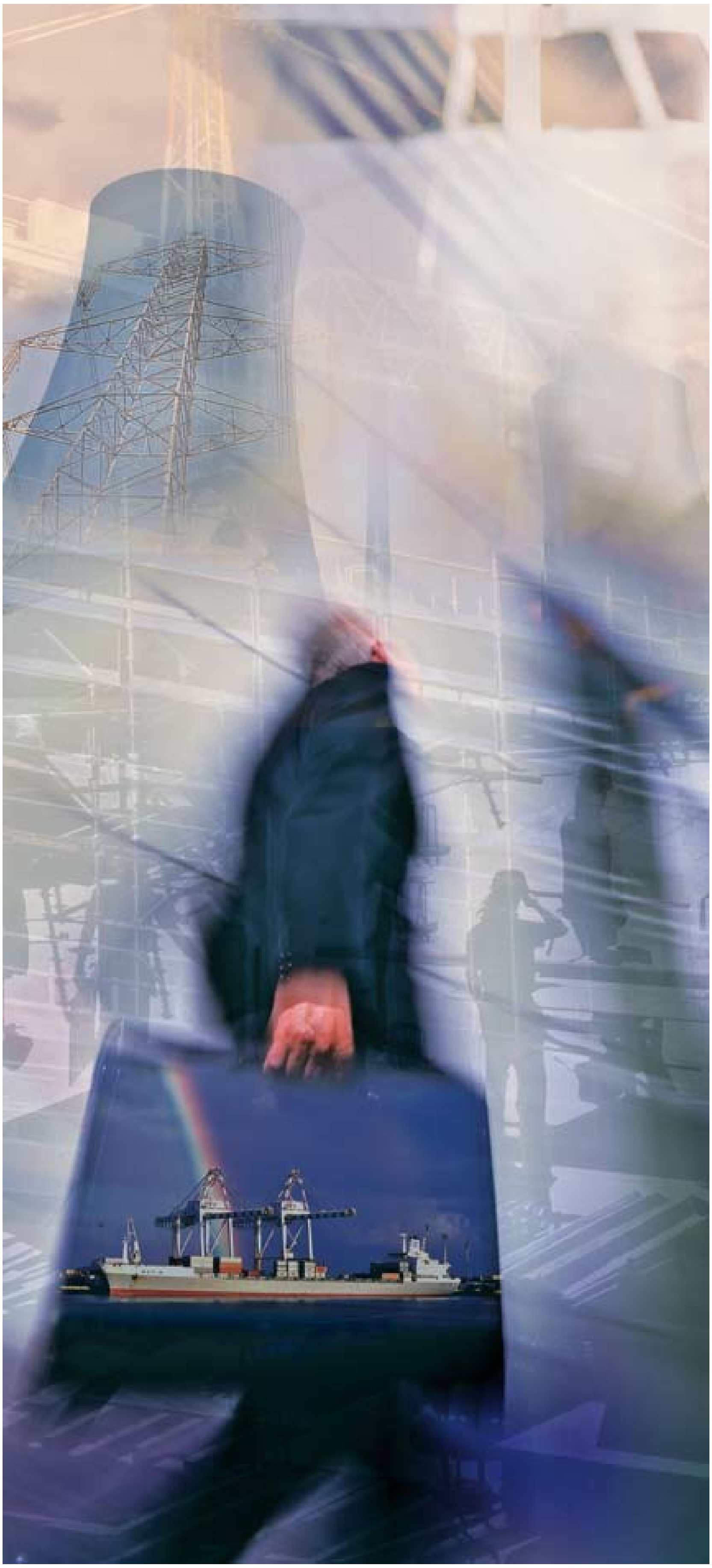

Queensland

The Economic Society of Australia Inc.

Proceedings of the 37th Australian Conference of Economists

Papers delivered at ACE 08

30th September to 4th October 2008
Gold Coast Queensland Australia

ISBN 978-0-9591806-4-0

Welcome

The Economic Society of Australia warmly welcomes you to the Gold Coast, Queensland, Australia for the 37th Australian Conference of Economists.

The Society was formed 83 years ago in 1925. At the time, the Society was opposed to declarations of policy and instead focused on open discussions and encouraging economic debate. Nothing has changed today, with the Society and the conference being at the forefront of encouraging debate.

This year we have a large number of papers dealing with Infrastructure, Central Banking and Trade.

Matters of the greatest global importance invariably boil down to be economic problems. Recent times have seen an explosion of infrastructure spending, after world-wide population growth has seen demand outpace aging supply. The world has become more globalised than at any time since World War I but the benefits of this (and the impact on our climate) has been questioned by some.

At the time of preparing for this conference we could not have known that it would have been held during the largest credit crisis since the Great Depression. The general public and politicians both look to central banks for the answers.

We are also very pleased to see a wide selection of papers ranging from applied economics to welfare economics. An A – Z of economics (well, almost).

Another feature of this conference is that we have gone out of our way to bring together economists from all walks of life, in particular from academia, government and the private sector. We are grateful to all of our sponsors, who are as diverse as the speakers.

The Organising Committee

James Dick
Khorshed Alam (Programme Chair)
Michael Knox
Greg Hall
Allan Layton
Rimu Nelson
Gudrun Meyer-Boehm
Jay Bandaralage
Paula Knight

Published November 2008

GPO Box 1170
Brisbane Queensland Australia
ecosocqld@optushome.com.au

Our Gold Sponsors

QIC

AUDAX AT FIDELIS
QUEENSLAND TREASURY CORPORATION

Reserve Bank of Australia

Queensland Government
Department of Infrastructure and Planning

Keynote Sponsors

Australian Government
Productivity Commission

Australian Government
The Treasury

QUT
Queensland University of Technology
Faculty of Business
School of Economics and Finance

The University of Queensland
Australia

Goldman Sachs JBWere

The University of Queensland
Australia

NCER National Centre for Econometric Research

ES Central Council

Special Session Sponsors

Griffith University

The Higher Education Academy Economics Network

Unless we have specifically been requested to do otherwise, all the papers presented at the conference are published in the proceedings in full. A small number of papers will have versions that have also been made available for special editions of Journals, Economic Analysis and Policy, and the Economic Record. Authors will retain the right to seek additional publication for papers presented at the conference so long as it differs in some meaningful way from those published here.

The opinions expressed in the papers included in the proceedings are those of the author(s) and no responsibility can be accepted by the Economic Society of Australia Inc, Economic Society of Australia (Queensland) Inc, the publisher for any damages resulting from usage or dissemination of this work.

The Paper following forms part of - *Proceedings of the 37th Australian Conference of Economists*
ISBN 978-0-9591806-4-0

Behavior of Online Shoppers

Uwe Dulleck, Franz Hackl, Bernhard Weiss, Rudolf Winter-Ebmer*

This Version May 2008

Abstract

In this article we propose a two stage procedure to model demand decisions by customers who are deciding about several dimensions of a product. We then test our procedure by analyzing the behavior of buyers from an Austrian price comparison site. Although here a consumer will typically search for the cheapest price for a given product, reliability and service of the supplier are other important characteristics of the retailer. In our data, consumers follow such a two stage procedure: they select a shortlist of suppliers by looking at their willingness to pay only; among these shortlisted suppliers finally, they trade off reliability and price to a large extent.

JEL Classifications L..

Keywords e-commerce, price comparison, decision theory, heuristics, seller reputation

*Hackl, Weiss and Winter-Ebmer: Department of Economics, JKU Linz, Austria, Dulleck: School of Economics and Finance, Queensland University of Technology, corresponding author (email): franz.hackl@jku.at.

1 Introduction

Consumers' decision making in realistic purchase situations is often more complex than a standard utility maximization model would predict. Due to time constraints, high search costs, lack of computational power or simply lack of cognitive knowledge in such situations, psychologists and decision theorists in marketing often refer to heuristical procedures (Gigerenzer et al., 1999) which are often able to perform better than more complicated algorithms suggested by economists.

In many cases, two-stage procedures are suggested, where the decision process is structured often with different governing rules in the different stages. For example, in a consumption situation, the consumer has to compare products along many different dimensions (e.g. a car): in a first stage consumers make a shortlist of potential varieties where only a limited number of features of the cars are taken into account. Once this shortlist is given, the consumer inspects the cars more in detail, arranges for a test drive, etc in order to come to a final decision("consider-then-choose model" by Gaskin et al. (2008), Yee et al. (2007)). Similar shortlist ideas are used for literature prizes, recruitment decisions or investment projects [1].

Two-stage decision procedures are related to Herbert Simon's satisficing behavior (Simon, 1957): individuals fix a satisfactory aspiration level and then take the first object which satisfies this criterium. Theories of elimination by aspects (Tversky, 1972) or "fast and frugal heuristics" (Gigerenzer and Goldstein, 1996) evoke lexicographic preferences [2]. In simple cases, where binary decisions on product characteristics are possible, individuals decide which characteristics are the most important ones and then they eliminate products step by step if required characteristics are missing. These heuristics are non-compensatory: bad features of one model cannot be compensated by good performance in another one. While they do describe many decision situations quite well, they suffer from some problems: How do individuals decide about the ranking of features in the lexicographic ordering? How can the non-compensatory principle be upheld if the varieties differ widely in some features, like quality and price?

In the following we will look at consumer decisions of online buyers. Relative to brick-and-mortar retailing, in online markets information is much

[1] Academics are familiar with these procedures from conference paper evaluations: referees are often asked to mark papers with A(definite accept), B(possible accept), C(definite reject), where B papers will be checked by the program committee - the committee decisions follow different rules.

[2] See also Kohli and Jedidi (2007) for experiments on computer buyers and Payne (1976) for an early application on apartment choice.

easier to get and search costs are much lower. An extreme case are so-called price comparison sites or shopbots. On these websites, shoppers can compare prices for thousands of products with one mouse-click: a comfortable list of suppliers with information about prices but also on the reliability of the seller - typically given by evaluations of past customers - is available at practically no cost.

Such markets are close to perfect competition, so we should expect all shoppers to buy at the cheapest price only. Despite of this, price dispersion on the internet is not much lower as in brick-and-mortar stores (Smith and Brynjolfsson, 2000) because buyers value accessability of the site and the shop, reliability of order fulfilment and the modalities of delivery (Betancourt and Gautschi (1993) and Pan et al. (2002)). As information is practically costless in such markets - shoppers can see all the relevant price and quality information on one screen - they have every incentive to make a well-informed decision: they should take all the specifics of the seller into account and trade off a higher price for higher reliability of the seller. In other words, such markets are the least to expect heuristical procedures.

Still we find consumers making shortcuts. Typically, shoppers decide on a shortlist of suppliers where the price is the most important determinant; reliability considerations of the supplier don't play a big role here. In the second stage quality, reliability, and other supplier specific characteristics are much more important aspects of the decision of the consumers.

2 A Decision Procedure

In this section we introduce a simplified model of a decision process. The process is a shortlist method (SM) (Manzini and Mariotti, 2007). We assume that a buyer knows which good he wants to buy. Retailers of this good are plenty and they differ in the additional service they offer. These services maybe availability of the good, pre and post sale services, payment options, etc. Without loss of generality we assume that the good in it's inert characteristics is homogeneous. Let $X = \{x_1, .., x_n\} \subset \mathbb{R}^2_+$, with $n > 2$, denoting the set of available offers, where n denotes the number of retailers offering this product. Each option represents the offer of one seller. We model the offer of a seller as having two components, the price and the service characteristic of the seller: $x_i = (p_i, s_i)$ where p_i is the price seller i charges for the product and s_i is the sellers service characteristic. For simplicity we assume that seller's characteristics can be represented by a single number. One can easily extent the model to allow for product differentiation (i.e. consumers compare several differentiated products) or several dimensions of sellers' characteristics. Given our focus on buyers' decision procedure and for reasons of simplicity, we assume X is exogenously given.

To model the decision of a buyer we assume each buyer has a complete binary preference relation $P \subseteq X \times X$ over all elements in X and we denote by $x \succeq y$, $x, y \in X$, that the customer likes option x is at least as much as option y, i.e. $(x, y) \in P$. This states that we assume that a buyer in principle has complete preferences when asked to choose between any sets of options. Our method is based on the assumption that a buyer first reduces his or her choice set before making a final decision. We need the following notation to identify utility maximizing choices from a set $Y \subseteq X$:

$$\max(Y; P) = \{x_i \,| \nexists z \in Y \text{ such that } z \succ x_i\}.$$

Even though our decision process will model boundedly rational choice, we assume that the preference relation is complete. A SM is characterized as a process that the preference relation is applied only to a shortlist, a set $S(v, X) \subset X$. The procedure we want to study, generates a shortlist S is based on a consumer's individual cut-off price v. This cut-off price maybe be determined by a maximum willingness to pay for the good or indeed it may be the price a consumer expects to see to consider to invest more time to look at the product. We assume v is independent of service characteristics of a retailer. Given that it is a characteristic of a consumer, we assume it is exogenous. The 2nd stage of the decision procedure is to apply the complete preference relation P to $S(v, X)$. Let $C(v, X)$ denote the choice based on this procedure.[3] Our procedure hence consists of two steps:

1. The customer generates a shortlist $S(v, X) := \{x_i \,| p_i < v\}$.
2. Customers choose according to their preferences from $S(v, X)$: $C(v, X) := \max(S(v, X); P)$.

To compare this procedure to the standard rational choice procedure, denote by $R(X)$ the rational choice: $R(X) := \max(X; P)$. If one interprets the service characteristic of a seller as a characteristic of the good, then the otherwise homogeneous product differentiates itself by the retailers service characteristic. To compare our decision process to the literature, the

[3] Mandler et al. (2008) propose a Checklist method that is can provide a choice method that mirrors complete preferences. We do not specify how consumers make their choice of among the items on the shortlist. The checklist method is a good candidate to model this decision, in this case our method just prescribes a first item on the checklist, namely the price.

standard approach in the literature is to follow Rosen (1974). His hedonic pricing approach in our case states that among all the offers one can derive a hedonic price function $p(s) = \min_{\{x_i | s_i \geq s\}} p_i$ which gives the best price for the product given a certain minimum characteristic of the seller. The hedonic pricing approach is another way to simplify the decision by eliminating all options which have a higher price given a certain level of service. A decision is then made among all elements of a restricted set. This yields the same result as $R(X)$. For our process this is not necessarily the case: $\exists (X, v)$ $C(v, X) \neq R(X)$. But, our process allows customers to consider all offers, if they do so then the choice will be the same as in a rational choice decision, i.e. $C(\max p_i, X) = R(X)$.

In the following, we state some of the characteristics of $C(v, X)$. Proofs are omitted, as they are straight forward. For notational purposes denote by $p(C(v, X))$ and $s(C(v, X))$ the respective element of $C(v, X)$:

- If for $\forall x_i \in X \quad p_i = p$, i.e. all sellers charge the same price then $C(v, X) = R(X)$.
- $C(v, X)$ does not fulfill the weak axiom of revealed preferences (WARP): $\exists X, Y$ with $Y \subseteq X$ such that $C(v, X) \neq C(v, Y)$ and $p(C(v, X)) < p(C(v, Y))$.[4]
- For $v_1 < v_2$, iff $s(C(v_2, X)) \geq s(C(v_1, X))$ then $p(C(v_2, X)) \geq p(C(v_1, X))$.
- For $X, Y \subset \mathbb{R}^2_+$, $C(v, X)$ fulfills the Expansion property, i.e. $C(v, X \cup Y) \in \{C(v, X), C(v, Y)\}$.

Note we differ from the rational shortlist methods by allowing for complete preferences over X. Manzini and Mariotti (2007) as well as Kfir and Ok (2006) assume that preferences are not complete, i.e. there exist options where neither $\preceq$ nor $\succeq$ applies. By allowing only two dimensions and having a natural ordering at least on the price, our choice method shares some features with these methods but it comes up with seemingly irrational choices for a different reason. While in the cited procedures WARP is contradicted because the procedures determine which options are eliminated from

[4]This result is the most likely to need a proof. Given that it is negative result, a counter example suffices. Note, given that we have unit demand, a choice process $W(X)$ fullfills WARP iff $Y \subseteq X \Rightarrow$if $W(X) \neq W(Y)$ then $p(W(X)) > p(W(Y))$. See Varian (2006) for a recent survey. This is a counter example: Let $P = \{(x, y) \,|p(x) - s(x) < p(y) - s(y)\}$ i.e. a higher value of s denotes a preferable quality and price and quality are perfectly substitutable. If $X = \{(1, 7), (2, 5), (3, 3)\}$ and $Y = \{(1, 7), (3, 3)\}$ then with $v = 2.1$ $C(2.1, X) = (2, 5)$ and $C(2.1, Y) = (1, 7)$ contradicting WARP.

the choice set given incomplete preferences in our case the short list depends just on one characteristic, namely the price. The main difference to the cited procedures is that our procedure is not affected by irrelevant alternatives.

The present choice procedure shares the dependence on irrelevant alternatives with the procedure suggested by Manzini and Mariotti (2007) or Kfir and Ok (2006) if v is determined endogenously, i.e. v is a function of X. Why? If $v(X)$ is such a function and $X_0 = X \cap \{x_0\}$ then consumers short list $S(v(X), X) \neq S(v(X_0), X_0)$ and thus the choice may change, even though $x_0 \notin S(v(X_0), X_0)$. To see the closer similarity of the procedures in this case, note in the notation of Manzini and Mariotti (2007) $\{x_i, x_j\} \in P_1$ iff $p_i \leq v$ and $p_j > v$, i.e. in our notation . If v is a function X, for example $v(X) = \frac{1}{|X|} \sum p_i$, then indeed irrelevant alternatives can affect the choice because they change $v(X)$, namely if $p(x_0) > \frac{1}{|X|} \sum p_i$.

Compared to the mentioned choice methods, our choice procedure - independent of whether v is exogenous or a function $v(X)$ - always generates a unique decision of a buyer given a set of alternatives X.

In the following empirical section we will take two views on shortlists. The first view is to exogeneously enforce a shortlist, ie. assuming that customers really consider only the - with respect to the price - lowest 5,10 or 20 offers. Our second view is that the decision to put an offer on the shortlist, i.e. to consider an offer further, and the purchase decision are two separate steps. Given that we observe clicks and distinguish them from purchases, we consider all offers to be clicked at least ones as those on the shortlist. Having our simple theory in mind, the following states hypotheses resulting from the proposed procedure.

Hypothesis 1 *If clicks determine the shortlist then prices strongly determine the selection of shops that are visited at least ones.*

Hypothesis 2 *Service will be more important among the sellers that are visited / are on the shortlist.*

3 Data and Estimation Strategy

For our empirical analysis we use the database of www.geizhals.at. This web-site offers a 'price search engine' which collects the price offers via standardized protocols from a predefined group of sellers and presents them electronically via it's web-platform. Typically the quality and reliability of price offers in price search engines are higher and more serious in contrast to 'shopbots' which do an arbitrary price search for products on the whole web and offer the results of this web search online.

Geizhals.at has contracts with about 3000 sellers which can list their price offerings for a total of 280000 products on the Geizhals.at website[5].

Since Geizhals.at is practically a monopolist in providing price comparisons in Austria, this price search engine is well-known and widely used by webshoppers. Hence, all e-tailers have an incentive to get their prices listed and we observe on geizhals.at practically the complete Austrian market for goods traded online. Due to computational limitations we have to restrict our data to an arbitrary week in the year 2006. The data sample in this analysis includes price offers for 43327 products from a total of 440 sellers. From sellers' price offers we know the exact name of the product and the producer together with the products' mapping into a hierarchical classification system for the products (categories, subcategories, and subsubcategories). Furthermore, sellers' price offers include information on availability and shipping charges. Customers have the possibility to evaluate the (service)quality of the firms: 32626 customers did so. Geizhals offers the possibility to evaluate the retailers' service quality on a 5-point scale between 1(=very satisfying and) 5 (=very unsatisfying)[6]. To a certain extent, these firm valuations can be interpreted as a form of vertical firm differentiation. Furthermore, the data comprise detailed information on about 556311 customer clicks requesting the referral to retail shops during this week.

For obvious reasons it is not possible to verify the above presented two-stage decision strategy with our dataset in a direct way - a combination of mind protocols together with the detailed clickstream of consumers would be necessary to do that. However, our dataset is perfectly suitable to pursue an indirect approach: By estimating an indirect hedonic price function we will show that the price is the dominant variable in the first stage of the decision process; but the closer we are moving towards the actual purchase decision, the more important are other firm characteristics (e. g. firms' valuation, country of origin, ...) relatively to the price. In particular, when we identify clicks with a higher purchase probability[7] we see that the importance of variables other than the price increases considerably. We will show that this is not an artefact of the way Geizhals.at presents their data on the website

[5]For the time span April 23, 2006 till June 25, 2007 we have in total 2.917 sellers, 279.973 products, 106.289.817 price offers, 49.594.757 clicks, 78.369 retailer valuations, 185.310 product valuations in our database.

[6]Customers who want to evaluate a shop have to register at the Geizhals website which enhances the reliability of the evaluations. Besides the identification of the shopper which is deterring rude behavior, Geizhals.at has a firm policy to check these evaluations: strategic evaluations coming from suspect IPs or from competitors are removed.

[7]For that purpose we use the known 'Last-Click-Through' concept and the newly suggested 'Most-Frequent-Click' idea discussed below.

but rather a heuristic which is used irrespective of the manner information is presented.

We use the following indirect hedonic price function:

$$\#clicks_{ij} = f(relprice_{ij}, valuation_j, ...) \tag{1}$$

In this equation $\#clicks_{ij}$ counts the consumers' referral requests on the Geizhals.at website to retailer j for product i. The variable $relprice_{ij}$ measures the price of product i of retailer j divided by the average price of product i across all firms offering this product (hence $relprice_{ij} = \frac{p_{i}j}{\sum_{j=1}^{N} p_{ij}/N})$[8]. Customers' average firm valuations are depicted with the variable $valuation_j$. Other control variables are included: $Shipping\ cost_j$ for retailers were calculated from the information given at Geizhals.at. Since this variable was not available (or was not unambiguously constructible) for all retailers, we interpolated with the average shipping cost and included additionally a *missing shipping cost* dummy. $Germany_j$ is equal to 1 if the online shop is located in Germany - as opposed to Austria, $avail_j$ is equal to 1 if the product is deliverable at short notice, $pickup_j$ is equal to 1 if the retailer has a pick up store as well. Controlling for the general type of the e-tailer (discounter versus high-priced online-shop) the $pricelevel_j$ indicates the average of $relprice_{ij}$ for firm j for all offered products divided by the average over all firms and products. As we observe relative prices and demand for different products, we include also product fixed effects to control for different demand conditions across products. Table 1 shows descriptive statistics for these variables.

4 Empirical Results

The number of clicks to a retailer is highly skewed. Around 91 percent of the products across the retailers are never selected from the customers due to the large number of offers available to them, the mean offer attracts 0.62 clicks during this week. As our dependent variable represents typical non negative count data, we are using fixed effects negative binomial panel estimations[9].

Table 2 includes our main results. We show marginal effects of the relative price, the firm valuation by the customers and some control variables. The first Column includes all product offers, whereas in Columns (2) to (4) we increasingly focus our analysis on top-listed firms, i.e. firms with the lowest

[8]Since a retailer can change the prices up to 10 times a day we are using the retailers' average price over the observation period.

[9]In all our estimations the likelihood ratio test for overdispersion rejects the poisson model.

prices among the shown list[10]. Due to the large number of observations all variables are significantly estimated and almost all have the expected sign. If we concentrate first on all product offers we see that both the relative price as well as the firm valuation (our service quality indicator) are important to explain demand for these products. Increasing the relative price by 10% would decrease demand by 0.13 clicks, which is considerable given a mean of 0.62 clicks per period. Likewise, increasing the numerical value of firm valuation (which is coded as a decrease of firm quality) by one standard deviation (0.47) will decrease demand by 0.013 clicks.

Columns (2) to (4) restricts our sample step by step: if we look only among the 20 cheapest, the ten cheapest or the five cheapest shops, the coefficient of the relative price increases steadily – from -2.2 to -3.8. This is not surprising, because the offers with the highest prices - getting no clicks at all - are eliminated. More remarkable is the development of the firm valuation coefficient. It increases by about tenfold and reaches the value -0.24. Among the five top-listed shops an increase of the firms' quality evaluation by half a grade will increase the number of clicks by 0.12. To show this realignment of coefficients more clearly, we calculate the "relative importance of price over service": while the marginal effect of relative price relative to firm valuation is 50.1 in Column (1) with all firms this relation falls to only 15.5 in Column (4) using only the top five firms. While this comparison is somewhat arbitrary as we compare different measurement units, the development across Columns is instructive. Apparently these results support our suggested decision strategy, in which consumers select a short list of potential retailers according to the price in the first step and consider carefully other variables on the second step through which the relativ importance of firm quality increases if we restrict our dataset to the products with a higher purchase probability.

Results from other product of firm specific characteristics corroborate this picture. We find that other quality or service components of an online shop have important effects on demand as well: these effects are in general several times more important than price if we look at the sub-sample of the cheapest shops. If the shop is located in Germany demand is considerably lower, presumably because customers fear warranty or delivery problems across borders. If the product in a shop is immediately available or if there is an additional pick up possibility, i.e. the online shop has also a brick-and-mortar store aside, these features are increasing the number of clicks: again, the importance of these effects increases with the focus to the cheapest price

[10] As Geizhals.at is ordering the offer increasing in prices at some places, the price search engine is anticipating this natural way to think about the attention and cognitive reasoning of potential shoppers.

offers. The number of firm valuations has a positive effect on demand, because customers might trust the reliability of the shop itself and also the valuations of the shop to a larger extent.

As a control for the general type of the e-tailer we included the general price level in the shop, which has a negative effect on buying a particular product, given the relative price of this particular product. This not surprising result indicates consumers' preferences for discounters rather than high priced shops. This might reflect the psychological effect that on average consumers believe more strongly in a good bargain if they buy in a web-shop well-known for their cheap prices than in a high priced shop even if they ask for the same price. The effect of shipping costs is inconclusive: If all offers are included, relative shipping costs have a small positive effect on demand; this coefficient turns duly negative once we concentrate upon the selection among the 20 or even five cheapest firms. It seems that customers only start looking at shipping costs once they consider seriously about buying from the shop.[11]

As a reverse check we look in Table 3 at firms which are bottom listed, i.e. whose prices are supposed to be above the unknown reservation price of the shoppers. If shoppers follow a two stage strategy - fixing the reservation price first and making a more elaborate evaluation among price and service quality later on - they should not care about firm evaluation in the case of offers above this reservation price. This is in fact, what we find: if we restrict firms to those ranked fortieth or above firm valuation turns minuscule and insignificant.

Another way to check the two stage procedure is to restrict the attention to products such that a heuristic is not necessary. If the number of shops is large, even the most meticulous shopper has to make shortcuts: she cannot check all details of the offers, a reservation price strategy might be a necessary first step in the decision process. However, if the number of shops is manageable, a full-fledged deliberation between price and service quality might be reasonable and possible. In Table 4 we restrict our analysis to products which are offered by a maximum of 20, 10 or only five firms. Column (1) shows again our baseline model from Table 2 using all products offered.

[11] There is no unambiguous expectation with respect to the dummy variable for the missing shipping cost. The coefficient should be negative if consumers will not be informed about this important variable in e-commerce business. It should be positive if the missing shipping cost are an indicator for special cheap rates. Shipping cost is the only variable which we have to parse from a text field. Obviously the parsing procedure requires some formal structure of the text field which will be not usable if the web-shop deviates from the default guideline because of special rates. Therefore we interpret the positive sign as an indicator for especially cheap shipping cost.

Looking again horizontally across Columns we see that the impact of firm valuation increases dramatically as we restrict the market size further and further: In markets with only 20, 10 or five firms, the coefficient for firm valuation increases from -0.0257 to -0.121, -0.226 up to -0.399. Comparing again this influence of relative price with the impact of service quality, we find the relative importance drop from 50.1 to 8.6 in the case of five firms. This is clear evidence for a different strategy in these markets: if many firms offer the good, the trade-off between price and service quality is less strong; the less firms there are, the more important quality dimensions get. The impact of other characteristics, like being a foreign (German) firm, having the product immediately available, having pickup possibilities and the number of evaluations for the firm, corroborates the conclusions: All these characteristics are much more important in small markets

A completely different view of the data offers additional support for our two stage decision process for the choice of the shortlist and the actual purchase. Looking at clicks to the website of an online shop - as we have done before - is the ideal way to define the shortlist of shops the customer is interested in. This should be contrasted with non observable actual purchases. Unfortunately, the actual act of purchasing a product is unknown. In the literature, the concept of 'Last-Click-Through' is often used as a proxy for the purchasing decision (e.g. Smith and Brynjolfsson (2001) or Bai (2004)). If a customer is searching for a product, she might meander around different web sites, comparing characteristics of the shops, but she will settle finally for the preferred shop and buy there online. The last click to a shop selling the product is usually identified as the click with a higher purchase probability.

In practice, it is not so simple to determine the 'Last-Click-Through' because buyers can shop for a specific product several times in a particular time interval. Analyzing the click behavior of a customer over time we have to define a 'searching period' which is finished with an actual purchase. If the costumer searches for several days, say, and then interrupts the search for a month or so, we might interpret the behavior as two distinct search and buying periods. Two approaches can be chosen for identification of such a search period. By hierarchical clustering which sequentially adds the clicks with respect to their minimal temporal distance we get a dendrogramm in which the fixing of a hierarchical level results in a certain amount of search intervals. Choosing a low level results in many search spells, choosing a high level gives us fewer intervals. Since the definition of the hierarchical level is arbitrary we decided to find the different search intervals with the Grubbs' Test for Outlier Detection. By choosing a significance level of 95 % those especially long time-differences can be found out which distinguish different search intervals[12]. Since by definition a search requires the comparison of several alternatives even a search period of one hour would have outliers we have to introduce additionally some minimal requirements - in one version a maximal time span has to be one week with at least 3 clicks, in a second version a maximal time span has to be one month with a least 5 clicks.

Another possibility to define the most probable purchase decision is the 'Most Frequent Click' (MFC) approach. If a favorite has emerged during the consumers' product search it will be necessary to compare repeatedly this favorite web-shop with the offers from other e-tailers. Hence, the last referral request to the most frequently clicked web-shop can be identified as the click with a high purchase possibility.

To complicate matters even more, customers might not only search for one specific product, they might look at substitutes during their search as well. The hierarchical mapping of the products into subsubcategories, subcategories and categories in the Geizhals.at data allows to cope with this issue since this classification scheme just describes the degree of substitutional relationship between the products (products in subsubcategory are close substitutes, products in categories reflect a looser substitutional relationship between products). Hence, the consumers' different search spells can be analyzed at the level of products, subsubcategories, and subcategories (categories seem to be a too general classification)[13].

Given the possibilities we come up with six different measures for the identification of actual purchase clicks indicating the choice between LCT and MFC, the length of the presumed search period[14], and the substitutional relationship of search products: 'LCT prod-week', 'LCT subsubc-week', 'LCT subc-week', 'MFC prod-week', 'MFC subsubc-week', 'MFC subc-week'.

In Table 5 we report estimates using these six definitions for counting the number of purchase clicks. For comparison reasons we also show our benchmark results from Table 2 using all clicks (Column 1). Panel A shows different variants of the Last-Click-Through and Panel B Most-Frequent-Clicks. Comparing all three LCT-variants with the benchmark results we see that in all cases the impact of firm valuation increases considerably up to twofold. For the Most-Frequent-Clicks we do not observe the increasing

[12]It can be shown that for each level in the hierarchical clustering a certain significance level for the Grubbs' Test for Outlier Detection can be found which results in identical search spells

[13]In total 539 subsubcategories and 53 subcategories are given. As an example the category 'Video/Foto/TV' contains the subcategory 'TV-Sets' and the subsubcategory 'LCD TV sets with 30-39 inches'.

[14]We do not report the results for the search period of one month. The results do not change our results

importance. Although the impact of relative price increases too, the relative importance of relative price versus firm valuation is substantially reduced, in particular if the Last-Click-Through is used as a proxy for an actual purchase. This result is not surprising if one has in mind that there should theoretically be a higher correlation between the MFC concept and the initially used number of clicks: If a retailer has a lot of clicks on his website it is more likely due to the fact that the retailer has a lot of repeated clicks by interested customers. This correlation should be smaller for the LCT approach. Hence, the LCT concept varies from the initial number of clicks more systematically than the MFC method.

Finally, when deciding whether to identify an actual purchase we restrict our data to those shops, where the potential buyer has already checked the firm-specific web-site. This corresponds to the idea, that making an actual purchase decision requires that the shop was shortlisted before (censored purchase clicks). Table 6 shows these results; again using the Last-Click-Through and the Most-Frequent-Click concept. The results are very similar, the additional importance of firm valuation is even higher as compared to Table 5. These results corroborate our previous findings: shoppers use different strategies when deciding about a shortlist of shops or when making a final buying decision.

The results with respect to the subsubcategorial and the subcategorial level are especially important in the context of our analysis. The price search engine Geizhals.at offers on its web-platform on several positions the possibility to rank product offers according to their price. One might have in mind that our empirical results are driven by the way how the data are presented by the price search engine. We want to indicate that this objection is not true for the subsub- and the subcategorial level. If someone is interested not in a specific product but rather in a specific type of product (eg. LCD TV sets with 30-39 inches) there is no way offered on geizhals.at to rank the different substitutes according to the price. If we can show that the relative importance of other characteristics of the good over the relative price increases in the final purchase decision for one of the substitute, this would again confirm the validity of our suggested heuristic. Table 7 is presenting exactly this result. In order to control for the substitutional relationship between the products on the subsubcategorial level we have collapsed the dataset by aggregating Last-Click-Through variable for all products and by averaging the right hand side variables on the subsubcategorial level for each firm. In the resulting dataset the firm's success measured with the sum of purchase clicks can be regressed again on firm specific variables (eg. average price of substitutes in the subsubcategory, firm evaluation, ...). Although no price ranking is available on the website for the substitutes within a subcategory we find our two stage decision strategy strongly confirmed. Whereas Column(1) includes the offers of all products Column (2) to (4) reduces to the 120 cheapest, the 100 cheapest and the 80 cheapest shops. Again the relative importance of the price variable is decreasing in support of a much stronger influence of non price related variables if we focus on the set of firms with a higher purchase probability. Consumers construct a short-list according to the price on the first stage, in the the second stage other variables again have a much stronger influence on the actual decision process.

5 Conclusions

This paper differs from previous work by (i) suggesting a new heuristic for consumer decisions where shoppers have the choice between several different dimensions of a product. (ii) For the empirical validation of the heuristic a new comprehensive data set from an Austrian price search engine is used. Compared to previous research this dataset allows a better control for substitutional relationships between the different products. (iii) Moreover, an alternative concept to identify the referral requests to web-shops in price search engines with higher purchase probabilities is suggested.

Our two stage decisions model proposes that consumers fix a reservation price and winnow all product offers with higher prices on the first stage in terms of a lexicographic. Within the remaining shortlist of product offers consumers will carry out a comprehensive consideration of the different choice alternatives. We found convincing evidence for this heuristic proceeding in data from an Austrian price search engine www.geizhals.at: Although we could not test the two stage decision procedure directly we can show with indirect hedonic price functions that the price variable is the dominant variable in the first stage of the decision process - however, the more we restrict our dataset to the set of product offers with a higher purchase probability the more important become other variables like the firms service valuation by customers, the products availability, the pick up facility of a web-shop etc. Identification of referral requests with higher purchase probability is being done by restricting our data to different subsamples as well as the usage of the so-called Last Click Through concept and newly suggested Most Frequent Click Approach, where those firm are selected as seller which are visited the most frequent during a search spell.

6 Appendix

Table 1: Descriptives

	Min	Max	Median	Mean	Std. Dev.	Description
#Clicks	0	3434	0	0.62	6.54	Number of clicks (referral requests) from consumers for firms' product offers
Rel. price	0.0056	5	0.985	1	0.18	Firm's product price relative to the market's mean product price
Firm valuation	1.05	3.55	1.68	1.74	0.47	Customers Valuations of the e-tailer (1 best and 5 worst)
Germany	0	1	1	0.69		Dummy: 1 if firms' country of origin is Germany, 0 if country of origin origin is Austria
Avail	0	1	0	0.29		Dummy: 1 if offered product is immediately available
Pick up	0	1	0	0.24		Dummy: 1 if e-tailer offers pick-up possibility
Pricelevel	0.24	1.47	0.992	1	0.068	General price level of the firm relativ to the average
#Valuations	5	1076	48	114.6	162.53	Number of customers valuations per firm
Rel. Shipping Cost	0.047	3.94	1.03	1	0.386	Firm's shipping cost (cash on delivery) relativ to the average (imputation with mean if not available)
Miss. ship. cost	0	1	0	0.11		Dummy: 1 if shipping cost are not available
LCT prod-week	0	158	0	0.045	0.45	Identification of Purchase Click with 'Last-Click Through' concept on product level
LCT subsubc-week	0	70	0	0.025	0.26	Identification of Purchase Click with 'Last-Click-Through' concept on subsubcategorial level
LCT subc-week	0	38	0	0.0087	0.128	Identification of Purchase Click with 'Last-Click-Through' concept on subcategorial level
MFC prod-week	0	290	0	0.045	0.58	Identification of Purchase Click with 'Most-Frequent-Click' concept on product level
MFC subsubc-week	0	175	0	0.026	0.36	Identification of Purchase Click with 'Most-Frequent-Click' concept on subsubcategorial level
MFC subc-week	0	99	0	0.0097	0.19	Identification of Purchase Click with 'Most-Frequent-Click' concept on subcategorial level

Table 2: Demand for Top-listed Firms

Data Sample	all product offers	top-listed 20 firms	top-listed 10 firms	top-listed 5 firms
Rel. Price	-1.287***	-2.194***	-2.841***	-3.820***
	-0.0093	-0.0221	-0.0370	-0.0706
Firm Valuation	-0.0257***	-0.0721***	-0.134***	-0.246***
	-0.0014	-0.00314	-0.00518	-0.00971
Rel. Shipping costs	0.0108***	-0.00545*	-0.0179***	-0.0244***
	-0.00138	-0.00303	-0.00489	-0.00881
Germany	-0.226***	-0.496***	-0.587***	-0.611***
	-0.00241	-0.00673	-0.0110	-0.0186
Availability	0.100***	0.162***	0.224***	0.322***
	-0.00163	-0.00345	-0.00565	-0.0105
Pick Up	0.0471***	0.0922***	0.149***	0.259***
	-0.00161	-0.00375	-0.00648	-0.0127
Pricelevel	-0.393***	-0.415***	-0.462***	-0.494***
	-0.0121	-0.0262	-0.0421	-0.0760
Miss. ship. cost	0.135***	0.208***	0.251***	0.318***
	-0.00248	-0.00497	-0.00771	-0.0138
#Valuations	0.00032***	0.00047***	0.00054***	0.00064***
	0.000	-0.00001	-0.00001	-0.00002
Observations	899959	457865	277144	153973
Products	39002	38345	37001	34840
χ^2	92729	46130	25255	11516
LL	-456040	-312447	-218610	-132023
rel. Importance of price over service	50.1	30.4	21.2	15.5

Method of Estimation: Negative Binomial with product fixed effects - marginal effects are shown. Standard errors are in parentheses. *, ** and *** indicate statistical significance at the 10-percent, 5-percent and 1-percent level. Constant is not shown in Table. Marginal effects for dummy variables represent discrete change from 0 to 1.

Table 3: Demand for Bottom-listed Firms

Data Sample	all product offers	offers with firm rank >20	offers with firm rank >40	offers with firm rank >60
Rel. Price	-1.287***	-0.321***	-0.102***	-0.0105
	-0.0093	-0.00868	-0.00898	-0.00969
Firm Valuation	-0.0257***	-0.0112***	-0.00468*	-0.00025
	-0.0014	-0.00198	-0.00284	-0.00451
Rel. Shipping costs	0.0108***	0.0142***	0.0175***	0.0182***
	-0.00138	-0.00187	-0.00261	-0.00412
Germany	-0.226***	-0.215***	-0.165***	-0.133***
	-0.00241	-0.00354	-0.0044	-0.0061
Availability	0.100***	0.0826***	0.0854***	0.0831***
	-0.00163	-0.00231	-0.00352	-0.00568
Pick Up	0.0471***	0.0311***	0.0360***	0.0454***
	-0.00161	-0.00190	-0.00274	-0.00451
Pricelevel	-0.393***	-0.188***	-0.207***	-0.195***
	-0.0121	-0.0164	-0.0226	-0.0342
Miss. ship. cost	0.135***	0.0909***	0.0867***	0.0810***
	-0.00248	-0.00369	-0.00562	-0.00906
#Valuations	0.00032***	0.00026***	0.00024***	0.00024***
	0.000	0.000	-0.00001	-0.00001
Observations	899959	297310	121362	46844
Products	39002	7680	3270	1373
χ^2	92729	32017	11833	3767
LL	-456040	-103052	-38310	-13858

Method of Estimation: Negative Binomial with product fixed effects - marginal effects are shown. Standard errors are in parentheses. *, ** and *** indicate statistical significance at the 10-percent, 5-percent and 1-percent level. Constant is not shown in Table. Marginal effects for dummy variables represent discrete change from 0 to 1.

Table 4: Demand for Products with Few Suppliers

Data Sample	all product offers	products with max. 20 firms	products with max. 10 firms	products with max. 5 firms
Rel. Price	-1.287***	-2.435***	-2.881***	-3.461***
	-0.0093	-0.0384	-0.0690	-0.139
Firm Valuation	-0.0257***	-0.121***	-0.226***	-0.399***
	-0.0014	-0.0074	-0.0156	-0.0347
Rel. Shipping costs	0.0108***	0.0201***	0.0107	-0.00002
	-0.00138	-0.00765	-0.0159	-0.0345
Germany	-0.226***	-0.490***	-0.552***	-0.463***
	-0.00241	-0.0112	-0.0201	-0.0350
Availability	0.100***	0.237***	0.321***	0.438***
	-0.00163	-0.00837	-0.0169	-0.0375
Pick Up	0.0471***	0.0885***	0.142***	0.279***
	-0.00161	-0.00812	-0.0170	-0.0390
Pricelevel	-0.393***	-0.949***	-0.821***	-0.227
	-0.0121	-0.0550	-0.106	-0.221
Miss. ship. cost	0.135***	0.213***	0.280***	0.376***
	-0.00248	-0.00912	-0.0174	-0.0377
#Valuations	0.00032***	0.00057***	0.00071***	0.00088***
	0.000	-0.00002	-0.00003	-0.00008
Observations	899959	186985	89484	37173
Products	39002	24801	18235	11480
χ^2	92729	17081	7150	2324
LL	-456040	-125011	-62620	-25830
rel. Importance of price over service	50.1	20.1	12.7	8.6

Method of Estimation: Negative Binomial with product fixed effects - marginal effects are shown. Standard errors are in parentheses. *, ** and *** indicate statistical significance at the 10-percent, 5-percent and 1-percent level. Constant is not shown in Table. Marginal effects for dummy variables represent discrete change from 0 to 1.

Table 5: Purchase Clicks

Type LCT/MFC	all clicks	prod-week	subsubc-week	subc-week
		PANEL A: Last Click Through		
Relprice	-1.287***	-1.485***	-1.641***	-1.559***
	-0.0093	-0.0366	-0.0578	-0.107
Firm valuation	-0.0257***	-0.0512***	-0.0453***	-0.0476***
	-0.0014	-0.00424	-0.00512	-0.00728
	...	...	...	...
Observations	899959	418657	320042	164472
Products	39002	12158	9300	4334
rel. Importance	50.1	29.0	36.2	32.8
		PANEL B: Most Frequent Clicks		
Relprice	-1.287***	-1.361***	-1.191***	-1.023***
	-0.0093	-0.0298	-0.0364	-0.05638
Firm valuation	-0.0257***	-0.0289***	-0.0249***	-0.0170***
	-0.0014	-0.00258	-0.00292	-0.00361
	...	...	...	...
Observations	899959	418081	323825	168793
Products	39002	12139	9458	4560
rel. Importance	50.1	47.1	47.8	60.2

Method of Estimation: Negative Binomial with product fixed effects - marginal effects are shown. Standard errors are in parentheses. *, ** and *** indicate statistical significance at the 10-percent, 5-percent and 1-percent level. Additional variables as in Table 2.

Table 6: Censored Purchase Clicks

TYPE LCT/MFC	ALL CLICKS	PROD-WEEK	SUBSUBC-WEEK	SUBC-WEEK
		PANEL 1: Last Click Through - Censored		
Relprice	-1.287***	-1.155***	-1.565***	-1.493***
	-0.0093	-0.0664	-0.0918	-0.131
Firm valuation	-0.0257***	-0.128***	-0.0830***	-0.0929***
	-0.0014	-0.0153	-0.0167	-0.0193
	...	...	...	...
Observations	899959	96749	78358	46119
Products	39002	11788	8685	4116
rel. Importance	50.1	9.0	18.9	16.1
		PANEL 2: Most Frequent Clicks - Censored		
Relprice	-1.287***	-2.207***	-1.773***	-1.405***
	-0.0093	-0.0711	-0.0742	-0.0910
Firm valuation	-0.0257***	-0.0714***	-0.0471***	-0.0240**
	-0.0014	-0.0105	-0.0107	-0.011
	...	...	...	...
Observations	899959	96815	79650	47986
Products	39002	11779	8847	4340
rel. Importance	50.1	30.9	37.6	58.5

Method of Estimation: Negative Binomial with product fixed effects - marginal effects are shown. Standard errors are in parentheses. *, ** and *** indicate statistical significance at the 10-percent, 5-percent and 1-percent level. Additional variables as in Table 2.

Table 7: Demand of top-listed firms for close substitutes

Data Sample	all offers within sub-subcategory	top-listed 120 firms within subsubcategory	top-listed 100 firms within subsubcategory	top-listed 80 firms subsubcategory
Rel. Price	-0.581***	-0.389***	-0.292***	-0.142***
	-0.0293	-0.0430	-0.0487	-0.0548
Firm Valuation	0.0201***	0.0231***	0.0206**	0.01704
	-0.00688	-0.00871	-0.00969	-0.0109
Rel. Ship. costs	0.0177***	0.0182**	0.0183**	0.0214**
	-0.00654	-0.00817	-0.0090	-0.00996
Germany	-0.180***	-0.196***	-0.211***	-0.224***
	-0.00972	-0.0134	-0.0155	-0.0186
Availability	0.0370***	0.0432***	0.0421***	0.0365***
	-0.00775	-0.0098	-0.0109	-0.0122
Pick Up	0.0294***	0.0377***	0.0415***	0.0452***
	-0.00716	-0.00932	-0.0104	-0.0120
Pricelevel	-0.237***	-0.356***	-0.414***	-0.574***
	-0.0570	-0.0790	-0.0894	-0.103
#Valuations	0.00042***	0.00046***	0.00048***	0.00048***
	-0.00001	-0.00002	-0.00002	-0.00002
Observations	42648	33866	29981	25054
Subsubcategories	373	373	373	370
χ^2	2476	1546	1274	997.9
LL	-26288	-20971	-18891	-16155
rel. Importance of price over service	28.9	16.8	14.2	8.33

Method of Estimation: Negative Binomial with product fixed effects - marginal effects are shown. Dependent Variable: Firms' sum of Last-Click-Throughs for all products within subsubcategories. for Standard errors are in parentheses. *, ** and *** indicate statistical significance at the 10-percent, 5-percent and 1-percent level. Constant is not shown in Table. Marginal effects for dummy variables represent discrete change from 0 to 1.

References

Bai, Xue, "Why Pay More? Why Charge Less?," 2004. Anindya Ghose's Collection, http://hdl.handle.net/2451/14974, Working Paper.

Betancourt, Roger R. and David Gautschi, "Two essential characteristics of retail markets and their economic consequences," *Journal of Economic Behavior & Organization*, August 1993, *21* (3), 277–294.

Gaskin, Steven, Thoedoros Evgeniou, Daniel Bailiff, and John Hauser, "Two-Stage Models: Identifying Non-compensatory Heuristics for the Consideration Set then Adaptive Polyhedral Methods within the Consideration Set," 2008. MIT Sloan School of Management, mimeo.

Gigerenzer, Gerd and Daniel G. Goldstein, "Reasoning the fast and frugal way: Models of bounded rationality," *Psychological review*, 1996, *103* (4), 650–669.

_ , Peter M. Todd, and the ABC Research Group, *Simple Heuristics that Make Us Smart*, Oxford University Press, 1999.

Kfir, Eliaz and Efe A. Ok, "Indifference or Indecisiveness? Choice Theoretic Foundations of Incomplete Preferences," *Games and Economic Behavior*, 2006, *56*, 61–86.

Kohli, Rajeev and Kamel Jedidi, "Representation and Inference of Lexicographic Preference Models and their Variants," *Marketing Science*, 2007, *26* (3), 380–399.

Mandler, Michael, Paola Manzini, and Marco Mariotti, "A Million Answers to Twenty Questions: Choosing by Checklist," 2008. IZA Discussion Paper No. 3377.

Manzini, Paola and Marco Mariotti, "Sequential Rationalizable Choice," *American Economic Review*, 2007, *97*, 1824–39.

Pan, Xing, Brian T. Ratchford, and Venkatesh Shankar, "Can price dispersion in online markets be explained by differences in e-tailerservice quality," *Journal of the Academy of Marketing Science*, 2002, *30* (4), 433–445.

Payne, John W., "Task Complexity and Contingent Processing in Decision Making: an Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*, 1976, *16*, 366–387.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 1974, *82* (1), 34–55.

Simon, Herbert A., *Models of man*, New York: Wiley, 1957.

Smith, Michael and Erik Brynjolfsson, "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, April 2000, pp. 563–585.

Smith, Michael D. and Erik Brynjolfsson, "Consumer Decision-making at an Internet Shopbot: Brand Still Matters," *Journal of Industrial Economics*, 2001, *49* (4), 541–558.

Tversky, Amos, "Elimination by Aspects: a Theory of Choice," *Psychological review*, 1972, *79* (4), 281–299.

Varian, Hal R., "Revealed Preference," 2006. University of California, Berkeley, mimeo.

Yee, Michael, Ely Dahan, John R Hauser, and James Orlin, "Greedoid-based Non-compensatory Inferences," *Journal of Marketing Science*, 2007, *26* (4), 532–549.