ESTIMATING OPTIMAL RECOMMENDATION SET SIZES FOR INDIVIDUAL CONSUMERS

Completed Research Paper

Michael Scholz
Faculty of Business Administration and Economics
University of Passau
Innstr. 43, 94032 Passau, Germany
michael.scholz@uni-passau.de

Verena Dorner
Faculty of Business Administration and Economics
University of Passau
Innstr. 43, 94032 Passau
verena.dorner@uni-passau.de

Abstract

Online consumers must burrow through vast piles of product information to find the best match to their preferences. This has boosted the popularity of recommendation agents promising to decrease consumers' search costs. Most recent work has focused on refining methods to find the best products for a consumer. The question of how many of these products the consumer actually wants to see, however, is largely unanswered. This paper develops a novel procedure based on signal detection theory to estimate the number of recommendable products. We compare it to a utility exchange approach that has not been empirically examined so far. The signal detection approach showed very good predictive validity in two laboratory experiments, clearly outperforming the utility exchange approach. Our theoretical findings, supported by the experimental evidence, indicate conceptual inconsistencies in the utility exchange approach. Our research offers significant implications for both theory and practice of modeling consumer choice behavior.

Keywords: Recommendation Agent, Decision Analysis, Multiple Attribute Utility Theory, Signal Detection Theory

Introduction

Online consumers today need to sift through vast piles of information to find the products that best match their preferences. This has boosted the popularity of recommendation agents which promise to decrease consumers' search costs. But do they? Recommendation agents certainly reduce the number of websites the consumer needs to search. If the agents structure product information homogeneously across a range of goods, they also reduce cognitive costs associated with processing that information (Montgomery et al. 2004). These reductions, however, are often offset by increased specification costs.

MyProductAdvisor.com, for example, lets consumers define criteria in a process consisting of up to 8 steps, each containing several questions. Consumers who specify levels for all criteria exert quite a lot of effort in the attempt to circumscribe their preferences as precisely as possible. If the recommendation agent returns a large product set, consumers need to spend considerable time on comparing the recommended products or narrowing their selection criteria. Some might even decide to switch to another recommendation agent (Punj and Moore 2009). Large recommendation sets not only lead to higher decision costs but also to decreased decision quality, to declining trust in the recommendation agent and an increasing inclination toward choice deferral (Xiao and Benbasat 2007; Kuksov and Villas-Boas 2010). In effect, large sets reduce the benefits from using the recommendation agent.

Small recommendation sets, on the other hand, do not necessarily lead to higher consumer satisfaction. Evidence from recent empirical studies shows that decision certainty and post-choice satisfaction can be adversely affected by reducing recommendation set size (Lapersonne et al. 1995; Dellaert and Häubl 2012). White and Hoffrage (2009) dubbed this phenomenon the 'tyranny of too much choice' versus the 'allure of more choice'. Although insights into the workings of this phenomenon are slowly forthcoming in informations systems research (Farag et al. 2003; Parra and Ruiz 2009; Gu et al. 2011) and marketing research (Häubl and Trifts 2000; Dellaert and Häubl 2012), empirical evidence is still incomplete and in some parts contradictory. In recent studies, context factors such as the intensity of information load (Farag et al. 2003; Parra and Ruiz 2009), the presence of a recommendation agent (Häubl and Trifts 2000) and personal factors such as age or education (Farag et al. 2003) have been examined. So far, no single indicator or sets of indicators have been confirmed to reliably predict how many recommendations a consumer would like to be presented with at a specific online shopping occasion. We therefore focus on approaches for predicting the recommendation set size independent of contextual or personal factors.

We develop and successfully test a novel approach to measuring and predicting optimal individual recommendation set sizes. Our approach is based on signal detection theory which, although it is widely known, has never been used before as the theoretical foundation for assessing the optimal size of individuals' recommendation sets. We also adapt and empirically examine the utility exchange approach introduced by Butler et al. (2001). Our findings not only contribute to recent research on predicting consideration set size but also provides two approaches that can be integrated into existing recommendation agents to estimate the consumers' willingness-to-pay. Marketing managers could estimate market segments more precisely if they knew individual recommendation set sizes. In this paper, we focus on the consumer perspective and demonstrate how accurately the two mentioned approaches can predict the number of recommendations a consumer will consider.

Our paper is organised as follows. First, we give a short overview over previous research in this area. Next, we briefly introduce the utility exchange approach (Butler et al. 2001) which we empirically examine for the first time in this paper. We then describe our novel approach that we compare with the utility exchange approach. Subsequently, we present our research methodology, data analysis and results. Finally, we point out some limitations to our research and opportunities for research and practice resulting from our findings.

Literature Review

We use the well-known two-stage model by Hauser and Wernerfelt (1990) to describe the consumer choice process. Consumers reduce information gathering and processing costs by first screening all accessible alternatives heuristically (Gilbride and Allenby 2004) and only then evaluating the remaining alternatives with cognitively more expensive compensatory decision rules (Bettman and Park 1980). The final purchase choice is made among the consideration set which consists of all alternatives that 'survive' both stages (Hauser and Wernerfelt 1990). A perfect recommendation agent would recommend just the consideration set, which contains the alternatives with the highest utilities. We will use the term "recommendation set" throughout the article to denominate a consumer's consideration set as identified by a recommendation agent (Häubl and Trifts 2000).

Research into consumer online shopping behaviour and recommendation agents thus needs to address two issues. The first issue is correctly idenfying (and recommending) products which correspond to the consumer's preferences. The second issue is finding the correct number of recommendations to present to the consumer in order to mitigate the dilemma of the 'tyranny of choice versus the allure of more choice' (White and Hoffrage 2009).

In this paper, we propose two approaches for predicting the number of recommendations a consumer is willing to evaluate in depth (i.e., the consideration set). Both approaches are feasible if it is possible to estimate the parth-worth functions for all product attributes. This is the case for recommendation agents which are built on multiple attribute utility theory (MAUT). MAUT-based recommendation agents have received increasing attention in recent research (Huang 2011; DeBruyn et al. 2008; Theetranont et al. 2007) because they have three major advantages over other – especially content-based and collaborative filtering-based (Ansari et al. 2000) – recommendation agents. First, it is not necessary to compile a shopping profile for the consumer before being able to give recommendations (start-up problem). Second, consumer preferences are up-to-date and purchase-related; other recommendation agents face the problems that preferences change over time and that preference estimates can be biased by gift purchases or purchases on behalf of someone else. Third, MAUT-based recommendation agents are able to provide transparent explanations to the customer why a certain product has been recommended. Consumer acceptance and usage of recommendation agents largely depends on whether they trust the agent. The transparency of the recommendation process is a major determinant of the degree of trust consumers put in an agent (Xiao & Benbasat 2007). Explicit preference elicitation is generally regarded as a more transparent task, which makes users more willing to follow the resulting recommendations (Kramer 2007). In addition, Xiao & Benbasat (2007) cite evidence that "compensatory recommendation agents lead to better decision quality and higher decision effort".

The two approaches we discuss here are extensions for existing MAUT-based recommendation agents (e.g. Dell Advisor, Online Insight, My Product Advisor, Plan Smart Choice) for predicting individual consumers' optimal recommendation set size.

While there are many methods for approaching the first issue of correctly identifying attractive products (see Xiao and Benbasat 2007 for a comprehensive survey), estimating the correct size of the recommendation set is a largely unsolved problem. Previous studies have shown that consumers who are confronted with large product assortments may suffer from "choice overload" (Iyengar and Lepper 2000). They defer their purchase decision because product comparisons become more difficult and time-consuming (Fasolo et al. 2009) or for fear of making suboptimal decisions (Iyengar et al. 2006). The recommendation agent extension we propose alleviates this problem. It chooses the most attractive products from the entire assortment but only presents the most promising options to the consumers. Since consumers have individual consideration set sizes (Farag et al. 2003), presenting a fixed-size recommendation set would decrease consumer satisfaction with the recommendation agent. Researchers in information systems and marketing have explored to which extent information systems, personal attributes and attitudes[1] determine consideration set size. Häubl & Trifts (2000) found that consumers using a recommendation agent have smaller consideration sets of higher quality, i.e. non-dominated alternatives with higher subjective utilities, than those who do not. This finding was confirmed by Parra and Ruiz (2009), but not supported by Pedersen (2000). Product expertise (also referred to as experience), age, gender, and the stages of the decision process have been shown to influence consumers' perceptions of product attributes (Farag et al. 2003; Gilbride and Allenby 2004; Karjaluoto et al. 2005; Ranjbarian and Kia 2010). Assuming that consumer preferences for products are aggregates of

[1] More complex constructs (e.g. Chakravarti and Janiszweski 2003; Paulssen and Bagozzi 2005; Kuksov and Villas-Boas 2010) require a very high level of effort and disclosure of very personal information from consumers. If they were implemented as mandatory precursors to the search process, consumer dissatisfaction with and distrust toward the recommendation agent would likely increase (Xiao and Benbasat 2007). This renders them ineligible for our purposes.

preferences for attribute levels (Keeney and Raiffa 1993), these factors determine the contents of the consideration set. The studies did not, however, provide conclusive evidence that these factors also determine the size of the consideration set.

Researchers in decision analysis examine the problem from another angle. They try to determine whether a product is a 'correct selection' (in our case, considerable) by applying procedures for statistical ranking and selection (Butler et al. 2001; Chick and Gans 2009). Most of them define a so-called indifference zone, which contains all products that are most likely to be the best product. The lower boundary of the indifference zone corresponds to the last product that looks attractive enough to the consumer to warrant the effort of evaluating it in detail. Although indifference zone procedures are usually used to find only one (best) product, the parameter $\delta^*$is a very convenient means to estimating the (best) number of recommendations. All products in the indifference zone are, per definition, considerable in the eyes of the consumer. Butler et al. (2001) couple the indifference zone approach with multiple attribute utility theory (MAUT, Keeney and Raiffa 1993), which they use to compute product and attribute utilites. The indifference zone δ^* can then be expressed in terms of utility units. Consumers state δ^* in terms of easily comprehensible attribute units such as costs, which are subsequently converted into utility differences.

The utility exchange approach has, to the best of our knowledge, never been examined empirically before. We compare it to our own approach to determining the best number of recommendations for a consumer, which is based on signal detection theory. This is an entirely novel approach based on the observation that deciding whether a particular product is within the indifference zone is similar to discerning signal and noise. How well a consumer performs on this decision task depends both on the difficulty of the task, i.e. how similar an unattractive and an attractive product are to each other, and on the ability of the consumer to correctly identify an attractive product as such (Tanner and Swets 1954). Like Butler et al. (2001), we used MAUT to obtain product utilities and specify δ^* in terms of utilities. In contrast to (Butler et al. 2001), our signal detection approach does not require consumers to go through the steps of the utility exchange, i.e. specifying the magnitude of perceived product differences in terms of costs.

Theoretical Foundations

At the core of all MAUT-based recommendation agents lies the method with which consumer-specific utility functions for all interesting product attributes are elicited. Recent research has investigated the applicability of methods such as direct specification (Cao and Li 2007), choice-based conjoint analysis (De Bruyn et al. 2008), SMARTER (Huang 2011) or radial basis function networks (Huang 2011) in recommendation agents. Although these methods are mathematically distinct, they share the same theoretical foundation (MAUT) which we will discuss in the next subsection. The two approaches we propose for estimating recommendation set size are also based on MAUT. Neither depends on the recommendation agent implementing any particular preference elicitation method. Both approaches are thus an extension of recent research into MAUT-based recommendation agents. In the next subsection but one, we present a theoretical framework of MAUT extensions for identifying the optimal numbers of decision objects (i.e. recommendable products) for individuals.

Multiple Attribute Utility Theory

Multiple attribute utility theory (MAUT) is one of the most frequently applied decision analysis instruments (Wallenius et al. 2008). Decision objects are analysed in a two-step procedure: their attributes are first evaluated in single-attribute (SAU) functions which are then combined into a multi-attribute (MAU) function. In our case, the prospective digital camera purchaser is asked to perform attribute or product comparisons. The recommendation agent then computes SAU and MAU functions. Based on the digital camera utility estimates which the MAU function yields, the recommendation agent finally sorts the digital cameras according to our consumer's preferences.

Let $u_i(x_i)$ be the SAU function for attribute i, x_i the outcome or level of attribute i, and w_i the weight for attribute i. We can formally express the additive function for assessing product utility $u(X)$ as

$$u(X) = \sum_{i=1}^{n} w_i u_i(x_i) \tag{1}$$

where $0 \leq w_i \leq 1$ and $\sum_{i=1}^{n} w_i = 1$. The MAU function is usually assumed to be additive. Although additive MAU functions seem somewhat unrealistic, they have been shown to be very robust in most situations (Butler 1997).

In MAUT $u_i(x_i)$ refers to the SAU function of attribute i. One simple form of a SAU function is

$$u_i(x_i) = a_i + b_i x_i \tag{3}$$

A linear relation between the attribute utility u_i and the attribute level x_i is often imprecise. A popular alternative expression that also covers attribute utility functions with diminishing or increasing utilities is

$$u_i(x_i) = a_i - b_i e^{x_i c_i} \tag{4}$$

where c_i is the consumers risk tolerance and a_i and b_i are scaling constants. In contrast to equation 3, equation 4 is more flexible and allows modelling concave as well as convex functions. Both types of SAU functions can be estimated with various methods such as multiple regression, AHP or direct trade-offs (see Schoemaker and Waid (1982) and Pöyhönen and Hämäläinen (2001) for an overview and comparison of some).

Selection Framework

Let us assume a set of K products $(X_1, X_2, \ldots, X_K)$ that are ranked according to their expected utility values $E[u(X_1)] \geq E[u(X_2)] \geq \ldots \geq E[u(X_K)]$. Each product's expected utility is computed based on the MAUT as described above. Due to elicitation errors and consumer uncertainty in specifying utility parameters, we cannot assume the best expected products to be a "correct selection". Recent literature suggests defining an indifference zone of products having a probability greater than 0 to be the best product (Butler et al. 2001, Branke et al. 2007). All products k for which $E[u(X_k)] \geq E[u(X_1)] - \delta^*$ holds true are within the indifference zone (i.e., recommendable). The difference between the best expected product and the indifference value δ^* defines the utility threshold u^* each product must surpass to be recommendable.

For example, let us assume digital cameras that are described by photo resolution and price. Each attribute's utility ranges between 0 and 1. If we assume a utility threshold of $u^* = 1.4$ for a particular consumer, all products with a utility higher than 1.4 are in the indifference zone and hence recommendable (see Figure 1).

Figure 1. Indifference Zone and Utility Threshold

In the next sections, we discuss two approaches – the well-known utility exchange approach and signal detection approach – that both estimate the number of recommendable products by computing an indifference zone. The goal of both approaches is arriving at an estimate for the indifference value δ^*.

Utility Exchange Approach

The basic idea of utility exchange is measuring δ^* in terms of attribute units instead of utility units (Butler et al. 2001). Consider a consumer in the process of purchasing a digital camera, for instance. He would be asked which price difference (between the actual price and a hypothetical price for the digital camera in question) would leave him feeling indifferent towards purchasing the digital camera at either price. This price difference is then transformed into a utility difference which is the consumer's δ^* value. The approach assumes that consumer are willing and able to specify a difference in terms of attribute units following the same train of thought as they would when specifying a difference in terms of utility units.

In the following subsections, we describe how δ^* is predicted by using the utility exchange approach as proposed in (Butler et al. 2001). The section concludes with a discussion on the theoretical advantages and disadvantages of the utility exchange approach.

Predicting δ^*

Having computed the expected utility of each product X_k, products are ranked in order of their expected utilities. All products with an expected utility equal to or greater than the best product's expected utility $E[u(X_1)]$ minus the indifference value δ^* are recommendable. Our digital camera purchaser would be shown a list of cameras sorted in descending order of their utilities by the recommendation agent. The last digital camera to be displayed is the one which he is just slightly favorably disposed towards; the next one's (which is not displayed anymore) utility is just too low to be of interest to our consumer. No assumptions about the form of the MAU and SAU functions used to estimate the expected utility values are necessary. We merely need to know attribute i^* and the constant levels of the other attributes.

Computing Utility Equivalents

Butler et al. (2001) suggest choosing cost or price as attribute i^* for scale transformation since they are most easily understood by consumers. All other attributes are set to constant levels. The level of i^* may become negative if unfortunate choices of values for constant levels lead to low utility values. We avoid this by using the best expected product's attribute levels for all other products.

Let i^* be the first attribute ($i = 1$). After estimating MAU and SAU functions, the level x'_{k1} for attribute i^* and product X_k can be computed by transforming the utility difference between the best expected product X_1 and X_k into a difference in terms of the units used to measure i^*. Given the level for i^* and for each product, we can prompt the consumer to specify an indifference zone δ_1^* for i^* if all other attributes are kept constant. If the levels of attributes other than i^* equal the best expected product's (x'_{11}) levels, the indifference zone is bounded by the best expected product's level of i^* and the indifference zone parameter δ_1^* as specified by the consumer. The utility indifference zone parameter δ^* is given by

$$\delta^* = u(X_1) - u_1(x_{11} - \delta_1^*) \tag{5}$$

where $u_1(x_{11})$ is the utility estimated for the first attribute (i.e. i^*) of the best expected product.

Discussion

The utility exchange approach is theoretically sound and applicable to all forms of MAU and SAU functions. All invertible functions can be used for modelling attribute utilities. Compared to measuring utility values directly, the additional effort involved for the consumer is marginal. He is only asked to specify which level of attribute i^* constitutes an imperceptible difference to the best product's level for i^*. But the assumption that attribute level intervals can be transformed into utility intervals directly is problematic. Let price be the attribute i^*. The consumer is asked to give a price toward which he feels indifferent when comparing it to the price of the best expected product. The resulting indifference interval is transformed into utility units without regard to the possibility that the consumer's sensitivities for price and utility values differ. Although the slope of the price SAU is taken into account for the transformation, there is no evidence that a consumer defines equal consideration set sizes when defining indifference zones based on prices and entire product utilities (third assumption of the utility exchange approach). If this is the case, the number of considerable products does not only depend on interval size, but also on the sensitivity for i^*, and the utility exchange approach will not give correct estimates.

Another problem is the fact that rational consumers feel indifferent only if both prices are exactly equal. If i^* is neither price nor cost, reliable measurements of δ_1^* are difficult to obtain. There may be few or no other attributes that are important to all consumers. Using the utility exchange approach with other attributes instead of price or cost is likely to distinctly limit its scope and usefulness.

Signal Detection Approach

Deciding whether a particular product is inside or outside the boundaries of the indifference zone is similar to discerning signal and noise. Signal detection theory states that a certain value exists for which a person is unable to decide whether it is a signal or noise. This value is equivalent to $E[u(X_1)] - \delta^*$ as previously defined. In the digital camera purchasing situation, the consumer tries to discern cameras that meet his preferences (signals) and unattractive cameras (noise). In order to estimate δ^* based on signal detection theory, the following assumptions must hold:

- Utility probability distribution functions (PDF) are given for all products X_k.
- Consumers rate products according to their individual δ^*.

In the next subsection, we present the theoretical foundations of signal detection theory and describe a novel procedure for predicting δ^*, which is based on the estimation of a utility value perceived to be 'neutral' by the consumer, followed by the estimation of utility PDFs. We conclude with a discussion on the advantages and disadvantages of our signal-detection-based approach.

Theoretical Background

Signal detection theory (SDT) was developed by Tanner and Swets to overcome the inability of traditional psychophysical methods to discriminate between sensitivity d and specificity τ (Tanner and Swets 1954; Macmillan and Creelman 2004). Sensitivity describes how difficult it is to distinguish between signal and noise, or attractive and unattractive products. The consumer's tendency to classify a product as signal or noise is called specificity. Both variables must be taken into account simultaneously, sensitivity being independent of specificity. The consumer evaluates for each product which of the two hypotheses - H_0: product is attractive or H_1: product is unattractive - is more likely to be valid. Product attractiveness for each product X_k is expressed in terms of utility $u(X_k)$.

The decision depends on the fraction $p(u(X_k)|H_1)/p(u(X_k)|H_0)$, which is called likelihood ratio $L(u(X_k))$. In the simplest case, H_1 is always supported if $L(u(X_k)) \geq 1$, and H_0 is always supported if $L(u(X_k)) < 1$. The ratio $p(H_1)/p(H_0)$ is called specificity τ and equals 1 if the subject is risk neutral. Risk-averse subjects have specificities τ greater than 1 because avoiding false alarms is paramount to them. The measurement level u that denotes τ is called the critical point u_c. Let us assume that $p(u(X_k)|H_0)$ and $p(u(X_k)|H_1)$ are given as normal distributions. The greater the difference between signal and noise and the smaller the standard deviation of both distributions, the greater is the sensitivity d. In the general case, where $\sigma(H_0) \neq \sigma(H_1)$, sensitivity is given as:

$$d = \frac{\mu(H_0) - \mu(H_1)}{\sqrt{\frac{\sigma(H_0)^2 + \sigma(H_1)^2}{2}}} \tag{6}$$

The specificity point is equal to the point where $p(H_0)$ and $p(H_1)$ cross if and only if $\tau = 1$. Otherwise, the point of specificity is to the left of the intersection point for risk-seeking and to the right for risk-averse subjects. Subjects will respond with H_1 if $L(u(X_k)) \geq \tau$, and with H_0 if $L(u(X_k)) < \tau$. If we know both probability functions $p(H_0)$ and $p(H_1)$, we can estimate the probabilities for hit, miss, correct rejection and false alarm. A recommendation agent using the signal detection theory approach tries to find a balance between hits and false alarms that best meets the preferences of a particular consumer. This has the practical implication for recommendation agents such as MyProductAdvisor.com of narrowing the recommendation set for risk-averse consumers and expanding it for risk-seeking consumers. The procedure explained in the following sections determines whether a consumer is risk-averse or risk-seeking by computing a consumer's specificity. The final result of the proposed procedure is a utility threshold that indicates which products are unattractive and therefore to be excluded from the recommendation set.

Predicting δ^*

We assume that the products X_k are ranked according to their utilities. Comparing two products X_a and X_b with $E[u(X_a)] > E[u(X_b)]$ is similar to comparing signal and noise. The closer to each other the products (stimuli in the previous section) are, the higher the probability that a respondent cannot distinguish between signal and noise. u is defined as a utility value, and the probability functions of signal and noise are defined by the utility PDF Θ of X_a and X_b, respectively.

If $\Theta[u(X_a)]$ and $\Theta[u(X_b)]$ are given, their intersection defines the critical point u_c if and only if $\tau = 1$. Otherwise, the critical point is defined such that the following equation holds

$$\frac{\Theta(t_c|X_b)}{\Theta(t_c|X_a)} = \tau \tag{7}$$

Before estimating τ, we need to specify X_a and X_b. Since we expect the best expected product X_1 to be recommendable, we set $X_a = X_1$. The product most likely to be not considerable is interpreted as noise. A particular subject's specificity τ gives us the correct critical point u_c, which is the utility value any product must exceed to be regarded as considerable. To estimate τ, we suggest evaluating the product X_c closest to the critical point u_c for $\tau = 1$. The evaluation of X_c allows us to recalculate τ and find the real critical point u_c, which we will call utility threshold u^*. δ^* corresponds to the distance between the expected utility of the best product and the utility threshold. δ^* is predicted as follows:

1. Rank all products in order of their expected utilities.
2. Set $\Theta[u(X_K)]$ as Θ_{noise} and $\Theta[u(X_1)]$ as Θ_{signal}.
3. Compute u_c for $\tau = 1$.
4. Select the product X_c nearest to u_c.
5. Let the consumer evaluate X_c with $h = \mathbb{Z} \in [-r; r]$.
6. Recalculate τ as a^{-h} with a as scaling constant.
7. Compute the critical point as utility threshold u^* for recalculated τ.
8. Estimate δ^* as $E[u(X_1)] - u^*$.

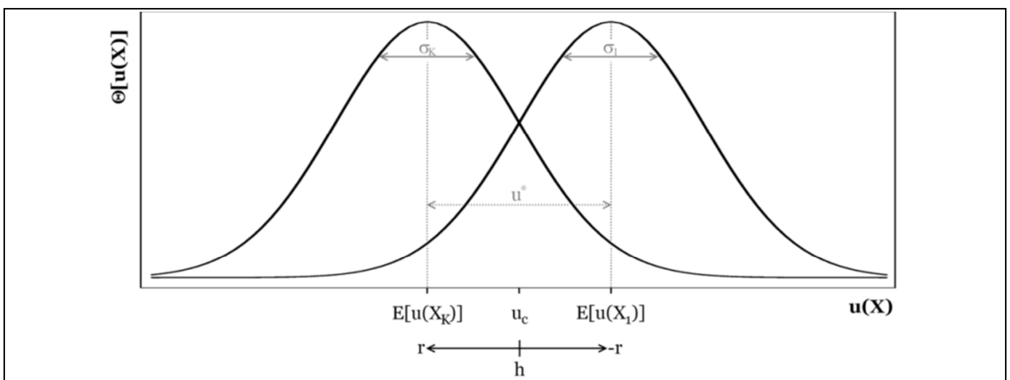

Figure 2. Signal Detection Approach

Like the utility exchange approach, this approach uses MAUT for assessing the expected utility values. Our approach requires, at the very least, information about the utility distributions of the two products which are estimated to be the best and worst product. We assume normally distributed utilities and refer to them as $\Theta[u(X_k)] = (E[u(X_k)], \sigma_k)$. Figure 1 visually demonstrates all variables that need to be computed for normally distributed utilities. The value range of u^* depends on the utility of the indifference product X_c, the scaling constant a and the evaluation scale of h. If the indifference product X_c is exactly between

X_1 and X_K (i.e. $\sigma_1 = \sigma_K$), u^* is scaled between $E[u(X_1)]$ and $E[u(X_K)]$. The location of X_c in turn depends on the expected utilities of X_1 and X_K as well as of their standard deviations.

As demonstrated in Figure 2, signal detection theory provides a theoretical framework for 1) predicting a utility threshold that a product should surpass in order to be considered by a risk-neutral consumer and 2) re-scaling the utility threshold for consumers that either are risk-averse or risk-aware. The validity of correctly estimating u^* depends on the validity of the utility distributions assessed for the best and the worst product and on the sensitivity of the evaluation of X_c. The more close the expected utilities of the best and the worst product, the lower the sensitivity d and thus the harder the evaluation of X_c. In the next section, we describe how u_c is computed (steps 3 and 7). We suggest a method for approximating $E[u(X_k)]$ and σ_k in the subsequent subsection.

Estimating t_c

Given $\Theta[u(X_1)]$ and $\Theta[u(X_K)]$, we can estimate the critical point u_c under the assumption that $\tau = 1$. In the case of normally distributed utilities, we have

$$\mathcal{N}(E[u(X_1)], \sigma_1) = \mathcal{N}(E[u(X_K)], \sigma_K) \tag{8}$$

If $\tau = 1$, the consumer is indifferent towards a product with $E[u(X_c)] = u_c$. We propose selecting that product X_c for evaluation for which the distance between its expected utility and u_c is smallest. Based on the evaluation h of X_c, we can compute τ as follows

$$\tau = a^{-h} \tag{9}$$

where a is a scaling constant. With $h = \mathbb{Z} \in [-r; r]$, we can compute a such that evaluation of $h = -r$ results in the best expected product, and evaluation of $h = r$ in the worst expected product.

If the utility of the products is normally distributed and $\sigma_1 = \sigma_K$, the utility threshold u^* (critical point u_c) is given as

$$u^* = \frac{2\ln(a^h)\sigma_1^2 - E[u(X_1)]^2 + E[u(X_K)]^2}{2E[u(X_K)] - 2E[u(X_1)]}. \tag{10}$$

We can compute the utility threshold u^* as follows if $\sigma_1 \neq \sigma_K$

$$u^* = -\frac{p}{2} + \sqrt{\left(\frac{p}{2}\right)^2 - q}, \tag{11}$$

where

$$p = \frac{2E[u(X_K)]\sigma_1^2 - 2E[u(X_1)]\sigma_K^2}{\sigma_K^2 - \sigma_1^2}$$

and

$$q = \frac{E[u(X_1)]^2\sigma_K^2 - E[u(X_K)]^2\sigma_1^2 - 2\ln\left(\frac{\sigma_K a^h}{\sigma_1}\right)\sigma_1^2\sigma_K^2}{\sigma_K^2 - \sigma_1^2}.$$

In the last step, δ^* is computed as $E[u(X_1)] - u^*$. All products with utility values greater than u^* are inside the boundaries of the indifference zone defined by δ^*, and ought to be recommended to the consumer.

Approximating $E[u(X_k)]$ and σ_k

In the previous sections, we assumed knowledge of the expected utilities and standard deviations of at least X_1 and X_K. Traditional MAUT offers an approach to computing crisp numbers but not distribution functions. We use triangular fuzzy numbers to capture the importance of attributes w_i (Cao and Li 2007).

A triangular fuzzy number $\tilde{q}$ is denoted as $\tilde{q} = \langle q_1, q_2, q_3 \rangle$ where q_1, q_2 and q_3 are real numbers with $q_1 \leq q_2 \leq q_3$. Real numbers in the interval $[q_1; q_3]$ are characterised by a grade of membership to q which is greater than 0. As proposed by Cao and Li (2007), we compute q_1 and q_3 by decrementing (incrementing) w_i by one unit where a rating scale is applied for measuring w_i. If a ranking method is used, we suggest using ranks ra, $ra + 1$ and $ra + 2$ for computing the weight w_i of attribute i ranked on position ra.

Based on the fuzzy weights, we compute overall fuzzy product utilities $u(X_k) = \langle u(X_{k1}), u(X_{k2}), u(X_{k3}) \rangle$ by extending the SAU functions. Each fuzzy utility consists of the 0.0 and 1.0 fractiles and the mode. We can now approximate expected utilities and standard deviations with three-point approximation methods. We chose the extended Pearson-Tukey approach, which Keefer and Bodily (1983) found to be more precise than other methods. We use the 0.05, 0.50 and 0.95 fractiles of one product's utility range to compute its expected utility and standard deviation:

$$E[u(X)] = 0.63u(X)_{0.50} + 0.185[u(X)_{0.05} + u(X)_{0.95}] \tag{12}$$

$$\sigma_K = \frac{u(X)_{0.95} - u(X)_{0.05}}{3.29 - 0.01\left(3.25 \frac{\Delta}{u(X)_{0.95} - u(X)_{0.05}}\right)} \tag{13}$$

where $\Delta = u(X)_{0.95} + u(X)_{0.05} - 2u(X)_{0.50}$.

Discussion

The signal detection approach is theoretically well-founded and can be adopted to different MAU and SAU functions. For estimating τ, we need to define the scaling interval $[-r; r]$. The larger r, the more levels of τ are possible and the better the prediction of δ^*. Compared to the utility exchange approach, the number of parameters specified in advance is virtually equal (a and r for the signal detection approach; attribute i^* and levels of other attributes than i^* for the utility exchange approach), but signal detection theory is based on more assumptions and consists of more processing steps. Thus the signal detection approach theoretically provides more possibilities for incorrect estimates of recommendation set sizes.

The effort required of the consumer is lower than or equal to the effort of using the utility exchange approach. Butler et al. (2001) suggest specifying more than one indifference zone, i.e. using more than one attribute. This would likely improve predictions at the cost of prolonging the recommendation process and requiring more effort on part of the consumer. The signal detection approach could be extended iteratively. We decided to focus on keeping the effort down to a minimum and used only one iteration (one product evaluation) in this paper. In contrast to the utility exchange approach, there are no conceptual inconsistencies between the task presented to the consumer and the computation of δ^*.

Empirical Investigation

We conducted two laboratory experiments, both with a within-subject design, to compare the utility exchange and signal detection approaches. In the first experiment, we used a product set with heterogeneous utility differences between products. In the second experiment, we chose a product set with homogeneous utility differences.

We designed a recommendation agent for the experiments that implements the two-stage decision process by Hauser and Wernerfelt (1990), measures utilities with an additive MAU function, and recommends a number of digital cameras. We chose conjunctive rules for the screening stage. The restrictions placed on the attribute levels by the participants during conjunctive screening were used to generate the stimuli set. The revealed attribute weights determined the order in which the stimuli attributes were presented. Three different attribute levels were generated for each attribute. We used the minimum and maximum acceptable attribute levels to compute the average as the third attribute level, which resulted in a 6x3 D-optimal conjoint design with 18 stimuli for the evaluation stage. We computed the attribute weights w_i with rank-ordered centroids (ROC) that have been shown to be the most precise weighting method for attribute ranking (Barron and Barret 1996). The fuzzy weight for an attribute ranked at position ra was assessed as $w_i = \langle ROC(ra), ROC(ra + 1), ROC(ra + 2) \rangle$.

Procedure

The participants were instructed to search for digital cameras with our recommendation agent. They used it both for initial non-compensatory screening (task A in Table 1) and subsequent evaluation (tasks B and

C). Each participant was first asked to specify upper and lower bounds for each attribute interval (photo resolution, video resolution, zoom factor, monitor size, weight, price). In task B, the participants were asked to rank these six attributes according to their importance and, in task C, to conduct a conjoint analysis by ranking the 18 stimuli.

Task D consisted in rating a number of digital cameras recommended by the system on a 7-point scale ranging from -3 (not considerable) to 3 (considerable). The product list the participants were shown in task D differed between the two experiments. In the first experiment, the participants were presented with product X_c, the best, the worst, and 8 randomly selected products (the product base contained 100 digital cameras). In the second experiment, we created a result set consisting of products with more homogeneous utilities by showing the participants the 10 best products and product X_c. We decided to show 10 products plus the product X_c because we found – in another experiment with 285 participants searching for a digital camera at Amazon.com – that approximately 90% of participants considered 11 or less products (mean=7.07, SD=3.44) before making their purchase decision. Our participants were told that the 11th product had the lowest utility value. Based on the product ratings of task D, we estimated the individually optimal recommendation set size with the signal detection approach.

Table 1. Experimental Procedure

	Screening	Evaluation		Consideration Set	Cost Evaluation
Task	A	B	C	D	E
System	Select Products with Minimal and Maximal Attribute Levels	Present Attributes in Random Order	Generate 6x3 D-optimal Stimuli	Generate Product List	Select Best Four Products
User	Restrict Attributes	Rank Attributes	Rank Stimuli	Rate Products	Specify Cost-Differences
Method	Conjunctive Model	Rank-Ordered Centroids	Ranking-based Conjoint Analysis	Signal Detection Approach	Utility Exchange Approach

Finally, the participants were asked (task E): 'At which price p would you perceive a product to be significantly better than an identical product for the (reference) price of ... ?' This corresponds to a utility exchange approach with the attribute 'price' as i^*. We used the best four products' recalculated prices as reference prices. All other attributes were set to the levels of the best expected product X_1 and the products' ranks were updated. The input of task E formed the basis for estimating the individually optimal recommendation set size with the utility exchange approach.

Pretest and Sample

We conducted one-on-one pre-tests with 8 students who did not take part in the final experiments. We used the think-aloud method for eliciting the participants' opinions of and thoughts on every step of the experiment. After the experiment, we interviewed each participant, asking specifically for suggestions to improve the prototype. All suggestions made by at least 2 participants were implemented.

For the first experiment, 93 students from the University of Passau were invited to a lab and given instructions how to proceed. For each participant's conjoint analysis results (task C), we computed a R^2 value indicating how well the estimated utility functions explained the participant's revealed stimuli ranking. Both the utility exchange and the signal detection approach assume that participants' preferences are consistent and reliable. High R^2 values indicate that this assumption holds true. We conducted an F-test on the R^2 value in order to identify participants that did not rank the 18 stimuli in task C consistently. 80 of the 93 participants passed the F-test. 88 other students from the same university were invited to the second experiment. 68 of these passed the F-test. Each participant was paid US$10. All participants who failed the F-test, thus violating both approaches' basic assumption, were excluded from further analysis. In total, 68.5% of the participants were female and average age was 23 years, ranging from 19 to 36. No significant differences between the participants of either experiment were explained by gender or age.

Analysis and Results

We first checked the reliability and validity of the product utilities' estimates to make sure that the assumption of products being ordered by their utility values (required by both approaches) holds. We then estimated an indifference zone parameter δ that fit the product ratings best. Having obtained the individual product utilities and δ values, we then compared the δ^* values predicted by both approaches to δ. We also used precision and recall to evaluate both approaches.

Utility Elicitation

Price was restricted most severely in both experiments, followed by zoom. Photo resolution displayed the lowest average restriction intensity in both experiments. After tasks B and C (ranking the attributes and conducting conjoint analysis), price emerged as the most important attribute, followed by photo resolution and zoom. All average weights and average non-standardized utility parameters for each attribute are presented in Table 2.

Table 2. Utility Parameters

Parameter	min (x_i)	max (x_i)	Experiment 1				Experiment 2			
			Estimate		Weight		Estimate		Weight	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Intercept	-	-	15.02	118.94	-	-	19.37	101.38	-	-
Photo Res.	5.00	15.90	4.97	6.46	0.20	0.18	5.00	5.86	0.23	0.21
Video Res.	0.31	2.07	5.40	11.61	0.09	0.12	4.78	13.43	0.09	0.14
Zoom	1	35	1.25	1.57	0.17	0.19	1.30	1.53	0.18	0.16
Display Size	2.40	3.50	1.67	19.32	0.05	0.05	0.01	13.23	0.04	0.03
Weight	114	732	-0.03	0.05	0.09	0.11	-0.02	0.05	0.08	0.10
Price	49	800	-0.21	0.22	0.40	0.24	-0.21	0.22	0.38	0.25

Average R^2 was 0.90 (adj. $R^2 = 0.84$) in the first and 0.90 (adj. $R^2 = 0.85$) in the second experiment, indicating reliable utility measurements. This alone, however, does not necessarily imply that participants are satisfied with the constructed utility functions. We measured predictive validity by comparing the predicted product ranks with observed product ratings. The results, with first-choice hit rates[2] of 0.71 (first experiment) and 0.74 (second experiment) and rank correlations of 0.57 and 0.52 respectively, were very satisfactory, especially when compared to other conjoint studies. (Green et al. 1993) report correlations between 0.53 and 0.67, and first-choice hit rates between 0.37 and 0.44. (Moore 2004) reports first-choice hit rates between 0.37 and 0.78. In addition, we found a high correlation between a product's estimated rank of and its observed rating (0.47 in the first and 0.57 in the second experiment).

We may safely conclude that the assumption that products are ordered according to their expected utilities holds for our experiments, and that we can use the utility PDF of the best expected product as signal PDF. Table 3 shows that $\Theta[u(X_1)] = \Theta_{signal}$ and $\Theta[u(X_K)] = \Theta_{noise}$ hold: the best expected product was evaluated as most considerable and the worst expected product as least considerable.

[2] The first-choice hit rate indicates what proportion of participants would like to purchase the best expected product.

Although the best expected product has the highest average rating, the first-hit choice rates of 0.71 and 0.74 respectively indicate that for more than 25% of each experiment's participants the best expected product was not a 'correct selection'. This underlines the importance of computing a set of recommendable products larger than 1.

Table 3. Average Ratings of Products in the Result List (Maximum=3, Minimum=-3)

Expected Rank	Experiment 1		Experiment 2	
	Mean(Rating)	SD(Rating)	Mean(Rating)	SD(Rating)
1	1.45	1.53	1.57	1.30
2	0.70	1.50	0.29	1.46
3	0.10	1.60	-0.09	1.30
4	-0.25	1.46	-0.28	1.35
5	-0.53	1.56	-0.47	1.34
6	-1.06	1.45	-0.79	1.53
7	-1.46	1.39	-0.84	1.42
8	-1.48	1.39	-1.03	1.43
9	-1.96	1.31	-1.29	1.60
10	-2.28	1.04	-1.44	1.49
11	-2.47	0.92	-1.90	1.35

We found sensitivities of d =2.25 (sd=1.47) in the first and d =1.45 (sd=1.22) in the second experiment, indicating that it was much harder for the participants in the second experiment to distinguish between signal and noise than for the participants in the first experiment. This is due to the fact that the products each participant was asked to rate (in addition to the best product) were selected randomly in the first experiment and in descending order of their expected utility values in the second experiment.

Indifference Zone δ

Products rated at least 4 out of 7 were interpreted as positive ratings (0), and all others as negative ratings (1). This coding permitted us to use logistic regression analysis for estimating each participant's δ value separating considerable (0) from not considerable (1) products.

We examined δ by means of evaluating precision, recall, and F-measures. We used precision to measure the fraction of products in the recommendation set that were rated 'considerable', obtaining a value of 0.77 in the first and the surprisingly high value of 0.78 in the second experiment. Recall, indicating the fraction of considerable products the consideration set contained, reached the high level of 0.81 in the first and the acceptable level of 0.67 in the second experiment. The F-measure amounted to 0.79 in the first and 0.72 in the second experiment, which points to an outstanding balance between precision and recall. The estimated threshold values are evidently appropriate for predicting whether a product is considerable for a consumer.

Indifference zone values δ were normalized to the interval [0;1] as follows for comparing δ between all participants:

$$\delta_{\text{normed}} = 1 - \frac{E[u(X_1)] - \delta}{E[u(X_1)] - E[u(X_K)]} \qquad (14)$$

The first experiment produced a mean δ_{normed} value of 0.30 (sd=0.22) and the second experiment of 0.18 (sd=0.20). Participants tended to consider a significantly larger number of products when the utilities were broadly distributed ($p < 0.001$). High standard deviations for δ_{normed} support the assumption that consumers have individual utility thresholds.

Predictive Validity of δ^*

We computed precision, recall, and the F-measure for each approach and experiment based on the participant's ratings and the δ^*_{normed} values from each approach. Additionally, we computed error values as the distance between δ_{normed} and δ^*_{normed}, which is $1 - (E[u(X_1)] - \delta^*)/(E[u(X_1)] - E[u(X_K)])$. The results (Table 4) show that the signal detection approach outperforms the utility exchange approach on all indicators for both heterogeneously and homogeneously distributed product utilities. The only case where the utility exchange approach proved superior was the recall value (with identical error values, p=0.29) for homogeneously distributed product utilities. In the case of heterogeneously distributed product utilities, we found a significant difference between the error values (p<0.01).

Table 4. Predictive Validity of Experiment 1 and 2

	Precision		Recall		F-Measure		Mean(Error)		SD(Error)	
	Exp1	Exp2	Exp1	Exp2	Exp1	Exp2	Exp1	Exp2	Exp1	Exp2
Utility Exchange	0.59	0.49	0.52	**0.56**	0.55	0.52	0.27	0.23	0.23	0.25
Signal Detection	**0.70**	**0.77**	**0.59**	0.47	**0.64**	**0.58**	**0.20**	0.21	**0.18**	0.15
Reference	0.77	0.78	0.81	0.67	0.79	0.72	--	--	--	--

The utility exchange approach tended to underestimate δ^* (Figure 3). The effect was more pronounced in the case of heterogeneously distributed product utilities. Figure 3 also demonstrates that signal detection theory fitted the δ values better on average, but that the δ^* produced larger variance than the reference and δ^* values predicted by the utility exchange approach.

If $\delta^* \rightarrow 0$, only the best expected product is part of the consideration set. In the first (second) experiment, 80% (81%) gave the best expected product a rating of at least 4 points, indicating that this product was considerable. Thus, if $\delta^* \rightarrow 0$, a precision value of around 0.8 is predicted. The utility exchange approach did not consistently underestimate δ^*. We found a negative correlation between underestimation of δ^* and the reference price (r=-0.18, p=0.001) in the first but not in the second experiment (r=0.04, p=0.552). A monotonically descending convex SAU function would improve the prediction of δ^* in the first experiment, but not in the second experiment. Convex utility functions for price point to irrational behaviour on part of the consumers. Compared to linear utility functions, they lead to larger utility differences between small prices and to smaller differences between large prices. This leads us to suspect that conceptual inconsistencies between defining price differences and specifying utility thresholds may exist.

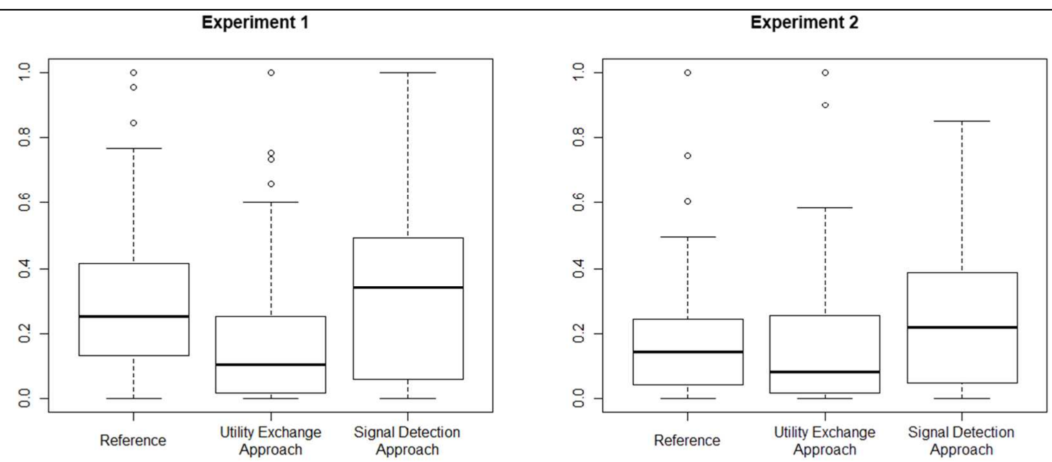

Figure 3. Boxplot of Normed Thresholds

Application

Our experiments revealed higher accuracy for the signal detection approach. In this section, we illustrate how this approach could be integrated into an existing MAUT-based recommendation agent. We consider the *computer advisor* at Dell.com (Figure 3). In a first step (i.e. screening phase), consumers can specify the price range for their new computer and select the type of computer they are planning to buy. In a second step (i.e. evaluation phase), consumers can use sliders to define their importance w_i for 6 attributes. Given a predefined relationship between an attribute's outcome level x_i and its utility $u_i(x_i)$, we can compute each product's utility with equation 1. The utility of the best outcome for an attribute i is then defined as 1, and the utility of the worst outcome 0.

Based on the information imparted to Dell's computer advisor by a consumer, we can compute the best expected product X_1 and the worst expected product X_K as well as the expected utilities and utility deviations with equations 12 and 13. We can then use equation 10 (if $\sigma_1 = \sigma_K$) or equation 11 (if $\sigma_1 \neq \sigma_K$) to estimate the utility u_c of the indifference product. Finally, we can select the computer X_c whose utilities are closest to u_c. The consumer using Dell's advisor will not have discerned any alterations to the recommendation process so far. The only novel step we now introduce to the process is asking the consumer to evaluate X_c. Once we have this last bit of information to feed to equation 10 or 11, we can compute the utility threshold u^* that a computer has to surpass to be included in the recommendation set.

If Dell were to extend its advisor to allow consumers to specify the importance of the price attribute, it would become possible to compute each consumer's willingness-to-pay. In Figure 4, we see that the consumer has rated the attribute "Email, Social Networking, Photo Sharing" with 3, "Mobility, Portability" with 4 and so on. If we knew how this consumer had rated price (let us say with 3), we would be able to estimate the utility for any computer. Let us assume that a hypothetical computer's attribute utilites $u_i(x_i)$ are the following: $u_{Email} = 0.5$, $u_{Mobility} = 0.6$, $u_{Media} = 0.3$, $u_{Photo} = 0.4$, $u_{Homework} = 0.8$, $u_{Gaming} = 0.2$, $u_{Price} = 0.6$. By multiplying the attribute's weightings from Figure 4 with their utilities, we get a total utility of 9.5. If the utility threshold u^* were, for example, 9.3, we could increase the price of the computer until the utility of the price attribute has decreased to 0.4. Let us say that the maximal price of our hypothetical computer is US$2,500 and the minimal price US$500. The utility for the price attribute is then computed by $u_{price} = -0.0005\text{price} + 1.25$. A decrease of 0.2 utility units equals, in this example, a price increase of US$400. We can therefore increase the price to up to US$1,700 for this particular consumer, for whom a price utility of 0.6 equals US$ 1,300.

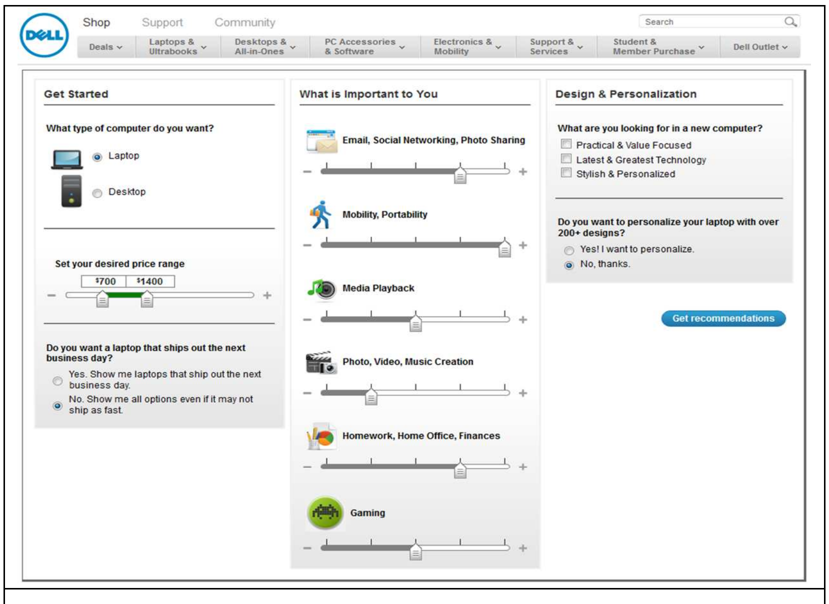

Figure 4. Screenshot of Dell's Computer Advisor

Discussion

Catering to consumer needs is not only a matter of capturing product preferences, but also of estimating how many products will be considered in a purchase decision. This paper introduces a novel approach – signal detection - for predicting the number of products a recommendation agent ought to present to a consumer. Another approach - utility exchange – is proposed as an alternative to estimating the number of recommendable products. Our empirical results indicate that the signal detection approach is better at predicting the number of recommendable products than the utility exchange approach.

Implications for Practice and Research

Our main theoretical contribution is the development of a novel approach that is based on the well-known signal detection theory. Our approach showed good predictive validity in two laboratory experiments. We also shed light on how the utility exchange approach by (Butler et al. 2001), which produced merely acceptable level of predictive validity, may be adapted in future research. We believe that by developing the utility exchange approach further, estimates of consumers' willingnes-to-pay[3] (WTP) could be vastly improved. They can be carried out if reliable and valid estimates of consumer MAU functions and utility thresholds are available (Jedidi and Zhang 2002).

From a recommendation agent provider's point of view, several of our findings hold interesting implications. Instead of being forced to overhaul their products to integrate consideration set size prediction, all they need to do is integrate an add-on to existing MAUT-based recommendation agents. We suggest adding the signal detection-based extension, which our experiments showed to have higher predictive validity than the utility exchange. The effort involved in the search process remains virtually constant. Our approach is a solution to the dilemma between the 'tyranny of too much choice' and the 'allure of more choice' (White and Hoffrage 2009). Let us consider MyProductAdvisor.com. This extension would change the recommendation process only very slightly. The 8 previous (optional) steps in the criteria specification process would be supplemented by a ninth step where the consumer is prompted to rate the product for which he is indifferent if $\tau = 1$. This would 1) decrease consumer information overload, 2) improve consumer perception of MyProductAdvisor.com as a recommendation agent that correctly identifies consumer preferences and therefore 3) improve MyProductAdvisor.com's credibility.

Our findings also benefit brand managers, product developers, and marketing managers. Previous research has shown that extending market models to include segment-level utility thresholds can significantly improve the accuracy of market share and market structure predictions (Jedidi et al. 1996). These models, however, assumed that the consumer attributes determining choice behaviour also determine consideration set formation. This assumption is as yet untested and, since previous findings (as indeed our own) show that attributes like gender or experience may have no impact on consideration set formation, will likely not hold in many scenarios. Our recommendation agent provides the information necessary for testing the old models' accuracy and for developing new models of market prediction. We believe that combining choice and consideration processes will improve segment-level based market estimates dramatically.

Popular market share prediction models use conjoint-based choice simulation. The widely used multinomial rule (Schön 2010) uses segment-based variables to estimate market shares. Each product is assigned a probability to be a consumer segment's choice. Even products that consumers in this segment would not even consider, let alone choose, are assigned a nonzero probability. Our research can help increase the accuracy of such models by improving choice rules. If one were to include information on individual-level thresholds, the models' accuracy would likely increase. Inconsiderable products could be identified and be (correctly) assigned a zero choice probability. Our recommendation agent provides this information.

---

[3]Methods for estimating consumers' willingness to pay that are based on the prediction of a utility threshold have been compared in Miller et al. (2011).

Limitations

First, the utility exchange approach produced surprisingly high error levels in our experiments. These results lead us to suspect that the utility exchange approach is based on a faulty assumption: specifying a difference in terms of attribute units and specifying a difference in terms of utility values are apparently not identical tasks in the eyes of consumers. We are, however, not able to explain (yet) what exactly constitutes the difference between cost/price units and utility units and how they may be made to feel more similar. This is a major obstacle towards practical implementations of this approach. We think that translating cost/price units and utility units in two distinct thresholds could improve the approach's validity. We will explore this and other possibilites in future research.

Second, we used digital cameras as products in our experiments and linear SAU functions for computing product utilities. This limits the generalizability of our findings somewhat. We will conduct further empirical studies on both approaches with non-linear SAU functions and other products (experience goods in particular).

Third, the recommendation process we implemented for our experiments allows only one iteration. This makes it very efficient to use for consumers but may not be ideal in terms of predictive accurary of δ. The signal detection approach could be adapted to allow multiple, iterative evaluations of product X_c. The utility threshold predicted in step 7 of the signal detection approach can be used as reference for another product X_c, to be evaluated by the consumer in a second iteration. We can then narrow the interval around the real δ by adapting Θ_{signal} or Θ_{noise}.

Appendix

Table of Notations

$u_i(x_i)$	SAU function for attribute i
x_i	outcome or level of attribute i
w_i	weight for attribute i
$u(X)$	product utility
c_i	consumers' risk tolerance
a_i and b_i	scaling constants
X_k	product k
X_c	product for which a risk neutral consumer is indifferent
δ^*	indifference value
x'_{k1}	level for attribute i* and product X_k $(X_1, X_2, \ldots, X_K)$
$E[u(X_k)]$	expected utility of product X_k
τ	specificity of a consumer *(tendency to classify a product as signal or noise)*
u_c	critical point *(measurement level u that denotes τ)*
d	sensitivity *(degree of difficulty of distinguishing between signal and noise)*
$\Theta[u(X_a)]$	utility PDF of X_a
h	evaluation of the product at u_c
u^*	utility threshold

Proof for Equation 10

$$\frac{1}{\sigma_1\sqrt{2\pi}} e^{-0.5\left(\frac{u^*-E[u(X_1)]}{\sigma_1}\right)^2} = a^{-h}\frac{1}{\sigma_1\sqrt{2\pi}} e^{-0.5\left(\frac{u^*-E[u(X_K)]}{\sigma_1}\right)^2}$$

$$a^h e^{-0.5\left(\frac{u^*-E[u(X_1)]}{\sigma_1}\right)^2} = e^{-0.5\left(\frac{u^*-E[u(X_K)]}{\sigma_1}\right)^2}$$

$$2\ln(a^h)\sigma_1{}^2 - (u^* - E[u(X_1)])^2 = -(u^* - E[u(X_K])^2$$

$$u^* = \frac{2\ln(a^h)\sigma_1^2 - E[u(X_1)]^2 + E[u(X_K)]^2}{2E[u(X_K)] - 2E[u(X_1)]}$$

Proof for Equation 11

$$\frac{1}{\sigma_1\sqrt{2\pi}} e^{-0.5\left(\frac{u^*-E[u(X_1)]}{\sigma_1}\right)^2} = a^{-h}\frac{1}{\sigma_K\sqrt{2\pi}} e^{-0.5\left(\frac{u^*-E[u(X_K)]}{\sigma_K}\right)^2}$$

$$\left(\frac{u^*-E[u(X_1)]}{\sigma_1}\right)^2 + 2\ln(\sigma_1) = \left(\frac{u^*-E[u(X_K)]}{\sigma_K}\right)^2 + 2\ln(a^h\sigma_K)$$

$$u^{*2} + u^*\left(\frac{2E[u(X_k)\sigma_1^2 - 2E[u(X_1)]\sigma_K^2}{\sigma_K^2 - \sigma_1^2}\right) + \frac{E[u(X_1)]^2\sigma_K^2 - E[u(X_K)]^2\sigma_1^2 - 2\ln\left(\frac{\sigma_K a^h}{\sigma_1}\right)\sigma_1^2\sigma_K^2}{\sigma_K^2 - \sigma_1^2} = 0$$

Since

$$\frac{p}{2} < \sqrt{\left(\frac{p}{2}\right)^2 - q} \qquad \forall E[u(X_1)], E[u(X_K)], \sigma_1, \sigma_K > 0$$

the only meaningful solution is

$$u^* = -\frac{p}{2} + \sqrt{\left(\frac{p}{2}\right)^2 - q}.$$

References

Ansari, A., Essegaier, S., and Kohli, R. 2000. "Internet Recommendation Systems," *Journal of Marketing Research* (37:3), pp. 363-375.

Barron, F., and Barret, B. 1996. "Decision Quality Using Ranked Attribute Weights," *Management Science* (42:11), pp. 1515–1523.

Bartolini, I., Ciaccia, P., and Patella, M. 2008. "Efficient Sort-Based Skyline Evaluation," *ACM Transactions on Database Systems* (33:4), pp. 31–49.

Bettman, J.R., and Park, C.W. 1980. "Effects of Prior Knowledge and Experience and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis," *Journal of Consumer Research* (7:3), pp. 234-248.

Branke, J., Chick, S.E., and Schmidt, C. 2007. "Selecting a Selection Procedure," *Management Science* (53:12), pp. 1916-1932.

Butler, J., Jia, J., and Dyer, J.S. 1997. "Simulation Techniques for the Sensitivity Analysis of Multi-criteria Decision Models," *European Journal of Operational Research* (103:3), pp. 531-545.

Butler, J., Morrice, D., and Mullarkey, P. 2001. "A Multiple Attribute Utility Theory Approach to Ranking and Selection," *Management Science* (47:6), pp. 800–816.

Cao, Y., and Li, Y. 2007. "An Intelligent Fuzzy-based Recommendation System for Consumer Electronic Products," *Expert Systems with Applications* (33:1), pp. 230–240.

Chakravarti, A., and Janiszewski, C. 2003. "The Influence of Macro-Level Motives on Consideration Set Composition in Novel Purchase Situations," *Journal of Consumer Research* (30:2), pp. 244-258.

Chick, S., and Gans, N. 2009. "Economic Analysis of Simulation Selection Problems," *Management Science* (55:3), pp. 421–437.

De Bruyn, A., Liechty, J.C., Huizingh, E.K.R.E., and Lilien, G.L. 2008. "Offering Online Recommendations with Minimum Customer Input Through Conjoint-Based Decision Aids," *Marketing Science* (27:3), pp. 443-460.

Dellaert, B.G.C. and Häubl, G. 2012. "Searching in Choice Mode: Consumer Decision Processes in Product Search with Recommendations," *Journal of Marketing Research*, forthcoming.

Farag, N.I., Smith, M., and Krishnan, M. 2003. "The Consumer Online Purchase Decision: A Model of Consideration Set Formation and Buyer Conversion Rate Across Market Leaders and Market Followers," *Proceedings of the 24th International Conference on Information Systems*, Seattle, USA.

Fasolo, B., McClelland, G. H., Todd, P. M. 2007. "Escaping the tyranny of choice: when fewer attributes make choice easier," *Marketing Theory* (7:1), pp. 13–26.

Green, P., Krieger, A., and Agarwal, M. 1993. "A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences," Marketing Letters (4:4), pp. 391-406.

Gilbride, T., and Allenby, G. 2004. "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science* (23:3), pp. 391-406.

Gu, B., Konana, P., and Chen, H. 2011. "Consumer Product Consideration and Choice at Purchase Time at Online Retailers," *Proceedings of the 32nd International Conference on Information Systems*, Shanghai, China.

Hauser, J., and Wernerfelt, B. 1990. "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research* (16:4), pp. 393–408.

Häubl, G. and Trifts, V. 2000. "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Marketing Science* (19:1), pp. 4-21.

Huang, S. 2011. "Designing Utility-based Recommender Systems for E-Commerce: Evaluation of Preference-Elicitation Methods," *Electronic Commerce Research and Applications* (10:4), pp. 398-407.

Iyengar, S. S. and Lepper, M. 2000. When choice is demotivating: "Can one desire too much of a good thing?" *Journal of Personality and Social Psychology* (79:6), pp. 995–1006.

Iyengar, S. S., Wells, R. E. and Schwartz, B. 2006. „Doing Better but Feeling Worse: Looking for the "Best" Job Undermines Satisfaction," *Psychological Science* (17:2), pp. 143–150.

Jedidi, K., and Zhang, Z. 2002. "Augmenting Conjoint Analysis to Estimate Consumer Reservation Price," *Management Science* (48:10), pp. 1350–1368.

Jedidi, K., Kohli, R., and DeSarbo, W. 1996. "Consideration Sets in Conjoint Analysis," *Journal of Marketing Research* (33:3), pp. 364–372.

Karjaluoto, H., Karvonen, J., Kesti, M., Koivumäki, T., Manninen, M., Pakola, J., Ristola, A., and Salo, J. 2005. "Factors Affecting Consumer Choice of Mobile Phones: Two Studies from Finland," *Journal of Euromarketing* (14:3), pp. 59-82.

Keefer, D., and Bodily, S. 1983. "Three-Point Approximations for Continuous Variables," *Management Science* (29:5), pp. 595–609.

Keeney, R., and Raiffa, H. 1993. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs:* Cambridge University Press.

Kramer, T. 2007. "The Effect of Measurement Task Transparency on Preference Construction and Evaluations of Personalized Recommendations," *Journal of Marketing Research* (44:2), pp. 224-233.

Kuksov, D., and Villas-Boas, J. M. 2010. "When More Alternatives Lead to Less Choice," *Marketing Science* (29:3), pp. 507-524.

Lapersonne, E., Laurent, G. and Le Goff, J.J. 1995. "Consideration Sets of Size One: An Empirical Investigation of Automobile Purchases," *International Journal of Research in Marketing* (12:1), pp. 55-66.

Macmillan, N., and Creelman, C. 2004. *Detection Theory: A User's Guide:* Lawrence Erlbaum Associates.

Miller, K.M., Hofstetter, R., Krohmer, H. and Zhang, Z.J. 2011. "How Should Consumers' Willingness to Pay be Measured," *Journal of Marketing Research* (48:1), pp. 172-184.

Montgomery, A.L., Hosanagar, K., Krishnan, R. and Clay, K.B. 2004. "Designing a Better Shopbot," *Management Science* (50:2), pp. 189-206.

Moore, W. 2004. "A Cross-Validity Comparison of Rating-based and Choice-based Conjoint Analysis Models," *International Journal of Research in Marketing* (21:3), pp. 299–312.

Papadias, D., Tao, Y., Fu, G., and Seeger, B. 2005. “Progressive Skyline Computation in Database Systems,” *ACM Transactions on Database Systems* (30:1), pp. 41–82.

Parra, J., and Ruiz, S. 2009. “Consideration Sets in Online Shopping Environments: The Effect of Search Tool and Information Load,” *Electronic Commerce Research and Applications* (8:5), pp. 1-19.

Paulssen, M., and Bagozzi, R. P. 2005. “A Self-regulatory Model of Consideration Set Formation,” *Psychology & Marketing* (22:10), pp. 785–812.

Pedersen, P.E. 2000. “Behavioral Effects of Using Software Agents for Product and Merchant Brokering: An Experimental Study of Consumer Decision Making,” *International Journal of Electronic Commerce* (5:1), pp. 125-141.

Punj, G., and Moore, R. 2009. “Information Search and Consideration Set Formation in a Web-based Store Environment,” *Journal of Business Research* (62:6), pp. 644-650.

Pöyhönen, M., and Hämäläinen, R.P. 2001. “On the Convergence of Multiattribute Weighting Methods,” *European Journal of Operational Research* (129:3), pp. 569-586.

Ranjbarian, B., and Kia, M. 2010. “The Influence of Personality Traits on Consideration Set Size,” *European Journal of Social Sciences* (15:2), pp. 262–269.

Schoemaker, P.J.H, and Waid, C.C. 1982. “An Experimental Comparison of Different Approaches to Determining Weights in Additive Utility Models,” *Management Science* (28:2), pp. 182-196.

Schön, C. 2010. “On the Optimal Product Line Selection Problem with Price Discrimination,” *Management Science* (56:5), pp. 896–902.

Tanner, W., and Swets, J. 1954. “A Decision-making Theory of Visual Detection,” *Psychological Review* (61:6), pp. 401–409.

Theetranont, C., Haddawy, P., Krairit, D. 2007. “Integrating Visualization and Multi-Attribute Utility Theory for Online Product Selection,” *International Journal of Information Technology & Decision Making* (6:4), pp. 723-750.

Wallenius, J., Dyer, J.S., Fishburn, P.C., Steuer, R.E., Zionts, S. and Deb, K. 2008. “Multiple Criteria Decision Making, Multiattribute Utility Theory: Recent Accomplishments and What Lies Ahead,” *Management Science* (54:7), pp. 1336-1349.

White, C., and Hoffrage, U. 2009. “Testing the Tyranny of Too Much Choice against the Allure of More Choice,” *Psychology & Marketing* (26:3), pp. 280–298.

Xiao, B., and Benbasat, I. 2007. “E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact,” *MIS Quarterly* (31:1), pp. 137-209.