant information. The importance of designing recommender systems that use evaluations of real instead of hypothetical alternatives and that require as little consumer input as possible has been highlighted by previous research (e.g., Ding et al., 2005; De Bruyn et al., 2008; Pfeiffer and Scholz, 2013). In the following section, we will develop a novel method for eliciting attribute weights that is specifically geared toward e-commerce shopping tasks and that exhibits all three characteristics. We base our method on a product configuration process.

4. Adapting MAVT-based recommender systems to cognitive processes

4.1. Estimating attribute weights from product configuration processes

Recalling the definition of attribute weights (see Section 2), the attribute that contributes most to an alternative's overall value is the attribute with the largest weight – assuming that all attributes are commensurable. Decision makers using a product configuration system are more likely to select a better level for an attribute with a large weight i than for an attribute with a small weight i': improvements in attribute i have a higher impact on the overall value than commensurate improvements in attribute i'. Let us introduce the following notation:

In a configuration system, decision makers assemble their desired product in T discrete configuration steps. In each step, a decision makers changes the level of exactly one attribute. Each attribute has τ_i available levels. The worst possible selection is defined as $s_i^\circ = 0$ and the best possible selection as $s_i^* = \tau_i - 1$. In each configuration step t, the decision maker changes the selected level of exactly one attribute i to $s_{i,t} \in [0, \tau_i - 1]$. We compute attribute weights $w_i^{(raw)}$ as the sum over all normalized attribute level selections $s_{i,t}$.

$$w_i^{(raw)} = \sum_{t=0}^{T} \frac{s_{i,t}}{\tau_i - 1} \tag{4}$$

τ_i is the number of totally available attribute levels. We divide the selected level by the number of available levels minus 1 in order to diminish a source of potential bias from attribute weight

Table 1: List of Variables

Symbol	Description
i	attribute index
t	configuration step index
x_i	level of attribute i
$w_i^{(raw)}$	unnormalized weight for attribute i
w_i	normalized weight for attribute i
$v_i(x_i)$	value function for attribute i
$s_{i,t}$	number of the selected level for attribute i at configuration step t
τ_i	number of totally available levels for attribute i
e_i	reference point for attribute i
α, β, λ	scaling constants

estimation: higher numbers of attribute levels lead to higher stated attribute weights (Weber and Borcherding, 1993).

The behavioral process underlying our proposed model for attribute level aggregation has two general implications for attribute weights. First, every attribute whose level remains set to the worst level throughout the configuration process has a weight of 0: it is irrelevant for the decision maker. Second, an attribute is considered the more important for a decision maker the more often she selects its best level, or the longer it remains set to the best level, during configuration.

To satisfy the constraints in Equation 3, attribute weights $w_i^{(raw)}$ are normalized in $[0, 1]$.

$$w_i = \frac{w_i^{(raw)}}{\sum_{i=1}^{I} w_i^{(raw)}} \tag{5}$$

We discuss necessary adaptations of the configuration process in the next subsection and thereafter present a numerical example for our proposed attribute weight estimation.

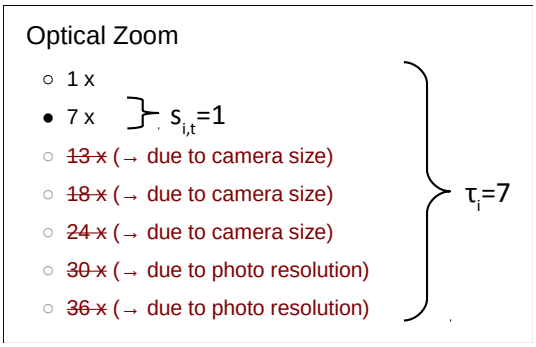

Figure 1: Exemplary configuration of an attribute

4.2. Adapting the product configuration process for attribute weight estimation

In product configuration systems, available combinations of attribute levels are determined by the set of available products, i.e., the market situation. We need two conceptual adaptations – estimation of prices and determination of available attribute level combinations – when applying product configuration systems to measuring attribute weights.

Prices. Products are characterized by a set of attributes including price (see Section 2). Price levels are not selected directly by the decision maker in order to guarantee that attribute level selections are made in trade-off to another attribute. The weight of attribute "price" is measured based on the expected attribute level $E(x_{i,t})$ instead of the selected level $s_{i,t}$. The expected price level $E(x_{i,t})$ is the mean price of all products that meet the configuration in round t (i.e., having attribute levels that are equal to or better than the attribute levels selected in the configuration system). Consider, for example, a market where cameras cost between 100 and 400 Euros, the mean price for a specific configuration is 200 Euros and the cameras that meet this configuration cost between 150 and 250 Euros. Normalizing in $[0, 1]$ gives an expected price of $E(x_i) = 0.667$ and a price interval of $[0.833, 0.5]$ for those cameras that meet the given configuration. The unnormalized weight for price equals the average expected price over all configuration rounds T and is $w_{i,t}^{(raw)} = 0.667$ in our example.

Available attribute level combinations. For some attributes, there might be almost as many levels as there are available products, for instance the weight of a notebook or the mileage of a car. Letting decision makers choose among hundreds of levels, however, is infeasible due to constraints

on cognitive capacity (von Nitzsch and Weber, 1993). Forcing decision makers to choose among very finely-grained attribute levels likely leads to greater attribute weight instability since decision makers usually have a range of attribute levels they consider acceptable (Wang et al., 2007). We therefore reduce the number of continuous attribute levels by aggregating them to intervals. For instance, an interval level for the attribute "weight" might be "below 200g".

Product configuration systems operate on a database of products that represents the entirety of available attribute level combinations. Once a decision maker has selected a particular level s_i of attribute i, the system i) selects those products P whose level of i equals s_i and ii) identifies those levels of all other attributes $i' \neq i$ that exist in P. Our system then assesses the best and worst available attribute level ($x^*_{i',P}$ and $x^\circ_{i',P}$) in P for each attribute i'. Levels of i' outside the interval $[x^\circ_{i',P}, x^*_{i',P}]$ are marked as unavailable for s_i and disabled if the system is configured to present the relationships between the attributes (Figure 1).

Summary of configuration process. The final configuration process is depicted in Figure 2. Decision makers start with a default configuration in which the worst level x°_i is set for each attribute i. In an iterative process, decision makers adapt the configuration such that it fits their attribute weights. Simultaneously, they learn about the attribute relations: if attribute levels become unavailable for certain configurations, they are visually and functionally disabled. This allows decision makers to become aware of the attribute relations and to adjust their attribute weights gradually, changing their attribute level selections until they finally arrive at stable attribute weights.

4.3. Numerical example for estimating attribute weights with a product configuration system

Let us assume a configuration system for cameras based on the attributes "photo resolution" and "zoom". The price for each configured product is computed based on cameras that are available in the market and that have attribute levels equal to or better than the selected attribute levels.

Consumers can select one out of four resolution levels: 5, 10, 15, or 20 megapixel and one out of four zoom levels: 3, 6, 9, or 12×. Hence, $\tau_i = 4$ for both photo resolution and optical zoom.

Let us also assume that cameras between 50 and 250 Euros are available. A consumer starts the configuration process in t_0, with resolution and zoom set to the worst levels (5 megapixel and 3× zoom) and may proceed with configuration steps as shown in Table 2. After each configuration

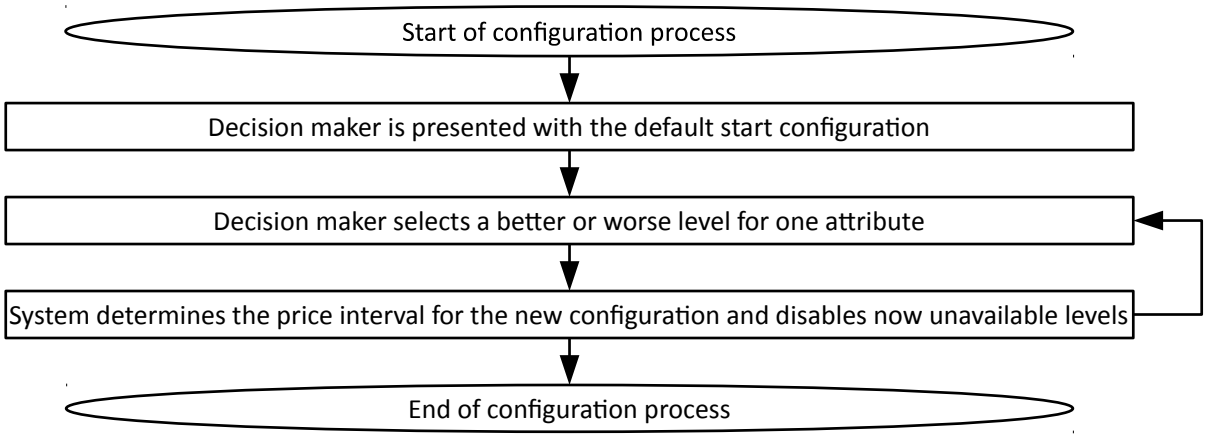

Figure 2: Summary of configuration process

step, the corresponding price range and the mean price are computed by the configuration system. Price ranges and mean prices are also shown in Table 2.

Table 2: Exemplary configuration steps t_i for resolution and zoom and the corresponding changes in price range and mean price

	t_0	t_1	t_2	t_3
resolution	5	15	10	10
zoom	3	3	3	9
price	50–250	80–250	70–250	100–200
mean price	120	135	130	140

Each configuration step is translated into normalized selected attribute levels $s_{i,t}/(\tau_i - 1)$ (Table 3). Unnormalized weights $w_i^{(raw)}$ are the sum over all normalized selected attribute levels for each attribute; final weights w_i are then normalized over all attributes' unnormalized weights.

For our example, the normalized weights w_i in the last column in Table 3 indicate that price is the most important attribute ($w_{price} = 0.543$), followed by resolution ($w_{resolution} = 0.305$) and zoom ($w_{zoom} = 0.152$).

Table 3: Computation of raw and normalized weights for the exemplary configuration in Table 2

	t_0	t_1	t_2	t_3	$w_i^{(raw)}$	w_i
$s_{resolution,t}$	0	2	1	1	–	–
$s_{resolution,t}/(\tau_{resolution} - 1)$	0.000	0.667	0.333	0.333	1.333	0.305
$s_{zoom,t}$	0	0	0	2	–	–
$s_{zoom,t}/(\tau_{zoom} - 1)$	0.000	0.000	0.000	0.667	0.667	0.152
$s_{price,t}$	0.650	0.575	0.600	0.550	2.375	0.543

4.4. Computing single-attribute values

For computing single-attribute product values, we use the S-shaped value function v_i as proposed by prospect theory (Tversky and Kahneman, 1992). The value function $v_i(x_i, e_i)$ implements three behavioral principles that determine its shape (loss aversion, reference point dependence, and diminishing sensitivity) and have been supported by many empirical investigations (e.g., Tversky and Kahneman, 1991, 1992; Wu and Markle, 2008). The function's gain and loss parts are separated by the reference point e_i which represents the inflection point. The shape of the value function $v_i(x_i, e_i)$ is given as

$$v_i(x_i, e_i) = G(x_i, e_i)^{\alpha} + \left[-\lambda\left(-L(x_i, e_i)\right)^{\beta}\right] \tag{6}$$

where α, β and λ are scaling constants. α represents the decision maker's risk aversion in the gain part of the value function; β the risk aversion in the loss part. λ expresses the degree of loss aversion. Values of $\lambda > 1$ indicate higher sensitivity towards losses than gains. In a series of experiments, Tversky and Kahneman (1992) found α and β to be 0.88 and λ to be 2.25 on average.

The gain function $G(x_i, e_i)$ computes the gain for a given attribute level x_i and the loss function $L(x_i, e_i)$ respectively computes the loss for a given attribute level x_i (Fan et al., 2013):

$$G(x_i, e_i) \quad = \quad \max\left(\frac{x_i - e_i}{\max(e_i, 1 - e_i)}, 0\right) \tag{7}$$

$$L(x_i, e_i) \quad = \quad \min\left(\frac{x_i - e_i}{\max(e_i, 1 - e_i)}, 0\right)$$

Both the gain and the loss function only depend on the reference point e_i, i.e. the attribute level which the consumer desires to achieve (Fan et al., 2013). Several methods for eliciting reference points have been proposed in recent research including direct elicitation (Fan et al., 2013) and interactive procedures based on quad trees (Sun and Steuer, 1996).

In the next section, we present the experiment we carried out to evaluate our proposed method.

5. Empirical evaluation

We conducted a laboratory experiment in the KD^{2}Lab at Karlsruhe Institute of Technology[7] in order to compare our proposed method to state-of-the-art methods in terms of their ability to be used in e-commerce recommender systems.

5.1. Treatments

We evaluated the performance of our proposed configuration-based method (CONF) by comparing it with SWING, TRADEOFF, and a ranking-based conjoint analysis (CONJOINT) – three methods that also measure attribute determinance and have been used in recent MAVT-based recommender systems. All treatments implemented a MAVT-based recommender system, each with a different method for attribute weights measurement[8]. Reference point elicitation for each attribute (i.e., the attribute level perceived neither as a loss nor as a gain) and product value computation (see Sections 2 and 4.4) were identical across treatments.

CONF was implemented based on equations 4 and 5. Each attribute (except price) was represented by seven levels. Upon selection of a particular level of an attribute, CONF immediately disabled and crossed out other attributes' levels that became unavailable as a result.

SWING started with informing participants that a camera with the worst levels for all attributes was available and then asked in which order participants would like to improve attributes from their

[7]The KD^{2}Lab is a professionally equipped and managed laboratory with soundproofed computer cubicles that allowed us to control several potential confounding variables such as communication between the participants. Further information about the Lab are available at http://www.kd2lab.kit.edu/english/index.php.

[8]Screenshots of all treatments are presented in the supplementary material.

worst to best levels. The most important attribute was awarded 100 points, and participants were then asked to assign points to the remaining attributes in order of their stated importances. The maximum number of points attributable to an attribute is limited by the number of points given to the previous attribute less one. Attribute weights were finally normalized to add up to 1.

TRADEOFF was implemented as a two-step procedure. The first step was identical to the first step of SWING: ranking attributes according to the order in which the participants would like to improve them. The second step consisted of $I(I-1)/2$ camera comparison tasks for I attributes. In each task, participants were asked to compare two hypothetical cameras (differing in two attributes only) and to adjust the level of the more important attribute such that they perceived the two cameras as equally attractive. Following Pöyhönen and Hämäläinen (2001), we asked the participants to compare all pairs of attributes in order to have some degrees of freedom when estimating attribute weights.

CONJOINT implemented a ranking-based conjoint analysis in which participants were asked to rank twelve hypothetical products. The products were generated such that the attribute level matrix over all products was D-optimal[9]. Attribute weights were computed such that the difference between the ranking vector and the normalized attribute level matrix, multiplied with the attribute weights, was minimal. We used a least squares estimator to compute optimal attribute weights.

All treatments operate on the same product database, i.e., 160 digital cameras, described by photo resolution, optical zoom, camera size, video resolution, photosensitivity and price. Reference point elicitation for each attribute (i.e., the attribute level perceived neither as a loss nor as a gain) and product value computation (see Sections 2 and 4.4) are identical across treatments. Reference points are directly specified by the subjects as those attribute levels that the subjects desire to achieve.

5.2. Sample

We invited 1500 undergraduate and graduate students of a large public university in Germany to take part in a laboratory experiment. 153 participated in and successfully completed the evaluation experiment. We administered the treatments in a between-subjects design. 38 participants

[9] A D-optimal design seeks to maximize the determinant of the attribute level matrix X times X^T.

used the configuration-based recommender (CONF), 40 the recommender with SWING, 42 the recommender with TRADEOFF and 33 the recommender with ranking-based conjoint analysis (CONJOINT). Differences in treatment group sizes are due to random assignment of participants to treatments. Each participant was paid 10 Euros. Participants' average age was 23.03 ($SD = 3.67$) and 25.49% of participants were female. The participants took on average 18.92 minutes ($SD = 5.32$) to complete the experiment.

ANOVA testing indicated no significant differences between participants' average age, proportion of females, and average experience between the four experimental groups (see Table 4).

Table 4: Mean (standard deviation) of participant age, gender and experience and ANOVA results (p-value) for differences between treatment groups

Variable	CONF	SWING	TRADEOFF	CONJOINT	p-value
Age	22.68 (2.47)	23.23 (2.44)	23.81 (5.50)	22.21 (3.03)	0.265
Females	28.95%	25.00%	19.05%	30.30%	0.674
Experience	3.14 (0.97)	3.29 (0.98)	3.13 (0.85)	2.89 (1.00)	0.371
n	38	40	42	33	–

5.3. Procedure

The experiment consisted of three tasks (Figure 3). Participants were given a short introduction to a fictitious purchase situation that required them to purchase a digital camera for personal purposes. They were instructed to use a virtual advisor (MAVT-based recommender system) to search for a new digital camera.

In the first task, participants were asked to denote their attribute reference points and assigned randomly to one of the four treatments (CONF, SWING, TRADEOFF or CONJOINT). We elicited participants' attribute weights using one of the MAVT-based recommender systems (CONF, SWING, TRADEOFF or CONJOINT) implemented for this experiment.

In the second task, participants were first asked to sort, in descending order of their attractiveness, seven cameras drawn at random from the available set of 160 digital cameras. Following this, they were shown a randomly drawn sample of cameras and asked to indicate which camera they preferred most. This holistic rating was then used as benchmark to compare the different treatments' accuracy.

In the third task, participants were asked to fill in a questionnaire on their perceptions of the system (treatment), which we later used to assess the systems' cognitive demands and cognitive fit to the experimental task. More specifically, the questionnaire measured the constructs cognitive load (NASA-TLX scale, Hart and Staveland, 1988), perceived difficulty (Dellaert and Dabholkar, 2009), fit to task (based on Van Der Land et al., 2013), and preference insights (based on Xu et al., 2014), and contained questions about product experience, age and gender[10].

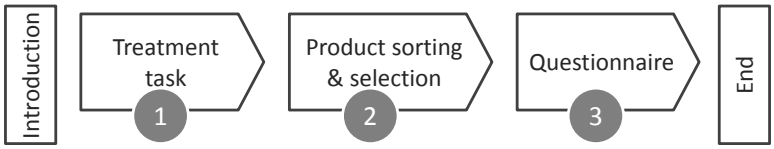

Figure 3: Summary of experimental procedure

5.4. Analysis and results

We estimated the attribute weights for each of the four treatments. Table 5 shows that, on average, i) participants in the CONJOINT treatment had significantly larger weights for video resolution than participants in any other treatment, ii) price was considered rather unimportant by participants across all treatments, and iii) photo resolution was very important according to the weights measured by CONF, SWING and CONJOINT.

The similarity between attribute weights across treatments was compared with Pearson correlation coefficients (Table 6). Results indicate that i) TRADEOFF attribute weights were very different from those elicited with the other methods and ii) CONF attribute weights were rather similar to those of SWING.

[10]The complete questionnaire is presented in the supplementary material.

Table 5: Means (standard deviations) of attribute importance weights (in %)

	Resolution	Zoom	Size	Video	Photosensitivity	Price
CONF	29.9 (8.3)	22.7 (8.2)	18.4 (11.3)	8.3 (9.9)	10.9 (11.0)	9.8 (6.9)
SWING	32.7 (21.8)	12.1 (10.9)	8.5 (11.6)	13.8 (16.3)	24.7 (23.4)	8.2 (12.6)
TRADEOFF	15.9 (10.2)	14.6 (11.5)	20.6 (13.5)	12.9 (8.8)	19.5 (17.0)	16.5 (11.4)
CONJOINT	19.6 (16.9)	13.7 (13.9)	10.4 (10.2)	30.4 (21.0)	15.5 (15.3)	10.5 (12.5)

Table 6: Correlations between the attribute weights estimated with different methods

	SWING	TRADEOFF	CONJOINT
CONF	0.49	0.01	-0.17
SWING		0.01	0.31
TRADEOFF			-0.68

As we argued in Section 3, a suitable method for eliciting attribute weights ought to i) present information about available attribute ranges, ii) base consumer input on the evaluation of really existing alternatives, and iii) demand as little time and cognitive effort as possible. We believe that meeting these characteristics will improve recommendation accuracy and perceptions of the recommender system.

Cognitive demand is measured as *cognitive load* (by using the NASA-TLX questionnaire) and *perceived difficulty*. In addition, we examine how suitable (fit to task) and helpful (preference insight) participants consider the recommender systems. Recommendation accuracy is measured as *sorting accuracy* and *ranking accuracy*, i.e. based on a holistic rating of real products. E-commerce consumers are usually interested in speeding up the process of exploring and evaluating (generally large numbers of) available products. Consumers who are not willing to evaluate all available products need only compare the top-ranked products and select one alternative holistically. Hence, recommender systems are commonly evaluated based on holistic product ratings

(Scholz et al., 2015; Xiao and Benbasat, 2007; Herlocker et al., 2004).

The correlation between the estimated and the actual ranks (Columns 3 and 4 in Table 7) reflects how accurately a method predicts the holistic ranks of all (displayed) products; the rank of the preferred product (Columns 5 and 6 in Table 7) indicates which rank was predicted for the most preferred product[11].

Product ranks were predicted based on product values (in descending order), which in turn were computed as the weighted sum of attribute values (Equation 2), using reference points e_i (as specified by the participants in task 1, Figure 3) to compute single-attribute values (Equation 6).

To compare the different methods' performance with regard to sorting accuracy, we computed Spearman's rank correlation coefficients between participants' rankings and the predicted ranks of the seven randomly chosen cameras for each treatment. We then transformed the coefficients into Fisher z-values (Silver and Dunlap, 1987) and performed a linear regression to determine whether treatments differed significantly in their sorting accuracy. Differences in ranking accuracy were determined with an ordered logit regression; CONF was the baseline in both regressions. Table 7 suggests that CONF performed significantly better than TRADEOFF and CONJOINT. We also compared the time required to complete the treatment tasks with a Gamma regression. Participants in the CONF treatment were significantly (and substantially) faster than in the other treatments (Columns 7 and 8 in Table 7).

That CONF and SWING exhibit similar recommendation accuracy fits the results of the correlation analysis of attribute weights (Table 6) that indicate similar weights. Figure 4 further illustrates these results. The probability that the preferred product of a participant appears in a list of the top-N cameras (x-axis in Figure 4) is clearly higher for CONF and SWING than for TRADEOFF and CONJOINT. Since consumers consider only a few (typically the top-N) alternatives when shopping online and even fewer in the presence of a recommender system (Häubl and Trifts, 2000), this is an important accuracy measure. Even if the list only contains 2 cameras, the probability that one of them is the preferred camera exceeds 70% for CONF and 65% for SWING – as opposed to less than 40% for TRADEOFF and CONJOINT. On average, the probability that

[11] Box plots on the correlation between the estimated and the actual ranks and on the rank of the preferred product are given in the supplementary material.

Table 7: Comparison between the actual and the predicted ranks (as correlation and rank of the preferred product) and time of treatments (means and standard deviations)

	n	Correlation (in %)		Rank of the Preferred Product		Time (in sec)	
		Mean	SD	Mean	SD	Mean	SD
CONF	38	**57.3**	33.6	**2.2**	1.7	**261.2**	109.0
SWING	40	48.5	38.8	2.6	1.9	349.6***	124.7
TRADEOFF	42	11.5***	54.8	3.9***	2.2	519.2***	163.9
CONJOINT	33	25.2*	56.5	3.8**	2.3	382.5***	156.2

$^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$

the preferred camera contained in the list is 7.4% higher when using CONF rather than SWING (TRADEOFF: 28.7%, CONJOINT: 26.7%).

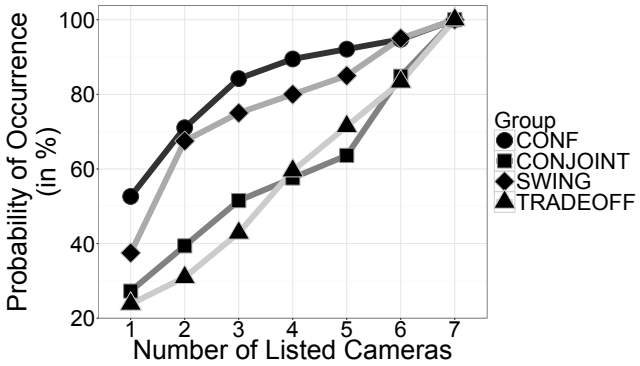

Figure 4: Probability that the preferred product occurs in the top-n recommendations

Participants reported the lowest cognitive load and perceived difficulty in CONF (Columns 2 – 5 in Table 8), but the difference to CONJOINT was not significant. Participant perceptions of fit to task similarly indicate that CONF fitted best, followed by CONJOINT (no significant difference), with significant differences to both SWING and TRADEOFF (Columns 6 and 7 in Table 8). Participants were also asked whether the treatment helped them gain better insight into their product preferences. CONF was, on average, rated better than all other methods, but not

significantly so (Columns 8 and 9 in Table 8).[12].

Table 8: Treatment evaluation by participants (means and standard deviations)

	Cognitive Load		Difficulty		Fit to Task		Preference Insights	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
CONF	**7.68**	2.65	**2.41**	1.14	**4.90**	1.07	**4.61**	1.63
SWING	9.23*	2.73	3.18*	1.36	4.22**	0.79	4.04	1.72
TRADEOFF	9.50**	3.26	3.70***	1.45	4.17***	1.04	3.99	1.77
CONJOINT	8.81	2.43	2.84	1.23	4.68	0.92	4.32	1.69

$^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$

The correlations between the self-reported measures suggest that better cognitive fit of the recommender system to the experimental task reduced cognitive load: perceived fit to task was negatively correlated with perceived cognitive load ($p = 0.010$) and perceived difficulty ($p < 0.001$) across and within treatments. This interpretation is supported by the fact that the time required by participants for the treatment task was not correlated with perceived cognitive load ($p = 0.294$), perceived difficulty ($p = 0.120$), or perceived fit to task ($p = 0.494$).

The more accurately the treatments predicted participants' product rankings, the higher were participants' perceived preference insights ($p = 0.020$). This is interesting, considering that participants received no feedback on recommender system accuracy. Providing information about the available attribute level ranges, as the CONF treatment did, apparently helped consumers obtain insights into their preferences, and ultimately resulted in a higher recommendation accuracy.[13]

[12]Box plots on all scales are presented in the supplementary material. Differences between the treatments in terms of cognitive load, difficulty, fit to task, and preference insights were tested in linear regressions with robust standard errors. All scales have been found to be reliable as indicated by Cronbach's alpha of 0.71 for cognitive load, 0.87 for difficulty, 0.77 for fit to task and 0.93 for preference insights.

[13]Recall that the CONF treatment provided the information about attribute relations by means of the configuration process: once a level of a particular attribute was selected, the system disabled unavailable levels of other attributes.

6. Discussion

We argue that MAVT-based recommender systems using attribute weighting methods, such as SWING or conjoint analysis, are not perfectly suited for application in contexts like e-commerce where decision makers are not willing or able to evaluate hypothetical alternatives and are not willing or able to spend a lot of time and cognitive effort on the task. In such contexts, attribute weights estimated with existing MAVT-based methods can be unreliable.

Our aim was to develop a MAVT-based recommender system with an attribute weight measurement method that supports such decision situations. At the core of our method is a configuration process in which consumers learn about the range of available attribute levels in a natural manner. Attribute weights are computed based on the behavior of the decision maker during the configuration process.

Our empirical investigation supports our reasoning in a number of ways. First, recommendation accuracy was higher with our proposed approach than with CONJOINT or TRADEOFF, indicating more reliable attribute weight estimates. SWING provided similar results as our approach with respect to attribute weights and accuracy. Second, cognitive load and perceived difficulty were lower than with TRADEOFF or SWING, indicating that attribute-level comparisons were easier with our proposed approach. The suggestion that this difference may be due to the more "natural" exploration of the attribute level ranges that are available is supported by the fact that better fit to task was associated with lower cognitive load. While lower cognitive load does not necessarily lead to higher decision quality (Hoeffler and Ariely, 1999), we may state that in our case, cognitive load was not correlated to recommendation quality: the configuration-based system performed as well as SWING but was cognitively less demanding.

From a managerial point of view, our method enables companies who already use a product configuration system on their websites to obtain better insights into their customers' preferences and to use this information for improving market shares and customer segment estimates. Implementing our proposed attribute weight elicitation method is relatively easy and probably would not require major alterations to configurator interfaces. If only market share estimation were the

goal, it would not even be necessary to measure reference points[14] and fit value functions, which would make the interaction with the configuration system even simpler.

Our theoretical contribution is threefold: (1) We provide a set of three characteristics that attribute weight elicitation methods should exhibit when used in e-commerce contexts, especially in MAVT-based recommender systems. (2) We design a method that exhibits all three characteristics and empirically compare this method to existing methods that do not meet all of these characteristics. Our empirical evaluation provides evidence that our proposed method and thus the derived characteristics help to improve MAVT-based recommender systems. (3) We provide a method that translates consumers' interactions with a configuration system into attribute weights. This method relies on both MAVT and findings from research on behavioral decision making.

In summary, our results suggest that measuring consumers' attribute weights with a configuration process that provides information about the relationships among product attributes in a natural way seems to fit consumers' exploratory-evaluative decision context very well, with comparatively low cognitive load, little required time and high recommendation accuracy.

Clearly, this study has limitations that need to be addressed in future research. First, there are a number of possible adaptations to our configuration-based approach that may improve its recommendation accuracy further. We investigated two possible adaptations in a supplementary online experiment (see supplementary material for details). The first adaptation refers to the question whether results can be improved by using a different reference point (Tversky and Kahneman, 1974) for attribute level normalization. In the model presented in this paper (see Equation 5 in Section 4.1), we assume that decision makers evaluate the attractiveness of attribute levels compared to the best overall level in this attribute. In the supplementary experiment, we tested whether using the best level available in a specific configuration as a reference point improves accuracy. Our results show no improvement.

The second adaptation refers to the question whether accuracy can be increased by accounting for possible anchoring effects (Tversky and Kahneman, 1974) during configuration. Anchoring reflects decision makers' "excess reliance on the starting point and insufficient adjustment for sub-

[14]Existing configurators like Lenovo's use up to three pre-configured alternatives (for one chosen notebook) that the customer can use as a reference point in her configuration.

sequently considered information" (Wansink et al., 1998, p. 72). We defined a sequence weighting function in order to test whether anchoring occurs, and if so, whether decision makers rely more heavily on early or late configurations. Our results indicate that for smaller numbers of attribute levels, the best accuracy was obtained for weighting late configuration steps more heavily than early steps. For higher numbers of attribute levels, decision makers simultaneously adjusted their attribute weights and configuration while learning about the availability of different attribute combination, and accuracy could not be improved by weighting the configuration sequence.

Second, our study focused on a product with (at least) ordinally scaled attributes. Since many decision objects have important properties that are nominally scaled, our proposed configuration-based method ought to be extended to such cases. This extension might, however, not be trivial when nominally scaled attributes have many levels.

Third, the results of our experiment indicate that better preference insights might be associated with higher recommendation accuracy, and that learning about the attribute level ranges between attributes in a configuration-based setting improves preference insights. Enriching other attribute weight measurement methods with this information and examining the effects on preference insights and recommendation accuracy thus provides another interesting avenue for future research.

Fourth, we implemented SWING and TRADEOFF as proposed in existing studies. There is, however, space for improvements of both methods. SWING, for example, starts with an alternative with the worst levels for all attributes. Since such an alternative is very unlikely to exist, an improvement might be to start with an alternative with medium levels for all attributes. Lahtinen and Hämäläinen (2016) show that TRADEOFF weights (as a variant of an even swaps decision analysis) might be path dependent and biased. They suggest a procedure to cope with this problem which provides an interesting extension to improve TRADEOFF performance. However, the effort associated with this method is comparable to the original TRADEOFF method and it is still based on hypothetical alternatives, which may make its application in an e-commerce context difficult. We used $I(I-1)/2$ trade-off tasks in order to better able to cope with inconsistent evaluations by decision makers (Pöyhönen and Hämäläinen, 2001). The absolute minimum of tasks required to estimate attribute weights for I attributes is $I-1$. Improving TRADEOFF by i) incorporating the procedure proposed by Lahtinen and Hämäläinen (2016) and ii) reducing the number of trade-off

tasks provides an interesting starting point for further research.

Fifth, we designed and tested our proposed attribute weighting method in a specific e-commerce context. Future research could adapt our method to other application areas in which decision makers need to explore (partly) unknown attribute level ranges (e.g., insurance selection), one of the major advantages of our method being its natural presentation of the attribute level ranges.

Finally, the experiment was carried out in a laboratory setting. Although this was necessary for establishing a sufficient level of control to ensure internal validity, it had the drawback that we could only observe decision makers' stated preferences and thus their (purchase) intentions rather than actual choices. Also, our sample was not drawn to establish representativeness with respect to any part of the consumer population. In future studies, the external validity of our proposed method ought to be tested and compared to that of SWING, TRADEOFF, and CONJOINT, for instance in field experiments.

References

Adomavicius, G., Tuzhilin, A., 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17 (6), 734–749.

Akaah, I. P., Korgaonkar, P. K., 1983. An empirical comparison of the predictive validity of self-explicated, huber-hybrid, traditional conjoint, and hybrid conjoint models. Journal of Marketing Research, 187–197.

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

Bettman, J., Johnson, E. J., Luce, M. F., Payne, J. W., 1993. Correlation, conflict, and choice. Journal of Experimental Psychology 19 (4), 931–951.

Borcherding, K., Eppel, T., Von Winterfeldt, D., 1991. Comparison of weighting judgments in multiattribute utility measurement. Management Science 37 (12), 1603–1619.

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

Dawes, R. M., 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.

Dellaert, B., Dabholkar, P., 2009. Increasing the attractiveness of mass customization: The role of complementary on-line services and range of options. International Journal of Electronic Commerce 13 (3), 43–70.

Dellaert, B., Häubl, G., 2012. Searchin in choice mode: Consumer decision processes in product search with recommendations. Journal of Marketing Research 49 (2), 277–288.

Ding, M., 2007. An incentive-aligned mechanism for conjoint analysis. Journal of Marketing Research 44 (2), 214–223.

Ding, M., Grewal, R., Liechty, J., 2005. Incentive-aligned conjoint analysis. Journal of Marketing Research 42 (1), 67–82.

Edwards, W., Barron, F. H., 1994. SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. Organizational Behavior and Human Decision Processes 60 (3), 306–325.

Eisenführ, F., Langer, T., Weber, M., 2010. Rational decision making. Springer.

Fan, Z., Zhang, X., Chen, F., Liu, Y., 2013. Multi-attribute decision making considering aspiration-levels: A method based on prospect theory. Computers & Industrial Engineering 65 (2), 341–350.

Fischer, G. W., 1995. Range sensitivity of attribute weights in multiattribute value models. Organizational Behavior and Human Decision Processes 62 (3), 252–266.

Fishburn, P. C., 1967. Methods of estimating additive utilities. Management Science 13 (7), 435–453.

Forsati, R., Mahdavi, M., Shamsfard, M., Sarwat, M., 2014. Matrix factorization with explicit trust and distrust side information for improved social recommendation. ACM Transactions on Information Systems 32 (4), 17:1–17:38.

Hart, S., Staveland, L., 1988. Development of nasa-tlx (task load index): Results of empirical and theoretical research. Advances in Psychology 52, 139–183.

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. R., 2014. Consideration set heuristics. Journal of Business Research 67 (8), 1688–1699.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., Riedl, J. T., 2004. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 22 (1), 5–53.

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.

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.

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

Keeney, R. L., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. Wiley, New York.

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.

Lahtinen, T. J., Hämäläinen, R. P., 2016. Path dependence and biases in the even swaps decision analysis method. European Journal of Operational Research 249 (3), 890–898.

Levi, A., Mokryn, O., Diot, C., Taft, N., 2012. Finding a needle in a haystack of reviews: Cold start context-based hotel recommender systems. In: ACM Conference on Recommender Systems.

Mustajoki, J., Hämäläinen, R. P., Salo, A., 2005. Decision support by interval SMART/SWING – incorporating imprecision in the SMART and SWING methods. Decision Sciences 36 (2), 317–339.

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.

Pöyhönen, M., Hämäläinen, R., 2001. On the convergence of multiattribute weighting methods. European Journal of Operational Research 129, 569–585.

Pu, P., Faltings, B., Chen, L., Zhang, J., Viappiani, P., 2011. Recommender Systems Handbook. Springer, Ch. Usability Guidelines for Product Recommenders Based on Example Critiquing Research, pp. 511–545.

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.

Silver, N., Dunlap, W., 1987. Averaging correlation coefficients: Should fisher's z transformation be used? Journal of Applied Psychology 72 (1), 146–148.

Sun, M., Steuer, R. E., 1996. InterQuad: An interactive quad tree based procedure for solving the discrete alternative multiple criteria problem. European Journal of Operational Research 89 (3), 462–472.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. Science 185, 1124–1131.

Tversky, A., Kahneman, D., 1991. Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics 106 (4), 1039–1061.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5 (4), 297–323.

Valenzuela, A., Dhar, R., Zettelmeyer, F., 2009. Contingent response to self-customization procedures: Implications for decision satisfaction and choice. Journal of Marketing Research 46 (6), 754–763.

Van Der Land, S., Schouten, A., Feldberg, F., Van Den Hooff, B., Huysman, M., 2013. Lost in space? cognitive fit and cognitive load in 3d virtual environments. Computers in Human Behavior 29 (3), 1045–1064.

Van Ittersum, K., Pennings, J. M., Wansink, B., van Trijp, H. C., 2007. The validity of attribute-importance measurement: A review. Journal of Business Research 60 (11), 1177–1190.

von Nitzsch, R., Weber, M., 1993. The effect of attribute ranges on weights in multiattribute utility measurement. Management Science 39 (8), 937–943.

von Winterfeldt, D., Edwards, W., 1986. Decision analysis and behavioral research. Vol. 604. Cambridge University Press Cambridge.

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

Wansink, B., Kent, R. J., Hoch, S. J., 1998. An anchoring and adjustment model of purchase quantity decisions. Journal of Marketing Research 35 (1), 71–81.

Weber, M., Borcherding, K., 1993. Behavioral influences on weight judgments in multiattribute decision making. European Journal of Operational Research 67 (1), 1–12.

Weng, L.-T., Xu, Y., Li, Y., Nayak, R., 2008. Exploiting item taxonomy for solving cold-start problem in recommendation making. In: IEEE International Conference on Tools with Artificial Intelligence.

Wu, G., Markle, A. B., 2008. An empirical test of gain-loss separability in prospect theory. Management Science 54 (7), 1322–1335.

Wyner, G. A., 1992. Uses and limitations of conjoint analysis-part ii. Marketing Research 4 (3), 46–47.

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

Xu, A. J., Wyer, R. S., 2010. Puffery in advertisement: The effects of media context, communications norms and consumer knowledge. Journal of Consumer Research 37 (2), 329–343.

Xu, J., Benbasat, I., Cenfetelli, R. T., 2014. The nature and consequences of trade-off transparancy in the context of recommendation agents. MIS Quarterly 38 (2), 379–406.

Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., Norick, B., Han, J., 2014. Personalized entity recommendation: A heterogeneous information network approach. In: ACM International Conference on Web Search and Data Mining.

Yue, S., Larson, M., Hanjalic, A., 2014. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. ACM Computing Surveys 47 (1), 3:1 – 3:45.

Zhao, W. X., Li, S., He, Y., Chang, E. Y., Wen, J.-R., Li, X., 2016. Connecting social media to e-commerce: Cold-start product recommendation using microblogging information. IEEE Transactions on Knowledge and Data Engineering 28 (5), 1147–1159.

Zigoris, P., Zhang, Y., 2006. Bayesian adaptive user profiling with explicit & implicit feedback. In: ACM International Conference on Information and Knowledge Management.