Essays on Modeling and Measurement of Consumers' Decision Strategies

By

Daria Dzyabura

S.B., Massachusetts Institute of Technology (2007)

Submitted to the Sloan School of Management

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Signature of Author...
Sloan School of Management
May 4, 2012

Certified by...
John R. Hauser
Kirin Professor of Marketing
Thesis Supervisor

Accepted by...
Ezra Zuckerman Sivan
Nanyang Technological University Professor
Chair, MIT Sloan PhD Program

Essays on Modeling and Measurement of Consumers' Decision Strategies

By

Daria Dzyabura

Submitted to the Sloan School of Management
on May 4, 2012 in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Management

ABSTRACT

This thesis consists of three related essays which explore new approaches to modeling and measurement of consumer decision strategies. The focus is on decision strategies that deviate from von Neumann-Morgenstern utility theory. Essays 1 and 2 explore decision rules that consumers use to form their consideration sets. Essay 1 proposes disjunctions-of-conjunctions (DOC) decision rules that generalize several well-studied decision models. Two methods are proposed for estimating the model. Consumers' consideration sets for global positioning systems are observed for both calibration and validation data. For the validation data, the cognitively simple DOC-based methods predict better than the ten benchmark methods on an information theoretic measure and on hit rates. The results are robust with respect to format by which consideration is measured, sample, and presentation of profiles. Essay 2 develops and tests an active-machine-learning method to select questions adaptively when consumers use heuristic decision rules. The method tailors priors to each consumer based on a "configurator." Subsequent questions maximize information about the decision heuristics (minimize expected posterior entropy). To update posteriors after each question the posterior is approximated with a variational distribution and uses belief-propagation. The method runs sufficiently fast to select new queries in under a second and provides significantly and substantially more information per question than existing methods based on random, market-based, or orthogonal questions. The algorithm is tested empirically in a web-based survey conducted by an American automotive manufacturer to study vehicle consideration. Adaptive questions outperform market-based questions when estimating heuristic decision rules. Heuristics decision rules predict validation decisions better than compensatory rules. Essay 3 proposes a model of product search when preferences are constructed during the process of search: consumers learn what they like and dislike as they examine products. Product recommendations, whether made by sales people or online recommendation systems, bring products to the consumer's attention <u>and</u> impact his/her preferences. Changing preferences changes the products the consumer will choose to search; at the same time, the products the consumer chooses to search will determine the future shifts in preferences. Accounting for this two-way relationship between products and preferences is critical in optimizing recommendations.

Thesis Supervisor: John R. Hauser
Title: Kirin Professor of Marketing

Acknowledgements

My first thanks go to my advisor, John Hauser, for his support, motivation, wisdom, and kindness throughout my PhD experience. He taught me what research is all about, how to balance theory and practice, approach real world problems, and helped me understand the world a little bit better. John gave me the freedom to explore on my own, while providing guidance and support. Our meetings always inspired me to return to my research with renewed enthusiasm, and I cannot thank him enough for all the lessons I have learned from him. It has been an honor to be his student, and hope that one day I might become as good a mentor for my students.

The other members of my committee, Natalie Mizik and Juanjuan Zhang, have been unfailingly generous with their time, and their mentorship has been invaluable to my research and career.

I had an amazing five years being part of the marketing group at MIT, and I will really miss this place. To the rest of the Marketing faculty, Birger Wernerfelt, Catherine Tucker, Drazen Prelec, Duncan Simester, Glen Urban, John Little, Josh Ackerman, Michael Braun, Renee Gosline, and Sharmila Chatterjee, thank you for all your insights and support. I would like to thank all my fellow PhD students, for being such a great group of intelligent and fun people.

I don't know where I would be without my friends and family. My brother, Maksim, has been my role model ever since I can remember. I can say that he is, hands down, the most incredible person I know: he is brilliant, creative, and extremely hard working, while at the same time compassionate and kind. I am so proud of everything he has accomplished and of being his little sister. My mother is a truly amazing woman who has overcome so many difficulties to help me get here, and who has been unfailingly supportive my entire life. I want to thank all my friends who have shared with me so many happy and difficult moments, especially Danielle Braun for helping me stay sane for the past 9 years.

Finally, huge thank you to my husband, Vasya, and my daughter, Nika, for making me so happy!

Table of Contents

Chapter 1: Disjunctions of Conjunctions, Cognitive Simplicity and Consideration Sets

Abstract

We test methods, based on cognitively-simple decision rules, that predict which products consumers select for their consideration sets. Drawing on qualitative research we propose disjunctions-of-conjunctions (DOC) decision rules that generalize well-studied decision models such as disjunctive, conjunctive, lexicographic, and subset conjunctive rules. We propose two machine-learning methods to estimate cognitively-simple DOC rules. We observe consumers' consideration sets for global positioning systems for both calibration and validation data. We compare the proposed methods to both machine-learning and hierarchical-Bayes methods each based on five extant compensatory and non-compensatory rules. On validation data the cognitively-simple DOC-based methods predict better than the ten benchmark methods on an information theoretic measure and on hit rates; significantly so in all but one test. An additive machine-learning model comes close on hit rate. Our results are robust with respect to format by which consideration is measured (four formats tested), sample (German representative vs. US student), and presentation of profiles (pictures vs. text). We close by illustrating how DOC-based rules can affect managerial decisions.

Keywords: *Consideration sets, non-compensatory decisions, consumer heuristics, statistical learning, machine learning, revealed preference, conjoint analysis, cognitive complexity, cognitive simplicity, environmental regularity, lexicography, logical analysis of data, decision trees, combinatorial optimization.*

CONSIDERATION SETS AND DECISION RULES

Consideration decisions are managerially important. For example, General Motors has invested heavily in product design and quality such that in 2007 Buick tied Lexus for the top spot in J. D. Power's vehicle dependability ranking and in 2008 Buick was the top US brand in Consumer Reports. However, roughly half of US consumers (and 64% in California) will not even consider a Buick. Because the typical consumer considers less than 10 vehicles when shopping for a new vehicle, top managers at General Motors are interested in understanding how consumers decide which 10 of the 350+ make-model combinations to consider further. To direct strategies, they would like to model the features consumers use to screen products for further consideration. They would like a model that can forecast changes in consideration as a function of changes in product lines or changes in the features that are emphasized in marketing activities.

Two-stage, consider-then-choose decision rules are particularly relevant in the automobile market, but modeling and forecasting such decision rules is of general interest. When consumers face a large number of alternative products, as is increasingly common in today's retail and web-based shopping environments, they typically screen the full set of products down to a smaller, more-manageable consideration set which they evaluate further (e.g., Bronnenberg and Vanhonacker 1996; DeSarbo et al., 1996; Hauser and Wernerfelt 1990; Jedidi, Kohli and DeSarbo, 1996; Mehta, Rajiv, and Srinivasan, 2003; Montgomery and Svenson 1976; Payne 1976; Roberts and Lattin, 1991; Shocker et al., 1991; Wu and Rangaswamy 2003). Consideration sets for packaged goods are typically 3-4 products rather than the 30-40 products on the market (Hauser and Wernerfelt 1990; Urban and Hauser 2004). Forecasting consideration sets can explain roughly 80% of the explainable uncertainty in consumer decision making (assuming equally likely choice within the consideration set, Hauser 1978). In complex product categories research suggests that at least some consumers use non-compensatory decision processes when evaluating many products and/or products with many features (e.g., Payne, Bettman and Johnson 1988, 1993).

In this paper we explore machine-learning algorithms based on non-compensatory decision rules that model decisions by consumers in the consideration stage of a consider-then-choose process. We measure consideration directly for a moderately-complex product, handheld Global Positioning Systems (GPSs) and, assuming a general form of non-compensatory decision rules, we attempt to model the non-compensatory patterns that best predict consumers' consideration decisions. The general form, disjunctions of conjunctions (DOC), is motivated by qualitative data and nests several previously-studied rules. We argue further that modeling and controlling for cognitive simplicity enhances predictive ability.

We compare the DOC-based machine-learning algorithms to two sets of benchmarks. The first set includes alternative machine-learning algorithms that assume either compensatory decision rules or previously published non-compensatory decision rules. The second set includes hierarchical Bayes (HB) methods for the same compensatory and non-compensatory rules. In this product category, the proposed DOC-based machine-learning methods predict consideration sets better than the benchmarks using two metrics – hit rates and an information-theoretic measure. In almost all comparisons, predictions are significantly better statistically.

We demonstrate that our basic conclusions are robust with respect to format by which consideration is measured (four formats tested), sample (German representative vs. US student), and presentation of profiles (pictures vs. text). We close by illustrating how the modeled non-compensatory patterns affect managerial decisions differently than additive decision rules.

NOTATION AND ESTABLISHED DECISION RULES

We focus on data in which respondents are asked to indicate which of several product profiles (32 in our experiments) they would consider. Respondents are free to select any size consideration set. In some formats respondents classify each profile as considered or not considered; in other formats they do not need to evaluate every profile.

We explore situations in which features are described by finitely many levels. Let j index the profiles, ℓ index the levels, f index the features (sometimes called "attributes" in the literature), and h index the respondents. Let J, L, F, and H be the corresponding numbers of profiles, levels, features, and respondents. For ease of exposition only, we do not write J, L, and F as dependent (e.g., L_f). Our models and estimation can (and do) handle such dependency, but the notation is cumbersome. Let $x_{jf\ell} = 1$ if profile j has feature f at level ℓ. Otherwise $x_{jf\ell} = 0$. Let $\vec{x}_j$ be the binary vector (of length LF) describing profile j. Let $y_{hj} = 1$ if we observe that respondent h considers profile j. Otherwise, $y_{hj} = 0$. Let $\vec{y}_h$ be the binary vector describing respondent h's consideration decisions.

Non-compensatory Decision Rules

Commonly-studied non-compensatory rules include disjunctive, conjunctive, lexicographic, elimination-by-aspects, and subset conjunctive rules (e.g., Gilbride and Allenby 2004, 2006; Jedidi and Kohli 2005; Montgomery and Svenson 1976; Ordóñez, Benson and Beach 1999; Payne, Bettman, and Johnson 1988; Yee, et. al. 2007). Subset conjunctive rules generalize disjunctive and conjunctive rules (Jedidi and Kohli 2005). For consideration decisions, they also generalize lexicographic rules and deterministic elimination-by-aspects, because any implied ranking of products by lexicographic feature-level orders is indeterminate if we observe only the consideration decision (Hogarth and Karelaia 2005; Johnson, Meyer and Ghose 1989; Montgomery and Svenson 1976; Payne, Bettman, and Johnson 1988; Tversky 1972).

Disjunctive rules. In a disjunctive rule, a profile is considered if at least one of the features is at an "acceptable" (or satisfactory) level. Let $a_{hf\ell} = 1$ if level ℓ of feature f is acceptable to respondent h. Otherwise, $a_{hf\ell} = 0$. Let $\vec{a}_h$ be the binary vector of acceptabilities for respondent h. A disjunctive rule states that respondent h considers profile j if $\vec{x}_j'\vec{a}_h \geq 1$.

Conjunctive rules. In a conjunctive rule, a profile is considered if <u>all</u> of the features are at an acceptable level. (Conjunctive rules usually assume a larger set of acceptable levels than disjunctive rules, but this is not required.) Because the use in each rule is clear in context, we use the same notation: in a conjunctive rule, respondent *h* considers profile *j* if $\vec{x}_j'\vec{a}_h = F$.

Subset conjunctive rules. In a subset conjunctive rule, a profile is considered if at least *S* features are at an acceptable level. Using the same notation, respondent *h* considers profile *j* if $\vec{x}_j'\vec{a}_h \geq S$. Clearly, a disjunctive rule is a special case where $S = 1$ and, because $\vec{x}_j'\vec{a}_h$ can never exceed *F*, a conjunctive rule is a special case where $S = F$. We denote subset conjunctive rules by Subset(*S*). (Subset conjunctive rules are mathematically equivalent to "image-theory" rules in organizational behavior, e.g., Ordóñez, Benson and Beach 1999.)

Additive and* q*-Compensatory Decision Rules

Perhaps the most pervasively studied decision rules are additive rules. In an additive rule, consumers consider a profile if its "utility" is above some threshold, T_h, which accounts for search and processing costs. If $\vec{\beta}_h$ is the vector of partworths for respondent *h*, then *h* considers profile *j* if $\vec{x}_j'\vec{\beta}_h \geq T_h$. For estimation we model errors in the decisions.

Many researchers demonstrate that an additive partworth rule can mimic lexicographic, subset conjunctive, and conjunctive rules (e.g., Jedidi and Kohli 2005; Kohli and Jedidi 2007; Olshavsky and Acito 1980; Yee, et al. 2007). To explore whether a model might predict better if it is constrained to be compensatory, we follow Bröder (2000) and Yee, et al. (2007) who specify a *q*-compensatory model by constraining the additive model so that no feature's importance is more than *q* times as large as another feature's importance. (Hogarth and Karelaia (2005) and Martignon and Hoffrage (2002) use related constraints. A feature's importance is the difference between the maximum and minimum partworths for that feature.)

DISJUNCTIONS OF CONJUNCTIONS (DOC)

To study consideration-set decisions we began with a qualitative study that used in-depth interviewing for 38 automobile consumers who were asked to describe their consideration decisions for 100 real automobiles that were balanced to market data. All interviews were video-recorded and the videos were evaluated by independent judges who were blind to any hypotheses about consumers' decision rules (Hughes and Garrett 1990; Perreault and Leigh 1989). Most respondents made consideration decisions rapidly (89% averaged less than 5 seconds per profile) and most used non-compensatory decision rules (76%). Typically, consumers used conjunctive-like criteria defined on specific levels of features. However, some consumers would consider an automobile if it satisfied at least one of multiple conjunctive criteria (i.e., a disjunction of two or more conjunctions).

For example, the following respondent considers automobiles that satisfy <u>either</u> of two criteria. The first criterion is clearly conjunctive (good styling, good interior room, excellent mileage). The second criterion allows cars that are "hotrods." "Hotrods" usually have poor interior room and poor mileage.

> *[I would consider the Toyota Yaris because]the styling is pretty good, lot of interior room, mileage is supposed to be out of this world.*

> *I definitely [would] consider [the Infinity M-Sedan], though I would probably consider the G35 before the "M". I like the idea of a kind of a hotrod.*

Depth interviewing is, by necessity, based on a small sample. From the sample we could not determine whether multiple conjunctions were pervasive or were limited to a subset of consumers. However, qualitative interviewing in the handheld GPS category also identified some consumers who used multiple conjunctions. A respondent might be willing to consider a GPS with a B&W screen if the GPS was small and the screen was high resolution, but would require a color screen on a large GPS. Such rules can be written as logical patterns: (B&W screen $\wedge$ small size $\wedge$ high resolution) $\vee$ (color screen

$\wedge$ large size), where $\wedge$ is the logical "and" and $\vee$ is the logical "or." Patterns might also include negations ($\neg$), for example, a consumer might accept a B&W screen as long as the GPS is less than the highest price of \$399: (B&W screen $\wedge$ $\neg$ \$399).

Formal Definition of DOC Rules

To study this phenomenon further, we formalize these qualitative insights with a class of decision rules that generalizes previously-proposed rules. First, following Tversky (1972) we define an aspect as a binary descriptor such as "B&W screen." A profile either has or does not have an aspect. A pattern is a conjunction of aspects or their negations such as (B&W screen $\wedge$ $\neg$ \$399). We define the size, s, of a pattern as the number of aspects in the pattern. For example, (B&W screen $\wedge$ $\neg$ \$399) has size $s = 2$. If p indexes patterns, then we say that a profile j matches pattern p if profile j contains all aspects (or negations) in pattern p.

We study rules where a respondent considers a profile if the profile matches one or more target patterns. Because each pattern is a conjunction, these logical rules are disjunctions of conjunctions (DOC). DOC rules generalize disjunctive rules (disjunctions of patterns of size 1), conjunctive rules (patterns of size F), and subset conjunctive rules (patterns of size S).[1]

Let $w_{hp} = 1$ if pattern p is one of the patterns describing respondent h's decision rule and let $m_{jp} = 1$ if profile j matches pattern p. Otherwise, w_{hp} and m_{jp} are zero. Let $\vec{w}_h$ and $\vec{m}_j$ be the corresponding binary vectors with length equal to the number of allowable patterns in a DOC rule. A DOC rule implies that respondent h considers profile j if and only if $\vec{m}_j'\vec{w}_h \geq 1$.

[1] We demonstrate formally, in the Web Appendix, that (1) disjunctive rules, subset conjunctive rules of pattern length 1, and DOC rules of maximum pattern length 1 are equivalent, (2) conjunctive rules, subset conjunctive rules of pattern length F are equivalent and a subset of DOC rules, and (3) subset conjunctive rules of pattern length S can be written as DOC rules but there exist DOC rules of maximum pattern length S that cannot be written as subset conjunctive rules of pattern length S.

Cognitive Simplicity

DOC rules generalize previously proposed non-compensatory decision rules, but they might be too general. For example, any profile can be described by a pattern of size F. Thus, any consideration set of size n can be fit perfectly with a disjunction of n conjunctions of size F. Fortunately, experimental evidence suggests that consumers make consideration decisions with relatively simple rules that enable them to make good decisions while avoiding excess cognitive effort (e.g., Bettman, Luce and Payne 1998; Bröder 2000; Gigerenzer and Goldstein 1996; Gigerenzer and Todd 1999; Hogarth and Karelaia 2005; Payne, Johnson and Bettman 1988, 1993; Martignon and Hoffrage 2002; Simon 1955; Shugan 1980). This perspective of simple, efficient, search-and-evaluation rules is consistent with economic theories of consideration-set formation which posit that consumers balance search costs and the option value of utility maximization (Hauser and Wernerfelt 1990; Roberts and Lattin 1991). To capture this "cognitive simplicity" we define DOC(S) rules as the set of DOC rules with maximum pattern length S. In addition, we either limit the number of patterns, P, or penalize DOC rules that have large P.

MACHINE LEARNING APPROACHES TO IDENTIFY DOC PATTERNS

The basic data we observe, for a set of respondents and profiles, is whether or not a respondent considers a profile (y_{hj}). We seek to identify the patterns that predict best how respondent h evaluates profiles. Using a calibration sample we seek patterns such that profile j is observed considered if $\vec{m}_j'\vec{w}_h \geq 1$ and not considered if $\vec{m}_j'\vec{w}_h = 0$. (Recall $\vec{m}_j$ and $\vec{w}_h$ are binary.)

The number of allowable DOC(S) patterns grows rapidly with S. For example, with the 16 binary features in our empirical test, there would be 32 patterns for $S = 1$, 512 for $S = 2$, 4,992 for $S = 3$, and 34,112 for $S = 4$. There would be almost 20 million patterns of length $S = 10$. With only 32 binary observations (consider vs. not consider) there is serious concern about over-fitting because the vector, $\vec{w}_h$, which we seek to estimate, has length equal to this large number of allowable patterns.

Machine learning is particularly suited to this pattern-matching task. Qualitative interviews suggest that it was not unreasonable for patterns to be up to length $S = 4$, which requires we search over 34 thousand patterns to find those that best fit the data. While we might place priors on each pattern and use Bayesian methods, we have not yet been able to develop a Bayesian representation in which the posterior is robust with respect to exogenously-set priors for the large number of parameters. We leave exploration of Bayesian DOC models to future research.

Rather than producing posterior probabilities of pattern inclusion, we seek binary indicators of whether or not a pattern is in the best-fit solution. If the data are too noisy or the solution space is too large (even controlling for cognitive simplicity), predictions could over fit the data and predict poorly. To be sensitive to this concern we compare models using predictive tests in which respondents face an entirely new set of profiles and report consideration for those profiles.

Cognitive Simplicity and Complexity Control

Although we used cognitive simplicity to motivate small S and P, such constraints or penalties have an alternative interpretation within machine learning – complexity control (e.g., Cucker and Smale 2002; Evgeniou, Boussios and Zacharia 2005; Hastie, Tibshirani and Friedman 2003; Langley 1996; Vapnik 1998). Limiting the complexity of a model often minimizes in-sample over-fitting and enhances out-of-sample prediction. Both the behavioral explanation and the complexity-control motivation are consistent with our DOC(S) models – we cannot rule out either with the data in this paper.

Sample Shrinkage

To further distinguish among potential patterns we use data from the entire sample to help select patterns for respondent h. In an analogy to shrinkage, which enhances accuracy in hierarchical Bayesian models (e.g., Rossi and Allenby 2003), we favor those patterns that fit the largest subset of respondents. While shrinkage alone is sufficient motivation for use in our models, shrinkage is consistent with behavioral theories which suggest that simple rules have evolved because they work well in the general environment in which a sample of consumers often make decisions (e.g., Chase, Hertwig and Gigerenzer 1998). These researchers hypothesize that consumers continue to use similar (simple) rules when faced with new decisions.

We now summarize briefly two machine-learning methods. Detailed equations are contained in the Web Appendix.

Mathematical Programming (DOCMP)

Because we seek the binary vector, $\vec{w}_h$, that best matches patterns in the calibration data, we formulate an integer program such that w_{hp} must be either 0 or 1 for all p. For respondent h, we define false positives, $FP_h(\vec{w}_h)$, as the number of profiles predicted to be considered but observed as not considered and we define false negatives, $FN_h(\vec{w}_h)$, as the number of profiles predicted to be not considered but observed to be considered. In its most basic form, the integer program (DOCMP) would choose the $\vec{w}_h$ that minimizes the sum of false positives and false negatives for respondent h.

We enforce cognitive simplicity (complexity control) by limiting the search to patterns of length S or less and by penalizing pattern length, P. We include shrinkage with terms proportional to the sum of false positives and false negatives in the sample (sum over all respondents). Formally, our objective function is:

(1) $$\min_{\{\vec{w}_h\}} \left\{ FP_h(\vec{w}_h) + FN_h(\vec{w}_h) + \gamma_M \sum_{i=1}^{H} [FP_i(\vec{w}_h) + FN_i(\vec{w}_h)] + \gamma_c P \right\}$$

DOCMP is equivalent to a set-covering problem and, hence, is an NP-hard problem (Cormen, et. al. 2001). Fortunately, efficient greedy approximation algorithms have been developed and tested for this class of problems (Fiege 1998; Lund and Yannakakis 1994). Alternatively, DOCMP can be solved approximately with a linear-programming relaxation in which we first allow $\vec{w}_h$ to be continuous on [0, 1], then round up any positive w_{hj} that is above a threshold (Hastie, Tisbshirani, and Friedman 2003, and references therein). In our estimations, we use both the greedy and the relaxation methods, choosing the solution that provides the best value of the objective function (using calibration data only; no data from the validation profiles).

DOCMP requires three exogenous parameters: g_M tells us how much to penalize lack of sample-level fit, g_c tells us how much to penalize the number of patterns, and S that sets the maximum pattern length. One method to select these parameters, is leave-one-out-cross-validation (e.g., Cooil, Winer and Rados 1987; Efron and Tibshirani 1997; Evgeniou, Pontil and Toubia 2007, Hastie, Tibshirani, and Friedman 2003; Kearns and Ron 1999; Kohavi 1995; Shao 1993; Toubia, Evgeniou and Hauser 2007; Zhang 2003). Specifically, for potential values of the exogenous "tuning" parameters we leave out one profile from the calibration data, estimate $\vec{w}_h$, predict consideration for the left-out profile, and choose "tuning" parameters to minimize prediction errors on the heldout profiles. (No data from any holdout or validation observations are used in leave-one-out cross validation.)

In our data, neither leave-one-out-cross-validation nor out-of-sample predictions are particularly sensitive to our choice of "tuning" parameters within ranges that roughly match a priori beliefs. Such robustness is consistent with Evgeniou, Pontil and Toubia (2007). Specifically, we can choose any g_M that is an arbitrarily small number such that sample-level consideration is used only to break ties among patterns. For g_c, cross-validation (and predictive tests) vary little in the range $g_c \in [1, 4.5]$. Similarly, we can select a cognitively-simple S to be within ranges that we observe in qualitative interviews ($S \sim 2, 3, 4$). We report $S = 4$ for ease of exposition.

Logical Analysis of Data (LAD-DOC)

Logical analysis of data (LAD), which seeks to distinguish "positive" events from "negative" events, is another approach to generate patterns (Boros, et. al. 1997; 2000). We control cognitive simplicity by limiting the search to at most P patterns of size at most S. We define positive patterns as patterns that match at least one considered profile, but no not-considered profile. Following the "bottom-up" approach described by Boros, et al, 2000, we begin by generating minimal patterns of length one that match some considered profiles. If such patterns are not contained in any non-considered profile, they are positive patterns. Otherwise, we add aspects to the patterns one by one until we generate positive pattens, or until we reach maximum length (S). We next use a greedy algorithm to identify up to P positive patterns that best fit the data, breaking ties first by giving preference to shorter patterns and then patterns that are positive most frequently in the sample. The union of these positive patterns is a DOC rule.

LAD-DOC provides a contrast to DOCMP. It is simpler to formulate and takes less time to run, but shares the characteristics of selecting those patterns that best fit the data subject to cognitive simplicity (S, P) and shrinkage (break ties to fit sample-level consideration). One potential weakness is that our implementation of LAD focuses primarily on avoiding false positives (in the calibration data) rather than a combination of false positives and false negatives. For comparability to DOCMP we set $S = 4$ and $P = 2$, but out-of-sample predictions are comparable for $P \sim 2, 3$, or 4 and $S \sim 4$ or 5.

BENCHMARKS

We choose as benchmarks five decision rules. These rules are estimated with both machine-learning and with hierarchical Bayes methods. The decision rules are:

- additive partworth rules
- additive q-compensatory rules
- disjunctive rules
- conjunctive rules
- subset conjunctive rules

The machine-learning estimations use objective functions comparable to Equation 1. For the additive and q-compensatory rules, we penalize the sum of the partworths rather than the number of patterns. Detailed formulations are available in the Web Appendix.

The hierarchical Bayes methods mimic extant methods to the greatest extent possible. For the additive and q-compensatory rules we use standard HB choice-based conjoint formulations adapted to our dependent variable (consideration vs. not). We use rejection sampling to enforce the q-compensatory constraint (e.g., Allenby, Arora and Ginter 1995). For subset conjunctive rules we modify an algorithm developed by Gilbride and Allenby (2004). The modifications reflect differences in data and generalization ($S = 1$ or F in Gilbride and Allenby 2004). As data, we observe consideration directly while it is a latent construct in the Gilbride-Allenby formulation. To address unordered multi-level features, we do not impose constraints that levels within a feature are ordered. Detailed HB formulations are available in the Web Appendix.

For the subset conjunctive rules, we select $S = 4$ to be consistent with the DOC rules. Predictive tests for other values of S are available from the authors.[2] In addition to detailed formulations, the Web Appendix also contains simulations which compare some of the benchmarks to DOC-based methods on synthetic data.[3]

EMPIRICAL APPLICATION – GLOBAL POSITIONING SYSTEMS (GPSs)

We chose to study GPSs because the number of features and the number of brands available is sufficiently large that we might expect some non-compensatory decision rules. Figure 1 illustrates sixteen features that consumers use to evaluate handheld GPSs. These features were chosen as the most important based on two pretests of 58 and 56 consumers, respectively. Ten of the features are represented by text and icons while the remaining six features are represented by text and visual cues.

[Insert Figures 1 and 2 about here.]

[2] The basic relative comparisons with DOC-based models are similar for $S \sim 1, 2, 3$, or 4.

[3] The simulations are consistent with intuition and are consistent with empirical results in the domain suggested by the empirical data. For example, when the data are generated with a particular decision rule, the estimation models which assume that decision rule tend to predict (out of sample) best.

Using the sixteen features we generated an orthogonal design of 32 GPS profiles.[4] We then developed four alternative formats by which to measure consideration. These respondent task formats were developed based on qualitative pretests to approximate the shopping environment for GPSs. Each respondent task format was implemented in a web-based survey and pretested extensively with over 55 potential respondents from the target market. At the end of the pretests respondents found the tasks easy to understand and felt that the task formats were reasonable representations of the handheld GPS market.

We invited two sets of respondents to complete the web-based tasks: a representative sample of German consumers who were familiar with handheld GPSs and a US-based student sample. We first describe results from our primary format using the German sample of representative consumers. We then discuss the other formats, the student sample, and a text-only version.

Figure 2 provides screen-shots in English and German for the basic format. A "bullpen" is on the far left. As respondents move their cursor over a generic image in the bullpen, a GPS appears in the middle panel. If respondents click on the generic image, they can evaluate the GPS in the middle panel deciding whether or not to consider it. If they decide to consider the GPS, its image appears in the right panel. Respondents can toggle between current consideration sets and their current not-consider sets. There are many ways in which they can change their mind, for example, putting a GPS back or moving it from the consideration set to the not-consider set, or vice versa. In this format respondents continue until all GPSs are evaluated.

Before respondents made consideration decisions, they reviewed screens that described GPSs in general and each of the GPS features. They also viewed instruction screens for the consideration task and instructions that encouraged incentive compatibility. Following the consideration task respondents ranked profiles within the consideration set (data not used in this paper) and then completed tasks designed to cleanse memory. These tasks included short brain-teaser questions that direct respondents'

[4] To make the task realistic and to avoid dominated profiles (Johnson, Meyer and Ghose 1989), price was manipulated as a two-level price increment. Profile prices were based on this increment plus additive feature-based costs. We return to the issue of orthogonal designs at the end of this section.

attention away from GPSs. Following the memory-cleansing tasks, respondents completed the consideration task a second time, but for a different orthogonal set of GPSs. These second consideration decisions are validation data and are not used in the estimation of any rules.

Respondents were drawn from a web-based panel of consumers maintained by the GfK Group. Initial screening eliminated respondents who had no interest in buying a GPS and no experience using a GPS. Those respondents who completed the questionnaire received an incentive of 200 points toward general prizes (Punkte) and were entered in a lottery in which they could win one of the GPSs (plus cash) that they considered. This lottery was designed to be incentive compatible as in Ding (2007) and Ding, Grewal, and Liechty (2005). (Respondents who completed only the screening questionnaire received 15 Punkte.)

In total 2,320 panelists were invited to answer the screening questions. The incidence rate (percent eligible) was 64%, the response rate was 47%, and the completion rate was 93%. Respondents were assigned randomly to one of the five task formats (the basic format in Figure 2, three alternative formats, and a text-only format). After eliminating respondents who had null consideration sets or null not-consider sets in the estimation task, we retained 580 respondents. The average size of the consideration set (estimation data) for the task format in Figure 2 was 7.8 profiles. There was considerable variation among respondents (standard deviation was 4.8 profiles). The average size of the consideration set in the validation task was smaller, 7.2 profiles, but not significantly different. Validation consideration set sizes had an equally large standard deviation (4.8 profiles).

PREDICTIVE TESTS

Criteria to Compare DOCMP, LAD-DOC and the Benchmarks

Hit rate is an intuitive measure which is used commonly when comparing predictive ability. However, with average consideration sets around 7.2 out of 32 (22.5%), a null model that predicts that no GPSs will be considered will achieve a hit rate of 77.5%. Thus, we follow Srinivasan (1988), Srinivasan and Park (1997), and Payne, Bettman and Johnson (1993, p. 128) and report the percent improvement relative to a random-prediction null model. Percent improvement is a linear transformation of hit rate, but it is easier to interpret.

More critically, the apparent strong performance of "predict nothing considered" suggests that we gain insight with statistics that reward models that actually try to predict consideration. The ability to predict the consideration-set size can reject bad models, but is not sufficient to evaluate a good model. A null model of random prediction (proportional to calibration consideration-set size) predicts the validation consideration-set size accurately but achieves a low hit rate of 65.3% and provides no useful information (0% relative hit-rate improvement).

Instead we consider a statistic that is sensitive to false positive predictions, false negative predictions, <u>and</u> predicted consideration-set sizes in the validation data. In particular, we use the Kullback-Leibler divergence (K-L) which measures the expected gain in Shannon's information measure relative to a random model (Chaloner and Verdinelli 1995; Kullback and Leibler 1951; Lindley 1956).[5] The K-L percentage is 0% for both the random null model and the "predict-nothing-considered" null model. It is 100% for perfect prediction. The K-L percentage rewards models that predict the consideration-set size correctly <u>and</u> favors a mix of false positives and false negatives that reflect true consideration sets over those that do not. It discriminates among models even when the hit rates might otherwise be equal. Together the three statistics, hit rate, relative hit rate improvement, and the K-L percentage, provide a means to assess relative predictive ability (DOC-based models vs. the benchmarks).

Predictive Tests

Table 1 summarizes the ability of each estimation method to predict consideration for the validation task. Focusing on the comparison of DOC-based models to the benchmarks, DOC-based predictions are best or not significantly different than best on both hit rates and K-L percentage measures

[5] Formulae for K-L percentage for consideration-set prediction are available in the Web Appendix. K-L acts for 0-vs.-1 predictions much like U^2 does for probabilistic predictions (Hauser 1978).

and better than all benchmark estimation methods on both measures. LAD-DOC predicts slightly better than DOCMP, but the difference is not significant.

Among the benchmarks the additive-rule models predict well, with the machine-learning version significantly better than the HB version on both hit rate and K-L percentage ($t = 2.6, p < 0.02$; $t = 3.7, p < 0.01$, respectively). While the DOC-based methods are best or not significantly different than best on all comparisons, the machine-learning additive model is within 1-2 percentage points on hit rate.[6] This is consistent with prior results on the robustness of the linear model for empirical data (e.g., Dawes 1979; Dawes and Corrigan 1974) and consistent with the ability of an additive rule to nest some non-compensatory rules.

Estimations based on the DOC generalization predict significantly better than the non-compensatory benchmarks suggesting the generalization improves predictions for at least some of our respondents.[7] The unconstrained additive models, which can represent both q-compensatory and many of non-compensatory models, predict better than the q-compensatory models on both measures, significantly so for the machine-learning algorithms ($t = 2.1, p < 0.04$ for hit rates; $t = 9.4,\ p < 0.01$ for K-L). At the level of the individual respondent, some respondents are fit much better with an unconstrained model and some much better with a q-constrained model. Future research might investigate correlates of these individual differences.

For brevity we do not elaborate further on comparisons <u>among</u> the benchmarks themselves. Our data are available for readers who wish to explore machine learning. HB, or other methods for the benchmark rules.

---

[6] LAD-DOC is significantly better than the best (machine-learning) additive model on both hit rate and K-L divergence ($t = 2.4, p < 0.02$; $t = 4.6, p < 0.01$), DOCMP is better, but not quite significantly so, on hit rate and significantly better on K-L divergence ($t = 1.9, p = 0.06$; $t = 4.1, p < 0.01$). One reason the additive model does less well on the K-L percentage is that it under-predicts the consideration-set size. We examine the predictive ability of the additive model further in the next section.

[7] We note the poor performance of the machine-learning subset conjunctive model with $S = 16$. With $S = 16$ and a goal of choosing 0 vs. 1 for w_{hp}, the subset-conjunctive integer program tends to over fit the calibration data.

Empirical Evidence is Consistent with Cognitive Simplicity

Although DOCMP and LAD-DOC are designed to favor cognitive simplicity, unconstrained estimation could conceivably predict better. We re-estimated DOCMP with the gs equal to zero and LAD-DOC without the S and P constraints. For both models, the hit rates are significantly better for the penalized/constrained estimation ($p < 0.01$ vs. 75.7% DOCMP without gs; $p < 0.01$ vs. 80.4% LAD-DOC without constraints, respectively). Cognitive simplicity also improves the K-L percentage, but the improvements are not quite significant ($p < 0.16$ vs. 29.6%; $p = 0.07$ vs. 32.5%, respectively for unconstrained DOCMP and LAD-DOC). These results are consistent with an hypothesis that predictions improve when cognitive simplicity is enforced, although the marginal significance for K-L percentages suggests that the cognitive-simplicity hypothesis is worth further testing in other contexts.

Despite the large number of potential patterns, DOC-based estimation chose relatively simple rules for our data. LAD-DOC predictions do not improve significantly, and often degrade, as we increase either pattern length (S) or the number of patters (P). For DOCMP, 7.1% of the respondents are represented as using two patterns; the remainder with a single pattern. It is interesting that the increased flexibility of the DOC-based estimation methods seems to improve predictive ability relative to alternative non-compensatory rules and their corresponding estimation methods even though only 7.1% of the respondents are modeled with two patterns.

Sensitivity to Orthogonal Designs

There has been significant research in marketing on efficient experimental designs for choice-based conjoint experiments (Arora and Huber 2001; Huber and Zwerina 1996; Kanninen 2002; Toubia and Hauser 2007), but we are unaware of any research on efficient experimental designs for consideration decisions or for the estimation of cognitively-simple DOC rules. When decisions are made with respect to the full set of 32 profiles, aspects are uncorrelated up to the resolution of the design and, if there were no errors, we should be able to identify DOC patterns accordingly. However, when profiles are removed aspects may no longer be uncorrelated and patterns may not be defined uniquely. As a mild test, we re-estimated three models, DOCMP, machine learning additive, and HB additive, with only 17 of 32 most-popular profiles (#'s 16-17 were tied). DOCMP remained significantly better on the K-L percentages and best or not significantly different than best on hit rates, even though we are now estimating the models with approximately half the observations per respondent.

Until the issue of optimal DOC-consideration experimental designs is resolved, the performance of DOC-based estimation methods remains a conservative predictive test. Improved or adaptive experimental designs might improve performance.

Summary of Empirical Results

DOC-based estimation appears to predict hit rates well and provide information (K-L percentage) about consideration decisions on validation data. Predictions appear to be better with DOC-based estimation than with any of the other five decision rules for both machine-learning and HB estimation, although an unconstrained machine-learning additive model (which can represent some non-compensatory rules) comes close. Some of this improvement is due to cognitive simplicity.

TARGET POPULATION, TASK FORMAT, AND PROFILE REPRESENTATION

We examine hypotheses that the predictive ability is unique to the task format, to the GfK respondents, or to the way we present profiles.

Variations in Task Formats

With the format analyzed in the previous section respondents must evaluate every profile ("evaluate all profiles"). However, such a restriction may be neither necessary nor descriptive. For example, Ordóñez, Benson and Beach (1999) argue that consumers screen products by rejecting products that they would not consider further. Because choice rules are context dependent (e.g., Payne, Bettman and Johnson 1993), the task format could influence the propensity to use a DOC rule.

To examine context sensitivity, we tested alternative task formats. One format asked respondents to indicate only the profiles they would consider ("consider only"); another asked respondents to indicate only the profiles they would reject ("reject only"). The tasks were otherwise identical to "evaluate all profiles." We also tested a "no browsing" format in which respondents evaluated profiles sequentially (in a randomized order). Representative screen shots for these formats, as well as example feature-introduction, and instruction screenshots, are provided in the Web Appendix.

The predictive results mimic the results in Table 1.[8] On the K-L percentages both DOC-based estimation methods were significantly better than all benchmarks on all four formats. On hit rate at least one of the DOC-based estimation methods was best on all formats, significantly better than all benchmarks for the majority of the formats (3 of 4), and significantly better than 9 of the 10 benchmarks for the remaining format. On hit rate, the only estimation method that did not differ significantly from a DOC-based estimation method on that one format was the machine-learning additive model – a result similar to that which we observed in Table 1. To test DOC-based methods further, we merged the data from the four formats and compared DOCMP and LAD-DOC hit rates to the additive machine-learning method. When the hit-rate data are merged, both DOCMP and LAD-DOC predict significantly better than the additive machine-learning method ($t = 4.4$, $p < 0.01$; $t = 3.0$, $p < 0.01$).

As predicted by the evaluation-cost theory of consideration-set formation, respondents considered fewer profiles when the relative evaluation cost (for consideration) was higher: 4.3 profiles in "consider only," 7.8 in "evaluate all," and 10.6 in "reject only." As predicted by the theory of context dependence, the propensity to use a second DOC pattern varied as well. Second disjunctions were more common when consideration sets were larger: 0% for "consider only," 7.1% for "evaluate all," and 9.8% for "reject only." While our data cannot distinguish whether these differences are due to the size of the consideration set or due to differential evaluation costs induced by task variation, these data illustrate how the predictive tests complement more direct (but possibly more intrusive) experimental measures.

[8] Tables for the other formats are provided in the Web Appendix.

US Student Sample vs. Representative German Sample

We replicated the "evaluate-all-profiles" GPS measurement with a sample of MBA students at a US university. Students were invited to an English-language website (e.g., first panel of Figure 1). As incentives, and to maintain incentive-compatibility, they were entered in a lottery with a 1-in-25 chance of winning a laptop bag worth $100 and a 1-in-100 chance of winning a combination of cash and one of the GPSs that they considered. The response rate for US students was lower, 26%, and consideration-set sizes were, on average, larger. Despite the differences in sample, response rate, incentives, and consideration-set size, DOCMP and LAD-DOC predicted validation data best (or were not significantly different than the best) on both hit rates and K-L percentages. (The best benchmark was again the additive machine-learning model. Lower sample sizes for the US sample made it more difficult to distinguish differences.)

Text-Only vs. Visual Representation of the GPS Profiles

The profile representations in Figure 1 were designed by a professional graphic artist and were pretested extensively. Pretests suggested which features should be included in the "JPEGs" and which features should be included as satellite icons. Nonetheless, it is possible that the relative predictive ability of the estimation methods might depend upon the specific visual representations of the profiles. To examine this hypothesis we included a task format that was identical to the task in "consider all profiles" except that all features were described by text rather than pictures, icons, and text. The DOC-based estimation methods are again the best predictive methods – significantly better on K-L percentages and best or not significantly different than the best on hit rates. Once again, the additive machine learning method does as well on hit rate but not the K-L percentage. We cannot distinguish with our data whether this is a text-only effect or a result consistent with the analyses of the other formats.

Interestingly, there is no significant difference in hit rates or K-L percentages between picture representations and text representations for either DOCMP or LAD-DOC.

Summary of Robustness Tests

The relative predictive ability of the tested methods appears to be robust with respect to:

- format of the respondent task (evaluate all profiles, consideration only, rejection only, or no browsing),
- respondent sample (representative German vs. US student),
- presentation of the stimuli (pictures vs. text).

MANAGERIAL IMPLICATIONS AND DIAGNOSTICS

We were motivated to study consideration-set decisions with a managerial challenge: "How can a firm increase the likelihood that its products will be considered?" We hope that by estimating DOC-based models we might gain insight to help a firm enhance consideration. If the improved predictive ability of DOC-based models holds up to further testing, then market-response simulators using DOC-based models might be more accurate than market-response simulators based on conjunctive, disjunctive, subset conjunctive, q-compensatory, or additive-rule decision rules. (See Geisser 1993 for a discussion of using predictive models to evaluate strategies.) To illustrate how models affect managerial decisions differently we compare the simulated value of feature improvements between estimated DOC rules and estimated additive rules. Our data are available for readers who wish to explore other comparisons.

Changes in Market Share as a Function of Feature Improvements

Ofek and Srinivasan (2002, p. 401) propose that a value of a feature be defined as "the incremental price the firm would charge per unit improvement in the product attribute (assumed to be infinitesimal) if it were to hold market share (or sales) constant." In DOC rules features and price levels are discrete, hence we modify their definition slightly. We compute the incremental improvement in market share if a feature is added for an additional $50 in price. Because this calculation is sensitive to the base product, we select the features of the base product randomly. We illustrate two of the many differences between DOC rules and additive rules. In both of these situations the recommended managerial decision depends upon whether consumers consider products based on the estimated DOC rules or based on the estimated additive rules.

(Example 1). DOC rules predict that that consideration share would increase if we switch to Garmin and raise the price by $50, but compensatory rules predict that consideration share would decrease. To understand this difference intuitively, we recognize that the estimated DOC rules imply that 12% of the respondents screen on brand and, of those, 82% screen on Garmin. The remaining respondents screen on other features. With an additive-partworth rule, 54% of the respondents have slightly higher partworths for Magellen. With DOC-rules the advantage to Garmin comes from the 12% who screen on brand, but with additive rules the advantage to Magellen comes a little from all the respondents in the sample.

(Example 2). Additive rules predict that "extra bright" is the highest-valued feature improvement yielding an 11% increase for the $50 price. However, DOC rules predict a much smaller improvement (2%) because many of the respondents who screen on "extra bright" also eliminate GPSs with the higher price.

Diagnostic Summaries of DOC Rules

Diagnostic summaries of additive partworths have been developed through decades of application. Recent developments have added heterogeneity with corresponding challenges in how best to summarize heterogeneity to managers. Diagnostic summaries of non-compensatory decision rules are relatively nascent. Academics and practitioners are still evolving the best way to summarize such rules for managers.

This challenge is exacerbated for DOC rules. Even with cognitive simplicity ($S = 4$) there are 34,112 potential DOC patterns. Listing each pattern that matches consideration in a sample of respondents is not nearly as diagnostic as the feature-improvement simulations which aggregate across identified patterns. As a first attempt, we examined summaries of first- and second-order inclusion.

(Gilbride and Allenby 2004 and Yee, et. al. report first-order inclusion.) For example, the mini-USB port appeared in at least one DOC conjunction for 36% of the respondents. Extra-bright displays (25%) and color displays (21%) were the next highest contributors. With second-order inclusions we see, for example, that those respondents who want a long battery life also want a mini-USB port (50%) and a color display (40%). Such first- and second-order conjunctive-inclusions provide insight which complement DOC-model-based market-response simulators. As in the market-response simulations these simple diagnostics vary from what one might infer from additive partworths.

We hope that such diagnostic information combined with market-response simulators will help managers evaluate product-line changes and marketing activities. With more experience, researchers might develop more intuitive ways to summarize DOC patterns for managers.

SUMMARY AND FUTURE DIRECTIONS

Consideration sets have become relevant to managerial decisions in many product categories and, whenever there are many products available and/or products are described by many features and levels, extant research suggests that consumers use non-compensatory decision rules to make consideration decisions. Research suggests further that such decision rules are often cognitively simple. We hope we have contributed to these literatures.

Drawing on qualitative research we propose a generalization of established non-compensatory decision rules: disjunctions of conjunctions (DOC). We posit further that DOC-rules will be cognitively simple and that models that attempt to represent cognitively-simple DOC rules will predict better than models that do not. We examine two machine-learning estimation methods, DOCMP and LAD-DOC, comparing predictions to five decision-rule models as implemented by both machine-learning and HB estimation methods.

The preponderance of the empirical evidence in this paper suggests that DOC rules and both estimation algorithms are worth further investigation. Both are significantly better on K-L percentages for all 10 benchmarks, all four respondent task formats, German and US data, and both highly visual and text-only stimuli. We get the same perspective with hit rates with one important exception. The machine-learning additive method does almost as well for some formats, a result consistent with the known robustness of the additive model and with ability of the additive model to represent some non-compensatory decision rules.

Our results must be considered hypotheses for further testing. The handheld GPS category has many features and, at the time of our test, was relatively new to our respondents. This provided a "proof-of-concept" test for DOC-based methods. In more-familiar or simpler categories, additive models might suffice. On the other hand, more complex categories, such as automobiles, might favor DOC rules.

We chose two methods to estimate DOC rules. There are likely others. For example, decision trees can also represent DOC rules (Breiman, et. al. 1984; Currim, Meyer and Le 1988). If researchers can develop a way to model cognitive simplicity on decision trees, this approach might prove promising. If features are continuous, then DOC rules are similar to specific interactions in a multilinear decision rule (Bordley and Kirkwood 2004; Mela and Lehmann 1995). With sufficient creativity and experimentation researchers might extend finite-mixture, Bayesian, simulated-maximum-likelihood, Markov, or kernel estimators to estimate cognitively simple continuous DOC analogs (Evgeniou, Boussios, and Zacharia 2005; Hauser and Wisniewski 1982; Mela and Lehmann 1995; Rossi and Allenby 2003; Swait and Erdem 2007).

Finally, we focused on the consideration stage of a consider-then-choice rule. DOC rules might also apply to the choice stage. One might also investigate a model that is DOC in the first stage and compensatory in the second stage. There is a rich history in marketing of two-stage models in which consideration is a latent, unobserved construct (e.g., Andrews and Srinivasan 1995; Gensch 1987; Gilbride and Allenby 2004; Siddarth, Bucklin, and Morrison 1995; Swait and Erdem 2007). We believe that DOC rules combined with cognitive simplicity could complement these lines of research.

REFERENCES

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1995), "Incorporating Prior Knowledge into the Analysis of Conjoint Studies," *Journal of Marketing Research*, 32, (May), 152-162.

Andrews, Rick L. and T. C. Srinivasan (1995), "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data," *Journal of Marketing Research*, 32, (February), 30-41.

Arora, Neeraj and Joel Huber (2001), "Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments," *Journal of Consumer Research*, 28, (September), 273-283.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25(3), 187-217.

Bordley, Robert F. and Craig W. Kirkwood (2004), "Multiattribute Preference Analysis with Performance Targets," *Operations Research*, 52, 6, (November-December), 823-835.

Boros, Endre, Peter L. Hammer, Toshihide Ibaraki, and Alexander Kogan (1997), "Logical Analysis of Numerical Data," *Mathematical Programming*, 79:163--190, August 1997

------, ------, ------, ------, Eddy Mayoraz, and Ilya Muchnik (2000), "An Implementation of Logical Analysis of Data," *IEEE Transactions on Knowledge and Data Engineering*, 12(2), 292-306.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone (1984), *Classification and Regression Trees*, (Belmont, CA: Wadsworth).

Bröder, Arndt (2000), "Assessing the Empirical Validity of the "Take the Best" Heuristic as a Model of Human Probabilistic Inference," *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 26, 5, 1332-1346.

Bronnenberg, Bart J., and Wilfried R. Vanhonacker (1996), "Limited Choice Sets, Local Price Response, and Implied Measures of Price Competition," *Journal of Marketing Research,* 33 (May), 163-173.

Chaloner, Kathryn and Isabella Verdinelli (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10, 3, 273-304. (1995)

Chase, Valerie M., Ralph Hertwig, and Gerd Gigerenzer (1998), "Visions of Rationality," *Trends in Cognitive Sciences*, 2, 6, (June), 206-214.

Cooil, Bruce, Russell S. Winer and David L. Rados (1987), "Cross-Validation for Prediction," *Journal of Marketing Research,* 24, (August), 271-279.

Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest and Clifford Stein (2001), *Introduction to Algorithms, 2E,* (Cambridge, MA: MIT Press).

Cucker, Felipe, and Steve Smale (2002), "On the Mathematical Foundations of Learning," *Bulletin of the American Mathematical Society*, 39(1), 1-49.

Currim, Imran S., Robert J. Meyer, and Nhan T. Le (1988), "Disaggregate Tree-Structured Modeling of Consumer Choice Data," *Journal of Marketing Research,* 25(August), 253-265.

Dawes, Robyn M. (1979), "The Robust Beauty of Improper Linear Models in Decision Making," *American Psychologist*, 34, 571-582.

------ and Bernard Corrigan (1974), "Linear Models in Decision Making," *Psychological Bulletin*, 81, 95-106.

DeSarbo, Wayne S., Donald R. Lehmann, Gregory Carpenter, and Indrajit Sinha (1996), "A Stochastic Multidimensional Unfolding Approach for Representing Phased Decision

Outcomes," *Psychometrika*, 61(3), 485-508.

Ding, Min (2007), "An Incentive-Aligned Mechanism for Conjoint Analysis," *Journal of Marketing Research*, 54, (May), 214-223.

------, Rajdeep Grewal, and John Liechty (2005), "Incentive-Aligned Conjoint Analysis," *Journal of Marketing Research*, 42, (February), 67–82.

Efron, Bradley and Robert Tibshirani (1997), "Improvements on Cross-Validation: The .632+ Bootstrap Method," *Journal of the American Statistical Association*, 92, 438, (June), 548-560.

Evgeniou, Theodoros, Constantinos Boussios, and Giorgos Zacharia (2005), "Generalized Robust Conjoint Estimation," *Marketing Science*, 24(3), 415-429.

------, Massimiliano Pontil, and Olivier Toubia (2007), "A Convex Optimization Approach to Modeling Heterogeneity in Conjoint Estimation," *Marketing Science*, 26, 6, (November-December), 805-818.

Feige, Uriel (1998), "A threshold of ln n for approximating set cover," *Journal of the Association for Computing Machinery*, 45(4), 634 – 652.

Gensch, Dennis H. (1987), "A Two-stage Disaggregate Attribute Choice Model," *Marketing Science*, 6, (Summer), 223-231.

Geisser, Seymour (1993), *Predictive Inference: An Introduction*, (New York, NY: Chapman & Hall).

Gigerenzer, Gerd and Daniel G. Goldstein (1996), "Reasoning the Fast and Frugal Way: Models of Bounded Rationality," *Psychological Review*, 103(4), 650-669.

------, Peter M. Todd, and the ABC Research Group (1999), *Simple Heuristics That Make Us Smart*, (Oxford, UK: Oxford University Press).

Gilbride, Timothy J. and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23(3), 391-406.

------ and ------ (2006), "Estimating Heterogeneous EBA and Economic Screening Rule Choice Models," *Marketing Science*, 25, 5, (September-October), 494-509.

Hastie, Trevor, Robert Tibshirani, Jerome H. Friedman (2003), *The Elements of Statistical Learning*, (New York, NY: Springer Series in Statistics).

Hauser, John R. (1978), "Testing the Accuracy, Usefulness and Significance of Probabilistic Models: An Information Theoretic Approach," *Operations Research*, Vol. 26, No. 3, (May-June), 406-421.

------ and Birger Wernerfelt (1990), "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research*, 16 (March), 393-408.

------ and Kenneth J. Wisniewski (1982), "Dynamic Analysis of Consumer Response to Marketing Strategies," *Management Science*, 28, 5, (May), 455-486.

Hogarth, Robin M. and Natalia Karelaia (2005), "Simple Models for Multiattribute Choice with Many Alternatives: When It Does and Does Not Pay to Face Trade-offs with Binary Attributes," *Management Science*, 51, 12, (December), 1860-1872.

Huber, Joel, and Klaus Zwerina (1996), "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research*, 33 (August), 307-317.

Hughes, Marie Adele and Dennis E. Garrett (1990), "Intercoder Reliability Estimation Approaches in Marketing: A Generalizability Theory Framework for Quantitative Data," *Journal of Marketing Research*, 27, (May), 185-195.

Jedidi, Kamel and Rajeev Kohli (2005), "Probabilistic Subset-Conjunctive Models for Heterogeneous Consumers," *Journal of Marketing Research*, 42 (4), 483-494.

------, ------ and Wayne S. DeSarbo (1996), "Consideration Sets in Conjoint Analysis," *Journal of Marketing Research*, 33 (August), 364-372.

Johnson, Eric J., Robert J. Meyer, and Sanjoy Ghose (1989), "When Choice Models Fail: Compensatory Models in Negatively Correlated Environments," *Journal of Marketing Research*, 26, (August), 255-290.

Kanninen, Barbara J. (2002), "Optimal Design for Multinomial Choice Experiments," *Journal of Marketing Research*, 39, (May), 214-227.

Kearns, Michael and Dana Ron (1999), "Algorithmic Stability and Sanity-Check Bounds for Leave-One-Out Cross-Validation," *Neural Computation*, 11, 1427–1453.

Kohavi, Ron (1995), "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. 2, 12, 1137-1143.

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

Kullback, Solomon, and Leibler, Richard A. (1951), "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79-86.

Langley, Pat (1996), *Elements of Machine Learning*, (San Francisco, CA: Morgan Kaufmann).

Lindley, Dennis V. (1956), "On a Measure of the Information Provided by an Experiment," *The Annals of Mathematical Statistics*, 27, 4 (December), 986-1005.

Lund, Carsten, and Mihalis Yannakakis (1994), "On the Hardness of Approximating Minimization Problems," *Journal of the Association for Computing Machinery*, 41(5), 960 - 981

Martignon, Laura and Ulrich Hoffrage (2002), "Fast, Frugal, and Fit: Simple Heuristics for

Paired Comparisons," *Theory and Decision*, 52, 29-71.

Mehta, Nitin, Surendra Rajiv, and Kannan Srinivasan (2003), "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation," *Marketing Science*, 22(1), 58-84.

Mela, Carl F. and Donald R. Lehmann (1995), "Using Fuzzy Set Theoretic Techniques to Identify Preference Rules From Interactions in the Linear Model: An Empirical Study," *Fuzzy Sets and Systems*, 71, 165-181.

Montgomery, H. and O. Svenson (1976), "On Decision Rules and Information Processing Strategies for Choices among Multiattribute Alternatives," *Scandinavian Journal of Psychology*, 17, 283-291.

Ofek, Elie, and V. Srinivasan (2002), "How Much Does the Market Value an Improvement in a Product Attribute?," *Marketing Science*, 21, 4, (Fall), 398-411.

Ordóñez, Lisa D., Lehmann Benson III, and Lee Roy Beach (1999), "Testing the Compatibility Test: How Instructions, Accountability, and Anticipated Regret Affect Prechoice Screening of Options," *Organizational Behavior and Human Decision Processes*, 78, 1, (April), 63-80.

Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search," *Organizational Behavior and Human Performance*, 16, 366-387.

------, James R. Bettman and Eric J. Johnson (1988), "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 534-552.

------, ------ and ------ (1993), *The Adaptive Decision Maker*, (Cambridge UK: Cambridge University Press)..

Perreault, William D., Jr. and Laurence E. Leigh (1989), "Reliability of Nominal Data Based on Qualitative Judgments," *Journal of Marketing Research*, 26, (May), 135-148.

Roberts, John H., and James M. Lattin (1991), "Development and Testing of a Model of Consideration Set Composition," *Journal of Marketing Research*, 28 (November), 429-440.

Rossi, Peter E., Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22(3), p. 304-328.

Shao, Jun (1993), "Linear Model Selection by Cross-Validation," *Journal of the American Statistical Association*, 88, 422, (June), 486-494.

Shocker, Allan D., Moshe Ben-Akiva, Bruno Boccara, and Prakash Nedungadi (1991), "Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions," *Marketing Letters*, 2(3), 181-197.

Shugan, Steven (1980), "The Cost of Thinking," *Journal of Consumer Research*, 27(2), 99-111.

Siddarth, S., Randolph E. Bucklin, and Donald G. Morrison (1995), "Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data," *Journal of Marketing Research*, 33, (August), 255-266.

Simon, Herbert A. (1955), "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics*, 69(1). 99-118.

Srinivasan, V. (1988), "A Conjunctive-Compensatory Approach to The Self-Explication of Multiattributed Preferences," *Decision Sciences*, 295-305.

------ and Chan Su Park (1997), "Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement," *Journal of Marketing Research*, 34, (May), 286-291.

Swait, Joffre and Tülin Erdem (2007), "Brand Effects on Choice and Choice Set Formation Under Uncertainty," *Marketing Science* 26, 5, (September-October), 679-697.

Toubia, Olivier, Theodoros Evgeniou, and John Hauser (2007), "Optimization-Based and Machine-Learning Methods for Conjoint Analysis: Estimation and Question Design," in Anders Gustafsson, Andreas Herrmann and Frank Huber, Eds, *Conjoint Measurement: Methods and Applications*, 4E, (New York, NY: Springer).

------ and John R. Hauser (2007), "On Managerial Efficient Designs," *Marketing Science*, 26, 6, (November-December), 851-858.

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

Urban, Glen L. and John R. Hauser (2004), "'Listening-In' to Find and Explore New Combinations of Customer Needs," *Journal of Marketing*, 68, (April), 72-87.

Vapnik, Vladimir (1998), *Statistical Learning Theory*, (New York, NY: John Wiley and Sons).

Wu, Jianan and Arvind Rangaswamy (2003), "A Fuzzy Set Model of Search and Consideration with an Application to an Online Market," *Marketing Science*, 22(3), 411-434.

Yee, Michael, Ely Dahan, John R. Hauser and James Orlin (2007) "Greedoid-Based Noncompensatory Inference," *Marketing Science*, 26, 4, (July-August), 532-549.

Zhang, Tong (2003), "Leave One Out Bounds for Kernel Methods," *Neural Computation*, 15, 1397–1437.

FIGURE 1

FEATURES OF HANDHELD GPSS

FIGURE 2

CONSIDERATION TASK IN ONE OF THE FORMATS (ENGLISH AND GERMAN)

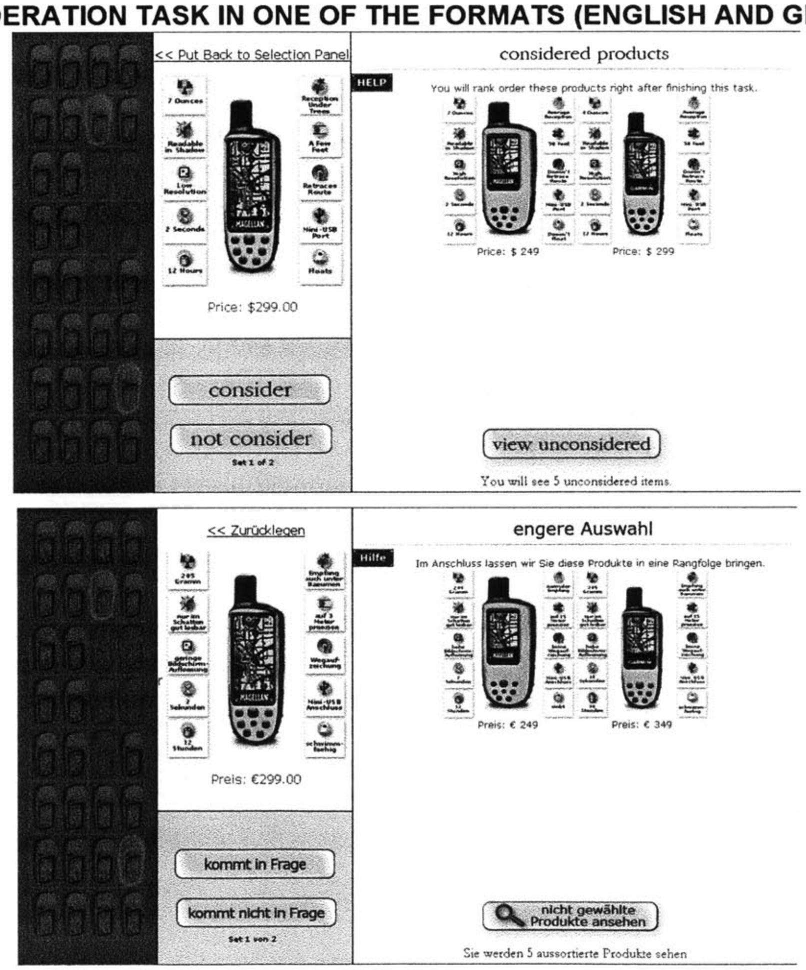


TABLE 1

EMPIRICAL COMPARISON OF ESTIMATION METHODS[a]

(Representative German Sample, Task Format in Figure 2)

Estimation method	Overall hit rate (%)	Relative hit-rate improvement (%)	K-L divergence percentage (%)
Hierarchical Bayes Benchmarks			
Conjunctive (S = 16)	77.7	35.6	6.2
Disjunctive (S = 1)	66.7	3.8	17.8
Subset Conjunctive (S = 4)	75.4	29.0	24.7
q-Compensatory	73.4	37.6	14.6
Additive	78.5	38.0	15.0
Machine-Learning Benchmarks			
Conjunctive (S = 16)	52.6	-36.8	13.3
Disjunctive (S = 1)	77.5	35.6	8.1
Subset Conjunctive (S = 4)	73.7	24.3	6.3
q-Compensatory	76.2	31.3	6.3
Additive	80.6	44.0	23.0
DOC-Based Estimation Methods			
DOCMP (S = 4)	**81.9***	**47.8***	**32.0***
LAD-DOC (S = 4, P = 2)	**82.2***	**48.6***	**34.6***

a Hit rate is the number of profiles predicted correctly, divided by 32.

* Best or not significantly different than the best at the 0.05 level.

Chapter 2: Active Machine Learning for Consideration Heuristics

Abstract

We develop and test an active-machine-learning method to select questions adaptively when consumers use heuristic decision rules. The method tailors priors to each consumer based on a "configurator." Subsequent questions maximize information about the decision heuristics (minimize expected posterior entropy). To update posteriors after each question we approximate the posterior with a variational distribution and use belief-propagation (iterative loops of Bayes updating). The method runs sufficiently fast to select new queries in under a second and provides significantly and substantially more information per question than existing methods based on random, market-based, or orthogonal questions.

Synthetic-data experiments demonstrate that adaptive questions provide close to optimal information and out-perform existing methods even when there are response errors or "bad" priors. The basic algorithm focuses on conjunctive or disjunctive rules, but we demonstrate generalizations to more-complex heuristics and to the use of previous-respondent data to improve consumer-specific priors. We illustrate the algorithm empirically in a web-based survey conducted by an American automotive manufacturer to study vehicle consideration (872 respondents, 53 feature-levels). Adaptive questions outperform market-based questions when estimating heuristic decision rules. Heuristic decision rules predict validation decisions better than compensatory rules.

Keywords: *Active learning, adaptive questions, belief-propagation, conjunctive models, consideration sets, consumer heuristics, decision heuristics, disjunctions-of-conjunctions, lexicographic models, variational-Bayes estimation.*

1. Problem Statement: Adaptive Questions to Identify Heuristic Decision Rules

We develop and test an active machine-learning algorithm to identify heuristic decision rules. Specifically, we select questions adaptively based on prior beliefs and respondents' answers to previous questions. To the best of our knowledge this is the first (near optimal) adaptive-question method focused on consumers' non-compensatory decision heuristics. Extant adaptive methods focus on compensatory decision rules and are unlikely to explore the space of non-compensatory decision rules efficiently (e.g., Evgeniou, Boussios and Zacharia 2005; Toubia, Hauser, and Garcia 2007; Toubia, Hauser and Simester 2004; Sawtooth 1996). In prior non-compensatory applications, question selection was almost always based on either random profiles or profiles chosen from an orthogonal design.

We focus on non-compensatory heuristics because of both managerial and scientific interest. Scientific interest is well-established. Experimental and revealed-decision-rule studies suggest that non-compensatory heuristics are common, if not dominant, when consumers face decisions involving many alternatives, many features, or if they are making consideration rather than purchase decisions (e.g., Gigerenzer and Goldstein 1996; Payne, Bettman and Johnson 1988, 1993; Yee, et al. 2007). Heuristic rules often represent a rational tradeoff among decision costs and benefits and may be more robust under typical decision environments (e.g., Gigerenzer and Todd 1999). Managerial interest is growing as more firms focus product development and marketing efforts on getting consumers to consider their products or, equivalently, preventing consumers from rejecting products without evaluation. We provide illustrative examples in this paper but published managerial examples include Japanese banks, global positioning systems, desktop computers, smart phones, and cellular phones (Ding, et al 2011; Liberali, et al. 2011).

Our focus is on adaptive question selection, but to select questions adaptively we need intermediate estimates after each answer and before the next question is asked. To avoid excessive delays in an online questionnaires, intermediate estimates must be obtained in a second or less (e.g., Toubia, et al. 2004). This is a difficult challenge when optimizing questions for non-compensatory heuristics because we must search over a discrete space of the order of 2^N decision rules, where N is the number of feature-levels (called aspects as in Tversky 1972). Without special structure, finding a best-fitting heuristic is much more difficult than finding best-fitting parameters for (an additive) compensatory model—such estimation algorithms typically require the order of N parameters. The ability to scale to large N is important in practice because consideration heuristics are common in product categories with large numbers of aspects (e.g., Payne, Bettman and Johnson 1993). Our empirical application searches over 9.0 x 10^{15} heuristic rules.

We propose an active-machine-learning solution (hereafter active learning) to select questions adaptively to estimate non-compensatory heuristics. The active-learning algorithm approximates the posterior with a variational distribution and uses belief propagation to update the posterior distribution. It then asks the next question to minimize expected posterior entropy by anticipating the potential responses (in this case, consider or not consider). The algorithm runs sufficiently fast to be implemented between questions in an online questionnaire.

In the absence of error, this algorithm comes extremely close to the theoretical limit of the information that can be obtained from binary responses. With response errors modeled the algorithm does substantially and significantly better than extant question-selection methods. We also address looking ahead *S* steps, generalized heuristics, and the use of population data to improve priors. Synthetic data suggest that the proposed method recovers parameters with fewer questions than extant methods. Empirically, adaptive question selection is significantly better at predicting future consideration than benchmark question-selection. Non-compensatory estimation is also significantly better than the most-commonly applied compensatory method.

We begin with a brief review and taxonomy of existing methods to select questions to identify consumer decision rules. We then review non-compensatory heuristics and motivate their managerial importance. Next we present the algorithm, test parameter recovery with synthetic data, and describe an empirical illustration in the automobile market. We close with generalizations and managerial implications.

2. Existing Methods for Question Selection to Reveal Consumer Decision Rules

Marketing has a long tradition of methods to measure consumer decision rules. Figure 1 attempts a taxonomy that highlights the major trends and provides examples.

[Insert Figure 1 about here.]

The vast majority of papers focus on compensatory decision rules. The most common methods include either self-explication, which asks respondents to self-state their decision rules, or conjoint analysis, which infers compensatory decision rules from questions in which respondents choose, rank, or rate bundles of aspects called product profiles. These methods are applied widely and have demonstrated both predictive accuracy and managerial relevance (e.g., Green 1984; Green and Srinivasan 1990; Wilkie and Pessemier 1973). In early applications, profiles were chosen from either full-factorial, fractional-factorial, or orthogonal designs but as hierarchical-Bayes estimation became popular many researchers moved to random designs to explore interactions better. For choice-based conjoint analysis, efficient designs are a function of the parameters of compensatory decision rules and researchers developed "aggregate customization" to pre-select questions using data from pre-studies (e.g., Arora and Huber 2001). More recently, faced with impatient online respondents, researchers developed algorithms for adaptive conjoint questions based on compensatory models (e.g., Toubia, et al. 2004). After data are collected adaptively, the likelihood principle enables the data to be reanalyzed with models using classical statistics, Bayesian statistics, or machine learning.

In some applications respondents are asked to self-state non-compensatory heuristics. Self-explication has had mixed success because respondents often chose profiles with aspects they had previously stated as unacceptable (e.g., Green, Krieger and Bansal 1988). Recent experiments with incentive-compatible tasks, such as having respondents write an e-mail to a friend who will act as their agent, are promising (Ding, et al. 2011).

Researchers have begun to propose methods to identify heuristic decision rules from directly-measured consideration of product profiles. Finding the best-fit decision rule requires solving a discrete optimization problem which is NP-hard (e.g., Martignon and Hoffrage 2002). Existing estimation uses machine-learning methods such as greedy heuristics, greedoid dynamic programs, logical analysis of data, or linear programming perturbation (Dieckmann, Dippold and Dietrich 2009; Hauser, et al. 2010; Kohli and Jedidi 2007; Yee, et al. 2007). Even for approximate solutions, runtimes are exponential in the number of aspects limiting methods to moderate numbers of aspects. Bayesian methods have been used to estimate parameters for moderate numbers of aspects (e.g., Gilbride and Allenby 2004, 2006; Hauser, et al. 2010; Liu and Arora 2011). To date, profiles for direct consideration measures are chosen randomly or from an orthogonal design.

Within this taxonomy Figure 1 illustrates the focus of this paper (thick box)—adaptive questions for non-compensatory heuristics. We also develop an estimation method for non-compensatory heuristics that scales to large numbers of aspects, even when applied to extant question-selection methods (dotted box). We focus on questions that ask about consideration directly (consider or not). However, our methods apply to any data in which the consumer responds with a yes or no answer and might be extendable to choice-based data where more than one profile is shown at a time. We do not focus on methods where consideration is an unobserved construct inferred from choice data (e.g., Erdem and Swait 2004; van Nierop, et al. 2010). There is one related adaptive method —the first stage of Adaptive Choice-Based Conjoint Analysis (ACBC, Sawtooth 2008) which is based on rules of thumb to select approximately twenty-eight profiles that are variations on a "bring-your-own" profile. Profiles are not chosen optimally and non-compensatory heuristics are not estimated.

3. Non-compensatory Decision Heuristics

We classify decision heuristics by simple and more-complex. The simpler heuristics include conjunctive, disjunctive, lexicographic, take-the-best, and elimination-by-aspects. The more complex heuristics include subset conjunctive and disjunctions of conjunctions. The vast majority of scientific experiments have examined the simpler heuristics with conjunctive the most common (e.g., Gigerenzer and Selten 1999; Payne, Bettman and Johnson 1988, 1993 and references therein). The study of more-complex heuristics, which nest the simpler heuristics, is relatively recent, but there is evidence that some consumers use the more complex forms (Jedidi and Kohli 2005; Hauser, et al. 2010). Both simple and complex heuristics apply for consideration, choice, or other decisions and for a wide variety of product categories. For simplicity of exposition we define the heuristics with respect to the consideration decision and illustrate the heuristics for automotive features.

Conjunctive heuristic. For some features consumers require acceptable ("must have") levels. For example, a consumer might only consider a sedan body type and only consider Toyota, Nissan, or Honda. Technically, for features not in the conjunction, such as engine type, all levels are acceptable.

Disjunctive heuristic. If the product has "excitement" levels of a feature, the product is considered no matter the levels of other features. For example, a consumer might consider all vehicles with a hybrid engine.

Take-the-best. The consumer ranks products on a single most-diagnostic feature and considers only those above some cutoff. For example, the consumer may find "brand" most diagnostic, rank products on brand, and consider only those with brands that are acceptable, say Toyota, Nissan, and Honda.

Lexicographic (by features). This heuristic is similar to take-the-best except the feature need not be most diagnostic. If products are tied on a feature-level, then the consumer continues examining features lower in the lexico-ordering until ties are broken. For example, the consumer might rank on brand, then body style considering only Toyota, Nissan, and Honda and, among those brands, only sedans.

Elimination-by-aspects. The consumer selects an aspect and eliminates all products with unacceptable levels, then selects another aspect and eliminates products with unacceptable levels on that aspect, continuing until only considered products are left. For example, the consumer may eliminate all but Toyota, Nissan, and Honda and all but sedans. Researchers have also examined acceptance-by-aspects and lexicographic-by-aspects which generalize elimination-by-aspects in the obvious ways.

When the only data are consider-vs.-not-consider, it does not matter in which order the profiles were eliminated or accepted. Take-the-best, lexicographic (by features), elimination-by-aspects, acceptance-by-aspects, and lexicographic-by-aspects are indistinguishable from conjunctive heuristics. The rules predict differently when respondents are asked to rank data and they differ in the underlying cognitive process, but they do not differ when predicting the observed consideration set. Disjunctive is a mirror image of conjunctive. Thus, any question selection algorithm that optimizes questions to identify conjunctive heuristics can be applied (perhaps with a mirror image) to any of the simple heuristics.

Subset conjunctive. The consumer considers a product if F features have levels that are acceptable. The consumer does not require all features to have acceptable levels. For example, the consumer might have acceptable brands (Toyota, Honda, Nissan), acceptable body types (sedan), and acceptable engines (hybrid), but only require that two of the three features have levels that are acceptable.

Disjunctions of conjunctions. The consumer might have two or more sets of acceptable aspects. For example, the consumer might consider [Toyota and Honda sedans] or [crossover body types with hybrid engines]. Disjunctions of conjunctions nests the subset conjunctive heuristic and all of the simple heuristics (for consideration). However, its generality is also a curse. Empirical applications require cognitive simplicity to avoid over-fitting data.

All of these decision heuristics are postulated as descriptions of how consumers make decisions. Heuristics are not, and need not be, tied to utility maximization. For example, it is perfectly reasonable for a consumer to screen out low-priced products because the consumer believes that he or she is unlikely to choose such a product if considered and, hence, does not believe that evaluating such a product is worth the time and effort. (Put another way, the consumer would purchase a fantastic product at a low price if he or she knew about the product, but never finds out about the product because the consumer chose not to evaluate low-priced products. When search costs are considered it may be rational for the consumer not to search the lower-priced product because the probability of finding an acceptable low-priced product is too low.)

In this paper we illustrate our question-selection algorithm with conjunctive decision rules (hence it applies to all simple heuristics). We later extend the algorithm to identify disjunctions-of-conjunctions heuristics (which nest subset conjunctive heuristics).

4. Managerial Relevance: Stylized Motivating Example

As a stylized example suppose that automobiles can be described by four features with two levels each: Toyota or Chevy, sedan or crossover body type, hybrid or gasoline engine, and premium or basic trim levels, for a total of eight aspects. Suppose we are managing the Chevy brand which makes only sedans with gasoline engines and basic trim and suppose it is easy to change trim levels but not the other features. If consumers are compensatory and their partworths are heterogeneous and not "too extreme," we can get some consumers to consider our vehicle by offering sufficiently premium trim levels. It might be profitable to do so.

Suppose instead that a segment of consumers is conjunctive on [Toyota $\wedge$ crossover]. (In our notation $\wedge$ is the logical "and;" $\vee$ is the logical "or.") No amount of trim levels will attract these conjunctive consumers. They will not pay attention to Chevy advertising, visit GM.com, or travel to a Chevy dealer—they will never evaluate any Chevy sedans no matter how much we improve them. In another example, if a segment of consumers is conjunctive on [crossover $\wedge$ hybrid] we will never get those consumers to evaluate our vehicles unless we offer a hybrid crossover vehicle no matter how good we make our gasoline-engine sedan. Even with disjunctions of conjunctions, consumers who use [(sedan $\wedge$ hybrid) $\vee$ (crossover $\wedge$ gasoline engine)] will never consider our gasoline-engine sedan. In theory we might approximate non-compensatory heuristics with compensatory partworth decision rules (especially if we include interactions), but if there are many aspects empirical approximations may not be accurate.

Many products just never make it because they are never considered; consumers never learn that the products have outstanding aspects that could compensate for the product's lack of a conjunctive feature. Our empirical illustration is based in the automotive industry. Managers at high levels in the sponsoring organization believe that conjunctive screening was a major reason that the automotive manufacturer faced slow sales relative to other manufacturers. For example, they had evidence that more than half of the consumers in the US would not even consider their brands. Estimates of non-compensatory heuristics are now important inputs to product-design and marketing decisions at that automotive manufacturer.

Non-compensatory heuristics can imply different managerial decisions. Hauser, et al. (2010) illustrate how rebranding can improve the share of a common electronic device if consumers use compensatory models, but not if consumers use non-compensatory models. Ding, et al. (2011) illustrate that conjunctive rules and compensatory rules are correlated in the sense that feature levels with higher average partworth values also appear in more "must have rules." However, the non-compensatory models identify combinations of aspects that would not be considered even though their combined partworth values might be reasonable.

5. Question Types, Error Structure, and Notation

5.1. Question Types and Illustrative Example

Figure 2 illustrates the basic question formats. The example is automobiles, but these types of questions have been used in a variety of product categories—usually durable goods where consideration is easy to define and a salient concept to consumers (Dahan and Hauser 2002; Sawtooth 2008; Urban and Hauser 2004). Extensive pretests suggest that respondents can accurately "configure" a profile that they would consider (Figure 2a). If respondents use only one conjunctive rule in their heuristic they find it difficult to accurately configure a second profile. If they use a disjunctions-of-conjunctions heuristic with sufficiently distinct conjunctions, such as [(Toyota $\wedge$ sedan) $\vee$ (Chevy $\wedge$ truck)], we believe they can configure a second profile "that is different from previous profiles that you said you will consider." In this section we focus on the first configured profile and the corresponding conjunctive heuristic. We return in a later section to address more conjunctions in a disjunctions-of-conjunctions heuristic.

[Insert Figure 2 about here.]

After configuring a considered profile, we ask respondents whether or not they will consider various profiles (Figure 2b). Our goal is to select the profiles that provide the most information about decision heuristics (information is defined below). With synthetic data we plot cumulative information (parameter recovery) as a function of the number of questions. In our empirical test, we ask 29 queries half of which are adaptive and half of which are chosen randomly (proportional to market share). We compare predictions based on the two types of questions. Although the number of questions was fixed in the empirical test, we address how stopping rules can be endogenous to the algorithm.

5.2. Notation, Error Structure, and Question-Selection Goal

Let M be the number of features (e.g., brand, body style, engine type, trim level, $M = 4$) and let N be the total numbers of aspects (e.g., Toyota, Chevy, sedan, crossover, hybrid, gasoline engine, low trim, high trim, $N = 8$). Let i index consumers and j index aspects. For each conjunction, consumer i's decision rule is a vector, $\vec{a}_i$, of length N, with elements a_{ij} such that $a_{ij} = 1$ if aspect j is acceptable and $a_{ij} = -1$ if it is not. For example, $\vec{a}_i = \{+1, -1, +1, -1, +1, -1, +1, +1\}$ would indicate that the i^{th} consumer finds hybrid Toyota sedans with both low and high trim to be acceptable.

Each sequential query (Figure 2b), indexed by k, is a profile, $\vec{x}_{ik}$, with N elements, x_{ijk}, such that $x_{ijk} = 1$ if i's profile k has aspect j and $x_{ijk} = 0$ if it does not. Each $\vec{x}_{ik}$ has exactly M non-zero elements, one for each feature. (In our stylized example, a profile contains one brand, one body type, one engine type, and one trim level.) For example, $\vec{x}_{ik} = \{1, 0, 1, 0,1,0, 1, 0\}$ would be a hybrid Toyota sedan with low trim.

Let X_{iK} be the matrix of the first K profiles given to a consumer; each row corresponds to a profile. Mathematically, profile $\vec{x}_{ik}$ *satisfies* a conjunctive rule $\vec{a}_i$ if whenever $x_{ijk} = 1$, then $a_{ij} = 1$ such that every aspect of the profile is acceptable. In our eight-aspect example, consumer i finds the hybrid Toyota sedan with low trim to be acceptable (compare $\vec{a}_i$ to $\vec{x}_{ik}$). This condition can be expressed as $\min_j\{x_{ijk}a_{ij}\} \geq 0$. It is violated only if a profile has at least one level ($x_{ijk} = 1$) that is unacceptable ($a_{ij} = -1$). Following Gilbride and Allenby (2004) we define a function to indicate when a profile is acceptable: $I(\vec{x}_{ik}, \vec{a}_i) = 1$ if $\min_j\{x_{ijk}a_{ij}\} \geq 0$, and $I(\vec{x}_{ik}, \vec{a}_i) = 0$ otherwise. We use the same coding for disjunctive rules, but modify the definition of $I(\vec{x}_{ik}, \vec{a}_i)$ to use $\max_j$ rather than $\min_j$.

Let y_{ik} be consumer i's answer to the k^{th} query, where $y_{ik} = 1$ if the consumer says "consider" and $y_{ik} = 0$ otherwise. Let $\vec{y}_{iK}$ be the vector of the first K answers. If there were no response errors we would observe $y_{ik} = 1$ if and only if $I(\vec{x}_{ik}, \vec{a}_i) = 1$. However, empirically, we expect response errors. Because the algorithm must run rapidly between queries, we choose a simple form for response error. Specifically, we assume that a consumer gives a false-positive answer with probability ϵ_1 and a false-negative answer with probability ϵ_2. For example, the i^{th} consumer will say "consider ($y_{ik} = 1$)" with probability $1 - \epsilon_2$ whenever the indicator function implies "consider," but he/she will also say "consider" with probability ϵ_1 if the indicator function implies "not consider." This error structure implies the following data-generating model:

(1)
$$\Pr(y_{ik} = 1 \mid \vec{x}_{ik}, \vec{a}_i) = (1 - \epsilon_2) I(\vec{x}_{ik}, \vec{a}_i) + \epsilon_1 \big(1 - I(\vec{x}_{ik}, \vec{a}_i)\big)$$
$$\Pr(y_{ik} = 0 \mid \vec{x}_{ik}, \vec{a}_i) = \epsilon_2 I(\vec{x}_{ik}, \vec{a}_i) + (1 - \epsilon_1)(1 - I(\vec{x}_{ik}, \vec{a}_i))$$

Each new query, $\vec{x}_{i,K+1}$ is based on our posterior beliefs about the decision rules ($\vec{a}_i$). After the K^{th} query, we compute the posterior $\Pr(\vec{a}_i | X_{iK}, y_{iK})$ conditioned on the first K queries (X_{iK}), the first K answers ($\vec{y}_{iK}$), and the priors. (Posterior beliefs might also reflect information from other respondents. See §6.6.) We seek to select the $\vec{x}_{ik}$'s to get as much information as feasible about $\vec{a}_i$, or, equivalently, to reduce uncertainty about $\vec{a}_i$ by the greatest amount. In §6.2 we define "information" and describe how we optimize it.

5.3. Error Magnitudes are Set Prior to Data Collection

We cannot know the error magnitudes until after data are collected, but we must set the ϵ's in order to collect data. Setting the ϵ's is analogous to setting "accuracy" parameters in aggregate customization. We address this conundrum in two ways. (1) We treat the ϵ's as "tuning" parameters and explore the sensitivity to these tuning parameters with synthetic data. Setting tuning parameters is common in machine-learning query-selection. (2) For the empirical test we rely on managerial judgment (Little 2004a, 2004b). Because the tuning parameters are set by managerial judgment prior to data collection, our empirical test is conservative in the sense that predictions might improve if future research allows updating of the error magnitudes within or across respondents.

To aid intuition we motivate the ϵ's with an illustrative micro-analysis of our stylized example. Suppose that a respondent's conjunctive heuristic is [Toyota $\wedge$ crossover]. This respondent should find a crossover Toyota acceptable and not care about the engine and trim. Coding each aspect as acceptable or not, and preserving the order Toyota, Chevy, sedan, crossover, hybrid, gasoline, premium trim, basic trim, this heuristic becomes $\vec{a}_i$ = [+1, -1, -1,+1,+1,+1,+1,+1]. Suppose when this respondent makes a consideration decision, he or she makes errors with probability η on each aspect, where an error involves flipping that aspect's acceptability. For example, suppose he or she is shown a Toyota crossover with a hybrid engine and premium trim, that is, $\vec{x}_{ik}$ = [1, 0, 0, 1, 1, 0, 1, 0]. He or she matches the heuristic to the profile aspect-by-aspect, making errors with probability η for each acceptable aspect in the profile. E.g., Toyota is acceptable per the heuristic, but may be mistaken for unacceptable with probability η. The respondent can make a false-negative error if any of the four aspects in the profile are mistaken for unacceptable ones. If these errors occur independently the respondent will make a false negative error with probability, $\epsilon_2 = 1 - (1 - \eta)^4$. If a profile is unacceptable, say $\vec{x}_{ik}$ = [0, 1, 1, 0, 1, 0, 1, 0], we easily compute $\epsilon_1 = \eta^2(1 - \eta)^2$.

In this illustration, any prior belief on the distribution of the heuristics and profiles implies expected ϵ's as a function of the η's. Whether one specifies the η's and derives expected ϵ's or specifies the ϵ's directly depends upon the researchers and managers, but in either case the tuning parameters are specified prior to data collection. With synthetic data we found no indication that one specification is preferred to the other. Empirically, we found it easier to think about the ϵ's directly.

6. Adaptive Question Selection

To select ask questions adaptively we must address four issues:

1. Initialize beliefs by generating consumer-specific priors.
2. Select the next query based on current posterior beliefs
3. Update posterior beliefs from the priors and the responses to all the previous questions.
4. Continuing looping Steps 2 and 3 until Q questions are asked (or until another stopping rule is reached).

6.1. Initialize Consumer-Specific Beliefs (Step 1)

Hauser and Wernerfelt (1990, 393) provide examples where self-stated consideration-set sizes are one-tenth or less of the number of brands on the market. Our experience suggests these examples are typical. If the question-selection algorithm used non-informative priors, the initial queries would be close to random guesses, most of which would not be considered by the consumer. When a consumer considers a profile we learn (subject to the errors) that all of its aspects are acceptable; when a consumer rejects a profile we learn only that one or more aspects are unacceptable. Therefore, the first considered profile provides substantial information and a significant shift in beliefs. Without observing the first considered profile directly, queries are not efficient, particularly with large numbers of aspects (N). To address this issue we ask each respondent to configure a considered profile and, hence, gain substantial information.

Prior research using compensatory rules (e.g., Toubia, et al. 2004), suggests that adaptive questions are most efficient relative to random or orthogonal questions when consumers' heuristic decision rules are heterogeneous. We expect similar results for non-compensatory heuristics. In the presence of heterogeneity, the initial configured profile enables us to tailor prior beliefs to each respondent.

For example, in our empirical application we tailor prior beliefs using the co-occurrence of brands in consideration sets. Such data are readily available in the automotive industry and for frequently purchased consumer goods. Alternatively, prior beliefs might be updated on the fly using a collaborative filter on prior respondents (see §6.6). Without loss of generality, let $j = 1$ index the brand aspect the respondent configures and, for other brand aspects, let b_{1j} be the prior probability that brand j is acceptable when brand 1 is acceptable. Let $\vec{x}_{i1}$ be the configured profile and set $y_{i1} = 1$. When co-occurrence data are available, prior beliefs on the marginal probabilities are set such that $\Pr(a_{i1} = 1|\vec{x}_{i1}, y_{i1}) = 1$ and $\Pr(a_{ij} = 1|\vec{x}_{i1}, y_{i1}, priors) = b_{ij}$.

Even without co-occurrence data we can set respondent-specific priors for every aspect on which we have strong prior beliefs. We use weakly informative priors for all other aspects. When managers have priors across features (e.g., considered hybrids are more likely to be Toyotas), we also incorporate those priors (Little 2004a; 2004b).

6.2. Select the Next Question Based on Posterior Beliefs from Prior Answers (Step 2)

The respondent's answer to the configurator provides the first of a series of estimates of his or her decision rule, $p_{ij1} = \Pr(a_{ij} = 1|X_{i1} = \vec{x}_{i1}, \vec{y}_{i1})$ for all aspects, j. (We have suppressed the notation for "priors.") We update these probabilities by iterating through Steps 2 & 3, computing updated estimates after each question-answer pair using all data collected up to and including that the K^{th} question, $p_{ijK} = \Pr(a_{ij} = 1|X_{iK}, \vec{y}_{iK})$ for $K > 1$. (Details in Step 3, §6.3.) To select the $K + 1^{st}$ query (Step 2) assume we have computed posterior values (p_{ijK}) from prior queries (up to K) and that we can compute contingent values ($p_{ij,K+1}$) one step ahead for any potential new query ($\vec{x}_{i,K+1}$) and its corresponding answer ($y_{i,K+1}$). We seek those questions that tell us as much as feasible about the respondent's decision heuristic. Equivalently, we seek to reduce uncertainty about $\vec{a}_i$ by the greatest amount.

Following Lindley (1956) we define the most informative question as the query that minimizes a loss function. In this paper, we use Shannon's entropy as the uncertainty measure (Shannon 1948), but other measures of uncertainty could be used without otherwise changing the algorithm. Shannon's entropy, measured in *bits*, quantifies the amount of information that is missing because the value of a random variable is not known for certain. Higher entropy corresponds to more uncertainty. Zero entropy corresponds to perfect knowledge. Shannon's entropy (hereafter, entropy) is used widely in machine-learning, has proven robust in many situations, and is the basis of criteria used to evaluate parameter recovery and predictive ability ($U^2$and Kullback-Leibler divergence). We leave to future implementations other loss functions such as Rényi entropy, suprisals, and other measures of information.[9] Mathematically:

$$H_{\vec{a}_i} = \sum_{j=1}^{N} -\{p_{ijK} \log_2 p_{ijK} + (1 - p_{ijK}) \log_2 (1 - p_{ijK})\} \tag{2}$$

If some aspects are more important to managerial strategy, we use a weighted sum in Equation 2.

To select the $K + 1^{st}$ query, $\vec{x}_{i,K+1}$, we enumerate candidate queries, anticipating the answer to the question, $y_{i,K+1}$, and anticipating how that answer updates our posterior beliefs about the respondent's heuristic. Using the p_{ijK}'s we compute the probability the respondent will consider the profile, $q_{i,K+1}(\vec{x}_{i,K+1}) = \Pr(y_{i,K+1} = 1 | X_{iK}, \vec{y}_{iK}, \vec{x}_{i,K+1})$. Using the Step 3 algorithm (described next) we update the posterior $p_{ij,K+1}$'s for all potential queries and answers. Let $p^{+}_{ij,K+1}(\vec{x}_{i,K+1}) = \Pr(a_{ij} = 1 | X_{iK}, \vec{y}_{iK}, \vec{x}_{i,K+1}, y_{i,K+1} = 1)$ be the posterior beliefs if we ask profile $\vec{x}_{i,K+1}$ and the respondent considers it. Let $p^{-}_{ij,K+1}(\vec{x}_{i,K+1}) =$
$\Pr(a_{ij} = 1 | X_{iK}, \vec{y}_{iK}, \vec{x}_{i,K+1}, y_{i,K+1} = -1)$ be the posterior beliefs if the respondent does not consider the profile. Then the expected posterior entropy is:

$$E[H_{\vec{a}_i}(\vec{x}_{i,K+1} | X_{iK}, \vec{y}_{iK})] = -q_{i,K+1}(\vec{x}_{i,K+1}) \Sigma_j \left\{ \begin{matrix} p^{+}_{ij,K+1}(\vec{x}_{i,K+1}) \log_2 [p^{+}_{ij,K+1}(\vec{x}_{i,K+1})] + \\ [1 - p^{+}_{ij,K+1}(\vec{x}_{i,K+1})] \log_2 [1 - p^{+}_{ij,K+1}(\vec{x}_{i,K+1})] \end{matrix} \right\}$$
$$-[1 - q_{i,K+1}(\vec{x}_{i,K+1})] \Sigma_j \left\{ \begin{matrix} p^{-}_{ij,K+1}(\vec{x}_{i,K+1}) \log_2 [p^{-}_{ij,K+1}(\vec{x}_{i,K+1})] + \\ [1 - p^{-}_{ij,K+1}(\vec{x}_{i,K+1})] \log_2 [1 - p^{-}_{ij,K+1}(\vec{x}_{i,K+1})] \end{matrix} \right\} \tag{3}$$

[9] Rényi's entropy reduces to Shannon's entropy when Rényi's $\alpha = 1$; the only value of α for which information on the a_{ij}'s is separable. To use these measures of entropy, modify Equations 2 and 3 to reflect Rényi's α.

When the number of feasible profiles is moderate we compute Equation 3 for every profile and choose the profile that minimizes Equation 3. However, in large designs such as the 53-aspect design in our empirical example, the number of potential queries (357,210) can be quite large. Because this large number of computations cannot be completed in less than a second, we focus our search using *uncertainty sampling* (e.g., Lewis and Gale, 1994). Specifically, we evaluate Equation 3 for the T queries about which we are most uncertain. "Most uncertain" is defined as $q_{i,K+1}(\vec{x}_{i,K+1}) \approx 0.5$. Profiles identified from among the T most-uncertain profiles are approximately optimal and, in some cases, optimal (e.g., Appendix 1). Uncertainty sampling is similar to choice balance as used in both polyhedral methods and aggregate customization (e.g., Arora and Huber 2001; Toubia, Hauser and Simester 2004). Synthetic data tests demonstrate that with T sufficiently large, we achieve close-to-optimal expected posterior entropy. For our empirical application setting $T = 1{,}000$ kept question selection under a second. As computing speeds improve researchers can use a larger T.

Equation 3 is myopic because it computes expected posterior entropy one step ahead. Extending the algorithm S steps ahead is feasible for small N. However S-step computations are exponential in S. For example, if there are 256 potential queries, a two-step ahead algorithm requires that we evaluate $256^2 = 65{,}536$ potential queries (without further approximations). Fortunately, synthetic data experiments suggest that one-step ahead computations achieve close to the theoretical maximum information of 1 bit per query (when there are no response errors) and do quite well when there are response errors. For completeness we coded a two-step-ahead algorithm in the case of 256 potential queries. Even for modest problems its running time was excessive (over 13 minutes between questions); it provided negligible improvements in parameter recovery. Our empirical application has over a thousand times as many potential queries—a two-step-ahead algorithm was not feasible computationally.

6.3. Update Beliefs About Heuristic Rules Based on Answers to the *K* Questions (Step 3)

In Step 3 we use Bayes' Theorem to update our beliefs after the K^{th} query:

$$(4) \qquad \Pr(\vec{a}_i | X_{iK}, \vec{y}_{iK}) \propto \Pr(y_{iK} | \vec{x}_{iK}, \vec{a}_i = \vec{a}) \Pr(\vec{a} = \vec{a}_i | X_{i,K-1}, \vec{y}_{i,K-1}).$$

The likelihood term, $\Pr(y_{iK} | \vec{x}_{iK}, \vec{a}_i = \vec{a})$, comes from the data-generating model in Equation 1. The variable of interest, $\vec{a}_i$, is defined over all binary vectors of length N. Because the number of potential conjunctions is exponential in N, updating is not computationally feasible without further structure on the distribution of conjunctions. For example, with $N = 53$ in our empirical example, we would need to update the distribution for 9.0 x 10^{15} potential conjunctions.

To gain insight for a feasible algorithm we examine solutions to related problems. Gilbride and Allenby (2004) use a "Griddy Gibbs" algorithm to sample threshold levels for features. At the consumer level, the thresholds are drawn from a multinomial distribution. The Griddy Gibbs uses a grid approximation to the (often univariate) conditional posterior. We cannot modify their solution directly, in part because most of our features are horizontal (e.g., brand) and thresholds do not apply. Even for vertical features, such as price, we want to allow non-threshold heuristics. We need algorithms that let us classify each level as acceptable or not.

For a feasible algorithm we use a variational Bayes approach. In variational Bayes inference, a complex posterior distribution is approximated with a variational distribution chosen from a family of distributions judged similar to the true posterior distribution. Ideally, the variational family can be evaluated quickly (Attias 1999, Ghahramani and Beal 2000). Even with an uncertainty-sampling approximation in Step 2 we must compute posterior distributions for $2T$ question-answer combinations and do so while the respondent waits for the next question.

As our variational distribution we approximate the distribution of $\vec{a}_i$ with N independent binomial distributions. This variational distribution has N parameters, the p_{ij}'s, rather than parameters for the 2^N potential values of $\vec{a}_i$. Because this variational approximation is within a consumer, we place no restriction on the empirical population distribution of the a_{ij}'s. Intercorrelation at the population level is likely (and allowed) among aspect probabilities. For example, we might find that those automotive consumers who screen on Toyota also screen on hybrid engines. In another application we might find that those cellular phone consumers who screen on Nokia also screen on "flip." For every respondent the posterior values of all p_{ijK}'s depend upon all of the data from that respondent, not just queries that involve the j^{th} aspect.

To calculate posteriors for the variational distribution we use a version of belief propagation (Yedidia, Freeman and Weiss 2003; Ghahramani and Beal 2001). The algorithm converges iteratively to an estimate of $\vec{p}_{iK}$. The h^{th} iteration uses Bayes Theorem to update each p_{ijK}^h based on the data and based on $p_{ij'K}^h$ for all $j' \neq j$. Within the h^{th} iteration the algorithm loops over aspects and queries using the data-generating model (Equation 1) to compute the likelihood of observing $y_k = 1$ conditioned on the likelihood for $k' \neq k$. It continues until the estimates of the p_{ijK}^{h}'s stabilize. In our experience, the algorithm converges quickly: 95.6% of the estimations converge in 20 or fewer iterations, 99.2% in 40 or fewer iterations, and 99.7% in 60 or fewer iterations. Appendix 2 provides the pseudo-code.

While variational distributions work well in a variety of applications, there is no guarantee for our application. Performance is an empirical question which we address in §7 and §8. Finally, we note that the belief propagation algorithm and Equation 4 appear to be explicitly dependent only upon the questions that are answered by consumer i. However, our notation has suppressed the dependence on prior beliefs. It is a simple matter to make prior beliefs dependent upon the distribution of the $\vec{a}_i$'s as estimated from previous respondents. See §6.6.

6.4. Stopping Rules (Step 4)

Adaptive-question selection algorithms for compensatory decision rules and fixed question-selection algorithms for compensatory or non-compensatory rules rely on a target number of questions chosen by prior experience or judgment. Such a stopping rule can be used with the adaptive question-selection algorithm proposed in this paper. For example, we stopped after $Q = 29$ questions in our empirical illustration.

However, expected posterior entropy minimization makes it feasible to select a stopping rule endogenously. One possibility is to stop questioning when the expected reduction in entropy drops below a threshold for two or more adaptive questions. Synthetic data provide some insight. In §7 we plot the information obtained about parameters as a function of the number of questions. In theory we might also gain insight from our empirical example. However, because our empirical example used only 29 questions for 53 aspects, for 99% of the respondents the adaptive-question selection algorithm would still have gained substantial information if the respondents had been asked a 30th question. We return to this issue in §11. Because we cannot redo our empirical example, we leave this and other stopping-rule extensions to future research.

6.5. Extension to Disjunctions of Conjunctions

Disjunctions-of-conjunctions heuristics nest both simple and complex heuristics. The extension to disjunctions-of-conjunctions is conceptually simple. After we reach a stopping rule, whether it be fixed a priori or endogenous, we simply restart the algorithm by asking a second configurator question, but requiring an answer that is substantially different from the profiles that the respondent has already indicated he or she will consider. If the respondent cannot configure such a profile, we stop. Empirically, cognitive simplicity suggests that respondents use a relatively few conjunctions (e.g., Gigerenzer and Goldstein 1996; Hauser, et al. 2010; Martignon and Hoffrage 2002). Most consumers use one conjunction (Hauser, et al. 2010). Hence the number of questions should remain within reason. We test this procedure on synthetic data and, to the extent that our data allow, empirically.

6.6. Using Data from Previous Respondents

We can use data from other respondents to improve priors for new respondents, but in doing so, we want to retain the advantage of consumer-specific priors. Collaborative filtering provides a feasible method (e.g., Breese, Heckerman, and Kadie 1998). We base our collaborative filter on the consumer-specific data available from the configurator (Figure 2a.)

Specifically, after a new respondent completes the configurator, we use collaboratively-filtered data from previous respondents who configured similar profiles. For example, if an automotive consumer configures a Chevy, we search for previous respondents who configured a Chevy. For other brands we compute priors with a weighted average of the brand posteriors (p_{ij}'s) from those respondents. (We weigh previous respondents by predictive precision.) We do this for all configured features. As sample sizes increase, population data overwhelms even "bad" priors; performance will converge to performance based on accurate priors (assuming the collaborative filter is effective). We test finite-sample properties on synthetic data and, empirically, with an approximation based on the data we collected.

7. Synthetic Data Experiments

To evaluate the ability of the active-learning algorithm to recover known heuristic decision rules we use synthetic respondents. To compare adaptive question selection to established methods, we choose a synthetic decision task with sufficiently many aspects to challenge the algorithm but for which existing methods are feasible. With four features at four levels (16 aspects) there are 65,536 heuristic rules—a challenging problem for extant heuristic-rule estimation methods. An orthogonal design is 32 profiles and, hence, in the range of tasks in the empirical literature. We simulate respondents who answer any number of questions $K \in [1, 256]$ where 256 profiles exhausts the feasible profiles. To evaluate the question-selection methods we randomly select 1,000 heuristic rules (synthetic respondents). For each aspect we draw a Bernoulli probability from a $Beta(1, 1)$ distribution (uniform distribution) and draw a +1 or -1 using the Bernoulli probability. This "sample size" is on the high side of what we might expect in an empirical study and provides sufficient heterogeneity in heuristic rules.

For each decision heuristic, $\vec{a}_i$, we use either the proposed algorithm or an established method to select questions. The synthetic respondent then "answers" the questions using the decision heuristic, but with response errors, ϵ_1 and ϵ_2, chosen as if generated by reasonable η's. To compare question-selection methods, we keep the estimation method constant. We use the variational-Bayes-belief-propagation method developed in this paper. The benchmark question-selection methods are orthogonal, random, and market-based. Market-based questions are chosen randomly but in proportion to profile shares we might expect in the market—market shares are known for synthetic data.

With synthetic data we know the parameters, a_{ij}. For any K and for all i and j, we use the "observed" synthetic data to update the probability, p_{ijK}, that $a_{ij} = 1$. An appropriate information-theoretic measure of parameter recovery is U^2 which quantifies the percent of uncertainty explained (empirical information/initial entropy, Hauser 1978). U^2= 100% indicates perfect parameter recovery.

We begin with synthetic-data which contain no response error. These data quantify potential maximum gains with adaptive questions, test how rapidly active-learning questions recover parameters perfectly, and bound improvements that would be possible with non-myopic S-step-ahead algorithms. We then repeat the experiments with error-laden synthetic data and with "bad" priors. Finally, we examine whether we can recover disjunctions-of-conjunctions heuristics and whether population-based priors improve predictions.

7.1. Tests of Upper Bounds on Parameter Recovery (no response errors)

Figure 3 presents key results. To simplify interpretation we plot random queries in Appendix 4, rather than Figure 3, because the results are indistinguishable from market-based queries on the scale of Figure 3. Market-based queries do about 3% better than random queries for the first 16 queries, about 1% better for the first 32 queries, and about ½% better for all 256 queries. (Queries 129 through 256 are not shown in Figure 3; the random-query and the market-based query curves asymptote to 100%.) Orthogonal questions are only defined for $K = 32$.

Questions selected adaptively by the active-learning algorithm find respondents' decision heuristics much more rapidly than existing question-selection methods. The adaptive questions come very close to an optimal reduction in posterior entropy. With 16 aspects and equally-likely priors, the prior entropy is 16 $\log_2(2)$, which is 16 bits. The configurator reveals four acceptable aspects (4 bits). Each subsequent query is a binary outcome that can reveal at most 1 bit. A perfect active-learning algorithm would require 12 additional queries to identify a decision rule (4 bits + 12 bits identifies the 16 elements of $\vec{a}_i$). On average, in the absence of response error, the adaptive questions identify the respondent's decision heuristic in approximately 13 questions. The variational approximation and the one-step-ahead question selection algorithm appear to achieve close to optimal information (12 bits in 13 questions).

[Insert Figure 3 about here.]

We compare the relative improvement due to question-selection methods by holding information constant and examining how many questions it takes to achieve that level of parameter recovery. Because an orthogonal design is fixed at 32 questions, we use it as a benchmark. As illustrated in the first line of data in Table 1, an orthogonal design requires 32 queries to achieve a U^2 of approximately 76%. Market-based questions require 38 queries; random questions require 40 queries, and adaptive questions only 9 queries. To parse the configurator from the adaptive questions, Appendix 4 plots the U^2 obtained with a configurator plus market-based questions. The plot parallels the plot of purely market-based queries requiring 30 queries to achieve a U^2 of approximately 76%. In summary, in an errorless world, the active-learning algorithm chooses adaptive questions which provide substantially more information per question than existing non-adaptive methods. The large improvements in U^2, even for small numbers of questions, suggests that adaptive questions are chosen to provide information efficiently.

[Insert Table 1 about here.]

7.2. Tests of Parameter Recovery When There are Response Errors or "Bad" Priors

We now add either response error or "bad" priors and repeat the synthetic-data experiments. The plots remain quasi-concave for a variety of levels of response error and/or bad priors.[10] We report representative values in Table 1. (Table 1 is based on false negatives occurring 5% of the time. False positives are set by the corresponding η. "Bad" priors perturb "good" priors with bias drawn from $U[0,0.1]$.) Naturally, as we add error or bad priors the amount of information obtained per question decreases, for example, 13 adaptive questions achieved a U^2 of 100% without response errors, but only 55.5% with response errors. On

[10] Although the plots in Figure 2 are concave, there is no guarantee that the plots remain concave for all situations. However, we do expect all plots to be quasi-concave. They are.

average it takes 12.4 adaptive questions to obtain a U^2 of 50% (standard deviation 8.7). The last column of Table 1 reports the information obtained by 32 orthogonal questions. Adaptive questions obtain relatively more information per question than existing methods under all scenarios. Indeed, adaptive questions appear to be more robust to bad priors than existing question-selection methods.

7.3. Tests of the Ability to Recovery Disjunctions-of-Conjunctions Heuristics

We now generate synthetic data for respondents who have two distinct conjunctions rather than just one conjunction. By distinct, we mean no overlap in the conjunctions. We allow both question-selection methods to allocate one-half of their questions to the first conjunction and one-half to the second conjunction. To make the comparison fair, all question-selection methods use data from the two configurators when estimating the parameters of the disjunctions-of-conjunctions heuristics. After 32 questions (plus two configurators), estimates based on adaptive questions achieve a U^2 of 80.0% while random questions achieve a U^2 of only 34.5%. Adaptive questions also beat market-based and orthogonal questions handily.

This is an important result. With random questions false positives from the second conjunction pollute the estimation of the parameters of the first conjunction and vice versa. The active-learning algorithm focuses questions on one or the other conjunction to provide good recovery of the parameters of both conjunctions. We expect the two-conjunction results to extend readily to more than two conjunctions. While this initial test is promising, future tests might improve the algorithm with endogenous stopping rules that allocate questions optimally among conjunctions.

7.4. Tests of Incorporating Data from Previous Respondents

To demonstrate the value of incorporating data from other respondents, we split the sample of synthetic respondents into two halves. For the first half of the sample, we use bad priors, ask questions adaptively, and estimate the $\vec{a}_i$'s. We use the estimated $\vec{a}_i$'s and a collaborative filter on two features to customize priors for the remaining respondents. We then ask questions of the remaining respondents using collaborative-filter-based priors. On average, U^2 is 17.8% larger on the second set of respondents (using collaborative-filter-based priors) than on the first set of respondents (not using collaborative-filter-based priors). Thus, even when we use bad priors for early respondents, the posteriors from those respondents are sufficient for the collaborative filter. The collaborative-filter-based priors improve U^2 for the remaining respondents.

7.5. Summary of Synthetic Data Experiments

The synthetic-data experiments suggest that:

- adaptive question selection via active learning is feasible and can recover the parameters of known heuristic decision rules
- adaptive question selection provides more information per question than existing methods
- one-step-ahead active-learning adaptive questions achieve gains in information (reduction in entropy) that are close to the theoretical maximum when there are no response errors
- adaptive question selection provides more information per question when there are response errors
- adaptive question selection provides more information per question when there are badly chosen priors
- it is feasible to extend adaptive-question selection to disjunctions-of-conjunctions heuristic decision rules
- incorporating data from other respondents improves parameter recovery.

These synthetic-data experiments establish that, if respondents use heuristic decision rules, then the active-learning algorithm provides a means to ask questions that provide substantially more information per question.

8. Illustrative Empirical Application with a Large Number of Aspects

In the spring of 2009 a large American automotive manufacturer (AAM) recognized that consideration of their vehicles was well below that of non-US vehicles. Management was interested in exploring various means to increase consideration. As part of that effort, AAM fielded a web-based survey to 2,336 respondents recruited and balanced demographically from an automotive panel maintained by Harris Interactive, Inc. Respondents were screened to be 18 years of age and interested in purchasing a new vehicle in the next two years. Respondents received 300 Harris points (good for prizes) as compensation for completing a 40 minute survey. The response rate was 68.2% and the completion rate was 94.9%.

The bulk of AAM's survey explored various marketing strategies that AAM might use to enhance consideration of their brands. The managerial test of communications strategies is tangential to the scope and focus of this paper, but we illustrate in §10 the types of insight provided by estimating consumers' non-compensatory heuristics.

Because AAM's managerial decisions depended upon the accuracy with which they could evaluate their communications strategies, we were given the opportunity to test adaptive question selection for a subset of the respondents. A subset of 872 respondents was not shown any communications inductions. Instead, after configuring a profile, evaluating 29 calibration profiles, and completing a memory-cleansing task (Frederick 2005), respondents evaluated a second set of 29 validation profiles. (A 30th profile in calibration and validation was used for other research purposes by AAM.) The profiles varied on 53 aspects: brand (21 aspects), body style (9 aspects), price (7 aspects), engine power (3 aspects), engine type (2 aspects), fuel efficiency (5 aspects), quality (3 aspects), and crash-test safety (3 aspects).

8.1. Adaptive Question Selection for Calibration Profiles

To test adaptive question selection, one-half of the calibration profiles were chosen adaptively by the active-learning algorithm. The other half were chosen randomly in proportion to market share from the top 50 best-selling vehicles in the US. To avoid order effects and to introduce variation in the data, the question-selection methods were randomized. This probabilistic variation means that the number of queries of each type is 14.5 on average, but varies by respondent.

As a benchmark we chose market-based queries rather than random queries. The market-based queries perform slightly better on synthetic data than purely-random queries and, hence, provide a stronger test. We could not test an orthogonal design because 29 queries is but a small fraction of the 13,320 profiles in a 53-aspect orthogonal design. (A full factorial would require 357,210 profiles.) Furthermore, even if we were to complete an orthogonal design of 13,320 queries, Figure 2 suggests that orthogonal queries do only slightly better than random or market-based queries. Following Sándor and Wedel (2002) and Vriens, Wedel and Sándor (2001) we split the market-based profiles (randomly) over respondents.

Besides enabling methodological comparisons, this mix of adaptive and market-based queries has practical advantages with human respondents. First, the market-based queries introduce variety to engage the respondent and help disguise the choice-balance nature of the active-learning algorithm. (Respondents get variety in the profiles they evaluate.) Second, market-based queries sample "far away" from the adaptive queries chosen by the active learning algorithm. They might prevent the algorithm from getting stuck in a local maximum (an analogy to simulated annealing).

8.2. Selecting Priors for the Empirical Application

AAM had co-occurrence data available from prior research, so we set priors as described in §6.1. In addition, using AAM's data and managerial beliefs, we were able to set priors on some pairwise conjunctions such as "Porsche ∧ Kia" and "Porsche ∧ pick-up." Rather than setting these priors directly as correlations among the a_{ij}'s, AAM's managers found it more intuitive to generate "pseudo-questions" in which the respondent was assumed to "not consider" a "Porsche ∧ pick-up" with probability q where q was set by managerial judgment. In other applications researchers might set the priors directly.

8.3. Validation Profiles Used to Evaluate Predictive Ability

After a memory-cleansing task, respondents were shown a second set of 29 profiles, this time chosen by the market-based question-selection method. Because there was some overlap between the market-based validation and the market-based calibration profiles, we have an indicator of respondent reliability. Respondents consistently evaluated market-based profiles 90.5% of the time. Respondents are consistent, but not perfect, and, thus, modeling response error (via the ϵ's) appears to be appropriate.

8.4. Performance Measures

While hit rate is an intuitive measure it can mislead intuition for consideration data. If a respondent were to consider 20% of both calibration and validation profiles, then a null model that predicts "reject all profiles" will achieve a hit rate of 80%. But such a null model provides no information, has a large number of false negative predictions, and predicts a consideration-set size of zero. On the other hand, a null model that predicts randomly proportional to the consideration-set size in the calibration data, would predict a larger validation consideration-set size and balance false positives and false negatives, but would achieve a lower hit rate (68%: $0.68 = (0.8)^2 + (0.2)^2$). Nonetheless, for interested readers, Appendix 5 provides hit rates.

We expand evaluative criteria by examining false positive and false negative predictions. A manager might put more (or less) weight on not missing considered profiles than on predicting as considered profiles that are not considered. However, without knowing specific loss functions to weigh false positives and false negatives differently, we cannot have a single managerial criterion (e.g., Toubia and Hauser 2007). Fortunately, information theory provides a commonly-used measure that balances false positives and false negatives: the Kullback-Leibler divergence (KL). KL is a non-symmetric measure of the difference from a prediction model to a comparison model (Chaloner and Verdinelli 1995; Kullback and Leibler 1951). It discriminates among models even when the hit rates might otherwise be equal. Appendix 3 provides formulae for the KL measure appropriate to the data in this paper. We calculate divergence from perfect prediction, hence a smaller KL is better.

In synthetic data we knew the "true" decision rule and could compare the estimated parameters, a_{ij}'s, to known parameters. $U^2$was the appropriate measure. With empirical data we do not know the true decision rule; we only observe the respondents' judgments about consider vs. not consider, hence KL is an appropriate measure. However, both attempt to quantify the information explained by the estimated parameters (decision heuristics).

8.5. Key Empirical Results

Table 2 summarizes KL divergence for the two question-selection methods that we tested: adaptive questions and market-based questions. For each question type, we use two estimation methods: (1) the variational-Bayes belief-propagation algorithm computes the posterior distribution of the non-compensatory heuristics and (2) a hierarchical Bayes logit model (HB) computes the posterior distribution for a compensatory model. HB is the most-used estimation method for additive utility models (Sawtooth 2004), and it has proven accurate for 0-vs.-1 consideration decisions (Ding, et al. 2011; Hauser, et al. 2010). The latter authors provide a full HB specification in online appendix 3. Both estimation methods are based only on the calibration data. For comparison Table 2 also reports predictions for null models that predict all profiles as considered, predict no profiles as considered, and predict profiles randomly based on the consideration-set size among the calibration profiles.

When the estimation method assumes respondents use heuristic decision rules, rules estimated from adaptive questions predict significantly better than rules estimated from market-based queries. (Hit rates are also significantly better.) Furthermore, for adaptive questions, heuristic rules predict significantly better than HB-estimated additive rules. Although HB-estimated additive models nest lexicographic models (and hence conjunctive models for consideration data), the required ratio of partworths is approximately 10^{15} and not realistic empirically. More likely HB does less well because its assumed additive model with 53 parameters over-fits the data, even with shrinkage to the population mean.

It is, perhaps, surprising that ~14.5 adaptive questions do so well for 53 aspects. This is an empirical issue, but we speculate that the underlying reasons are (1) consumers use cognitively-simple heuristics with relatively few aspects, (2) the adaptive questions search the space of decision rules efficiently to confirm the cognitively-simple rules, (3) the configurator focuses this search quickly, and (4) consumer-specific priors keep the search focused.

There is an interesting, but not surprising, interaction effect in Table 2. If the estimation assumes an additive model, non-compensatory-focused adaptive questions do not do as well as market-based questions. Also, consistent with prior research using non-adaptive questions (e.g., Dieckmann, et al. 2009; Kohli and Jedidi 2007; Yee, et al. 2007), non-compensatory estimation is comparable to compensatory estimation using market-based questions. Perhaps to truly identify heuristics, we need heuristic-focused adaptive questions.

But are consumers compensatory or non-compensatory? The adaptive-question-non-compensatory-estimation combination is significantly better than all other combinations in Table 2. But what if we estimated both non-compensatory and compensatory models using all 29 questions (combining ~14.5 adaptive questions and ~14.5 market-based questions)? The non-compensatory model predicts significantly better than the compensatory model when all 29 questions are used (KL = 0.451 vs. KL = 0.560, $p < 0.001$ using a paired t-test). Differences are also significant at $p < 0.001$ using a related-samples Wilcoxon signed rank test. Because we may not know a priori whether the respondent is non-compensatory or compensatory, collecting data both ways gives us flexibility for post-data-collection re-estimation. (In the automotive illustration prior theory suggested that consumers were likely to use non-compensatory heuristics.)

8.6. Summary of Empirical Illustration

Adaptive questions to identify non-compensatory heuristics are promising. We appear to be able to select questions to provide significantly more information per query than market-based queries. Furthermore, it appears that questions are chosen efficiently because we can predict well with ~14.5 questions even in a complex product category with 53 aspects. This is an indication of cognitively simplicity. Finally, consumers appear to be non-compensatory.

9. Initial Tests of Generalizations: Disjunctions of Conjunctions and Population Priors

Although data were collected based on the conjunctive active-learning algorithm, we undertake exploratory empirical tests of two proposed generalizations: disjunctions of conjunctions and prior-respondent-based priors. These exploratory tests complement the theory in §6.5 and §6.6 and the synthetic-data tests in §7.3 and §7.4.

9.1. Disjunctions of Conjunctions (DOC)

AAM managers sought to focus on consumers' primary conjunctions because, in prior studies sponsored by AAM, 93% of the respondents used only one conjunction (Hauser, et al. 2009). However, we might gain additional predictive power by searching for second and subsequent conjunctions using the methods of §6.5. Ideally this requires new data, but we get an indicator by (1) estimating the best model with the data, (2) eliminating all calibration profiles that were correctly classified with the first conjunction, and (3) using the remaining market-based profiles to search for a second conjunction. As expected, this strategy reduced false negatives because there were more conjunctions. It came at the expense of a slight an increase in false positives. Overall, using all 29 questions, KL increased slightly (0.459 vs. 0.452, $p < 0.001$) suggesting that the re-estimation on incorrectly classified profiles over-fit the data. Because the DOC generalization works for synthetic data, a true test awaits new empirical data.

9.2. Previous-Respondent-Based Priors

The priors used to initialize consumer-specific beliefs were based on judgments by AAM managers and analysts, however, we might also use the methods proposed in §6.6 to improve priors based on data from other respondents. As a test, we used the basic algorithm to estimate the p_{ijK}'s, used the collaborative filter to reset the priors for each respondent, re-estimated the model (p'_{ijK}'s), and compared predicted consideration to observed consideration. Previous-respondent-based priors improved predictions, but not significantly (0.448 vs. 0.452, $p = 0.082$), suggesting that AAM provided good priors for this application.

10. Managerial Use

The validation reported in this paper was part of a much larger effort by AAM to identify communications strategies that would encourage consumers to consider AAM vehicles. At the time of the study, two of the three American manufacturers had entered bankruptcy. AAM's top management believed that overcoming consumers' unwillingness to consider AAM vehicles was critical if AAM was to become profitable. Table 2, combined with ongoing studies by AAM, was deemed sufficient evidence for managers to rely on the algorithm to identify consumers' heuristic decision rules. AAM is convinced of the relevancy of consumer heuristics and is actively investigating how to use non-compensatory data routinely to inform management decisions. We summarize here AAM's initial use of information on consumer heuristics.

The remaining 1,464 respondents each answered 29 adaptive plus market-based questions, were shown an experimental induction, and then answered a second set of 29 adaptive plus market-based questions. Each induction was a communications strategy targeted to influence consumers to (1) consider AAM vehicles or (2) consider vehicles with aspects on which AAM excelled. Details are proprietary and beyond the scope of this paper. However, in general, the most-effective communications strategies were those that surprised consumers with AAM's success in a non-US reference group. AAM's then-current emphasis on J.D. Power and *Consumer Reports* ratings did not change consumers' decision heuristics.

AAM used the data on decision heuristics for product development. AAM recognized heterogeneity in heuristics and identified clusters of consumers who share decision heuristics. There were four main clusters, high selectivity on brand and body type, selectivity on brand, selectivity on body type, and (likely) compensatory. There were 2-6 sub-clusters within each main cluster for a total of 20 clusters.[11] Each sub-cluster was linked to demographic and other decision variables to suggest directed communications and product-development strategies. Decision rules for targeted consumer segments are proprietary, but the population averages are not. Table 3 indicates which percent of the population uses elimination rules for each of the measured aspects.

[Insert Table 3 about here.]

While some brands were eliminated by most consumers, larger manufacturers have many targeted brands. For example, Buick was eliminated by 97% of the consumers and Lincoln by 98%, but these are not the only GM and Ford brands. For AAM, the net consideration of its brands was within the range of more-aggregate studies. Consumers are mixed on their interest in "green" technology: 44% eliminate hybrids from consideration, but 69% also eliminate large engines. Price elimination illustrates that heuristics are screening criteria, not surrogates for utility: 77% of consumers will not investigate a $12,000 vehicle. This means that consumers' knowledge of the market tells them that, net of search costs, their best strategy is to avoid

[11] AAM used standard clustering methods on the posterior p_{ij}'s. By the likelihood principle, it is possible to use latent-structure models to reanalyze the data. Post hoc clustering is likely to lead to more clusters than latent-structure modeling. Comparisons of clustering methods are beyond the scope and tangential to our current focus on methods to select questions efficiently for the estimation of heuristic decision rules.

investing time and effort to evaluate $12,000 vehicles. It does not mean that consumers would not buy a top-of-the-line Lexus if it were offered for $12,000. Table 3 provides aggregate summaries across many consumer segments—AAM's product development and communications strategies were targeted within segment. For example, 84% of consumers overall eliminate sports cars indicating the sports-car segment is a relatively small market. However, the remaining 16% of consumers constitute a market that is sufficiently large for AAM to target vehicles for that market.

11. Summary and Challenges

We found active machine learning to be an effective methodology to select questions adaptively in order to identify consideration heuristics. Both the synthetic data experiments and the proof-of-concept empirical illustration are promising, but many challenges remain.

Question selection might be improved further with experience in choosing "tuning" parameters (ϵ's, T), improved priors, an improved focus on more-complex heuristics, and better variational-Bayes belief-propagation approximations. In addition, further experience will provide insight on the information gained as the algorithm learns. For example, Figure 4 plots the average expected reduction in entropy for adaptive questions and for market-based questions. We see that, on average, adaptive questions provide substantially more information per question (5.5 times as much). Prior to the 10^{th} question the increasingly accurate posterior probabilities enable the algorithm to ask increasingly more accurate questions. Beyond 10 questions the expected reduction in entropy decreases and continues to decrease through the 29^{th} question. It is likely that AAM would have been better able to identify consumers' conjunctive decision rules had they used 58 questions for estimation rather than split the questions between calibration and validation. Research might explore the mix between adaptive and market-based questions.

[Insert Figure 4 about here.]

The likelihood principle implies that other models can be tested on AAM's and other adaptive data. The variational-Bayes belief-propagation algorithm does not estimate standard errors for the p_{ij}'s. Other Bayesian methods might specify more complex distributions. Re-estimation or bootstrapping, when feasible, might improve estimation.

Active machine learning might also be extended to other data-collection formats including formats in which multiple profiles are shown on the same page or formats in which configurators are used in creative ways. The challenge for large N is that we would like to approximate decision rules in less than N queries per respondent.

References

Arora, Neeraj and Joel Huber (2001), "Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments," *Journal of Consumer Research*, 28, (September), 273-283.

Attias, Hagai (1999), "Inferring Parameters and Structure of Latent Variable Models by Variational Bayes," Kathryn B. Laskey and Henri Prade, eds., *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (UAI-99), 21-30.

Breese, John S., David Heckerman, and Carl Kadie (1998), "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence* (UAI-98), (Morgan Kaufmann: San Francisco, CA)43-52.

Chaloner, Kathryn and Isabella Verdinelli (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10, 3, 273-304.

Dahan, Ely and John R. Hauser (2002), "The Virtual Customer," *Journal of Product Innovation Management*, 19, 5, (September), 332-354.

Dieckmann, Anja, Katrin Dippold and Holger Dietrich (2009), "Compensatory versus Noncompensatory Models for Predicting Consumer Preferences," *Judgment and Decision Making*, 4, 3, (April), 200-213.

Ding, Min, John R. Hauser, Songting Dong, Daria Dzyabura, Zhilin Yang, Chenting Su and Steven Gaskin (2011), "Unstructured Direct Elicitation of Decision Rules," *Journal of Marketing Research*, 48, (February), 116-127.

Erdem, Tülin and Joffre Swait (2004), "Brand Credibility, Brand Consideration, and Choice," *Journal of Consumer Research*, 31, (June), 191-98.

Evgeniou, Theodoros, Constantinos Boussios and Giorgos Zacharia (2005), "Generalized Robust Conjoint Estimation," *Marketing Science*, 24, 3, 415-429.

Frederick, Shane (2005), "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19, 4, 25-42.

Ghahramani, Zoubin and Beal, Matthew J (2000), "Variational Inference for Bayesian Mixtures of Factor Analyzers." *Advances in Neural Information Processing Systems*, 12, (Cambridge, MA: MIT Press).

------ and ------ (2001), "Propagation Algorithms for Variational Bayesian Learning," *Advances* *in Neural Information Processing Systems*, 13, (Cambridge, MA: MIT Press).

Gigerenzer, Gerd and Daniel G. Goldstein (1996), "Reasoning the Fast and Frugal Way: Models of Bounded Rationality," *Psychological Review*, 103, 4, 650-669.

------, Peter M. Todd and the ABC Research Group (1999), *Simple Heuristics That Make Us Smart*, (Oxford, UK: Oxford University Press).

Gilbride, Timothy J. and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23, 3, 391-406.

------ and ------ (2006), "Estimating Heterogeneous EBA and Economic Screening Rule Choice Models," *Marketing Science*, 25, 5, (September-October), 494-509.

Green, Paul E., (1984), "Hybrid Models for Conjoint Analysis: An Expository Review," *Journal of Marketing Research*, 21, 2, (May), 155-169.

Green, Paul E., Abba M. Krieger and Pradeep Bansal (1988), "Completely Unacceptable Levels in Conjoint Analysis: A Cautionary Note," *Journal of Marketing Research*, 25, (August), 293-300.

------ and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments With Implications for Research and Practice," *Journal of Marketing*, 54, 4, (October), 3-19.

Hauser, John R. (1978), "Testing the Accuracy, Usefulness and Significance of Probabilistic Models: An Information Theoretic Approach," *Operations Research*, 26, 3, (May-June), 406-421.

------, Olivier Toubia, Theodoros Evgeniou, Rene Befurt and Daria Dzyabura (2009), "Cognitive Simplicity and Consideration Sets," *Journal of Marketing Research*, 47, (June), 485-496.

------ and Birger Wernerfelt (1990), "An Evaluation Cost Model of Consideration Sets," *Journal of Consumer Research*, 16, (March), 393-408.

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

Kullback, Solomon, and Leibler, Richard A. (1951), "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79-86.

Lewis, David D. and William A Gale (1994), "Training Text Classifiers by Uncertainty Sampling," *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, 3-12.

Liberali, Guilherme, Glen L. Urban, and John R. Hauser (2011), "When Does Providing Competitive Information to Your Own Customers Increase Sales?," (Cambridge, MA: MIT Sloan School of Management).

Little, John D. C. (2004a), "Managers and Models: The Concept of a Decision Calculus," *Management Science*, 50, (12, Supplement), 1841-1853. Reprinted from 1970.

------ (2004b), "Comments on 'Models and Managers: The Concept of a Decision Calculus:' Managerial Models for Practice" *Management Science*, 50, (12, Supplement), 1854-1860.

Liu, Qing and Neeraj Arora (2011), "Efficient Choice Designs for a Consider-then-Choose Model," *Marketing Science*, 30, 2, (March-April), 321-338.

Martignon, Laura and Ulrich Hoffrage (2002), "Fast, Frugal, and Fit: Simple Heuristics for Paired Comparisons," *Theory and Decision*, 52, 29-71.

Payne, John W. , James R. Bettman and Eric J. Johnson (1988), "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 3, 534-552.

------, ------ and ------ (1993), *The Adaptive Decision Maker*, (Cambridge UK: Cambridge University Press).

Sándor, Zsolt and Michel Wedel (2002), "Profile Construction in Experimental Choice Designs for Mixed Logit Models," *Marketing Science*, 21, 4, (Fall), 455-475.

Sawtooth Software, Inc. (1996), "ACA System: Adaptive Conjoint Analysis," ACA Manual, (Sequim, WA: Sawtooth Software, Inc.)

------ (2004), "CBC Hierarchical Bayes Analysis Technical Paper," (Sequim WA; Sawtooth Software, Inc.)

------ (2008), "ACBC Technical Paper," (Sequim WA; Sawtooth Software, Inc.)

Shannon, C.E. (1948), "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27, (July & October) 379–423 & 623–656.

Toubia, Olivier and John R. Hauser (2007), "On Managerial Efficient Designs," *Marketing Science*, 26, 6, (November-December), 851-858.

------, ------, and Rosanna Garcia (2007), "Probabilistic Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis: Theory and Application," *Marketing Science*, 26, 5, (September-October), 596-610.

------, ------, and Duncan I. Simester (2004), "Polyhedral Methods for Adaptive Choice-based

Conjoint Analysis," *Journal of Marketing Research*, 41, (February), 116-131.

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

Urban, Glen L. and John R. Hauser (2004), "'Listening-In' to Find and Explore New Combinations of Customer Needs," *Journal of Marketing*, 68, (April), 72-87.

van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel, and Philip Hans Franses (2010), "Retrieving Unobserved Consideration Sets from Household Panel Data," *Journal of Marketing Research*, 47, (February), 63-74.

Vriens, Marco, Michel Wedel, and Zsolt Sándor (2001), "Split-Questions: A New Tool in Survey Design," *Marketing Research*, 13, 2, (Summer), 14-19.

Wilkie, William L. and Edgar A. Pessemier (1973), "Issues in Marketing's Use of Multi-attribute Attitude Models," *Journal of Marketing Research*, 10, (November), 428-441.

Yedidia, Jonathan S., William T. Freeman and Yair Weiss (2003), *Understanding Belief Propagation and Its Generalizations: Exploring Artificial Intelligence in The New Millennium*, (San Francisco, CA: Morgan Kaufmann Publishers Inc.), 239-269.

Yee, Michael, Ely Dahan, John R. Hauser and James Orlin (2007) "Greedoid-Based Noncompensatory Inference," *Marketing Science*, 26, 4, (July-August), 532-549.

Table 1
Synthetic-Data Experiments
Number of Questions Necessary to Match Predictive Ability of 32 Orthogonal Questions[†]

	Adaptive Questions	Random Questions	Market-Based Questions	Orthogonal-Design	Percent Uncertainty[‡]
Base Comparison	9	40	39	32	76.1%
Error in answers	11	38	38	32	53.6%
"Bad" priors	6	42	41	32	50.4%

[†] Number of questions in addition to the configurator question.

[‡] U^2 (percent uncertainty explained) when heuristics estimated from 32 orthogonal questions. U^2 for other question-selection methods approximately the same subject to integer constraints on the number of questions.

Table 2
Illustrative Empirical Application
KL Divergence for Question-Selection-&-Estimation Combinations (Smaller is Better)

	Non-compensatory Heuristics	Compensatory Decision Model
Question-selection method		
Adaptive questions	0.475[*†‡]	0.537[‡]
Market-based questions	0.512[‡]	0.512[‡#]
Null Models		
Consider all profiles	0.565	
Consider no profiles	0.565	
Randomly consider profiles	0.562	

[*] Significantly better than market-based questions for non-compensatory heuristics ($p < 0.001$).

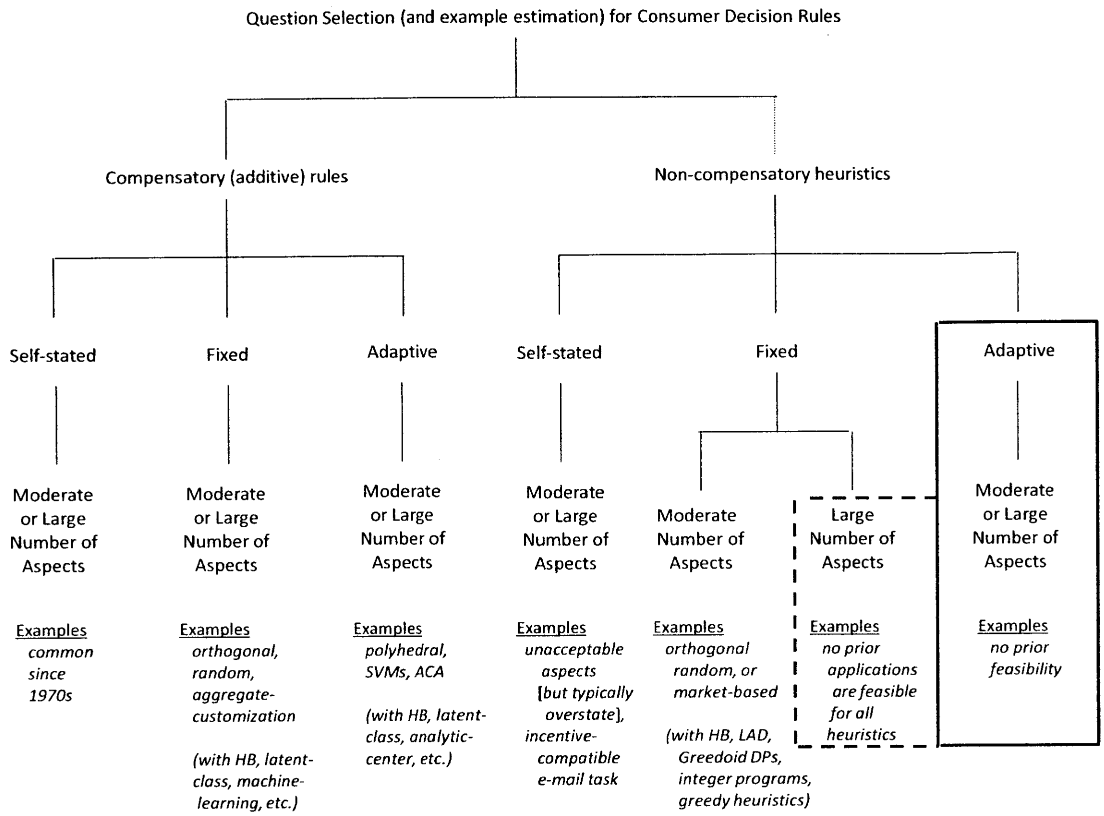
† Significantly better than compensatory decision model ($p < 0.001$).
‡ Significantly better than null models ($p < 0.001$).
# Significantly better than adaptive questions for compensatory decision model ($p < 0.001$).

Table 3
Percent of Respondents Using Aspect as an Elimination Criterion

Brand	Elimination Percent	Body Type	Elimination Percent	Engine Type	Elimination Percent
BMW	68%	Sports Car	84%	Gasoline	3%
Buick	97%	Hatchback	81%	Hybrid	44%
Cadillac	86%	Compact Sedan	62%	**Engine Power**	
Chevrolet	34%	Standard Sedan	58%	4 cylinders	9%
Chrysler	66%	Crossover	62%	6 cylinders	11%
Dodge	60%	Small SUV	61%	8 cylinders	69%
Ford	23%	Full-size SUV	71%		
GMC	95%	Pickup Truck	82%	**EPA Rating**	
Honda	14%	Minivan	90%	15 mpg	79%
Hyundai	89%			20 mpg	42%
Jeep	96%	**Quality**		25 mpg	16%
Kia	95%	Q-rating 5	0%	30 mpg	5%
Lexus	86%	Q-rating 4	1%	35 mpg	0%
Lincoln	98%	Q-rating 3	23%		
Mazda	90%			**Price**	
Nissan	14%	**Crash Test**		$12,000	77%

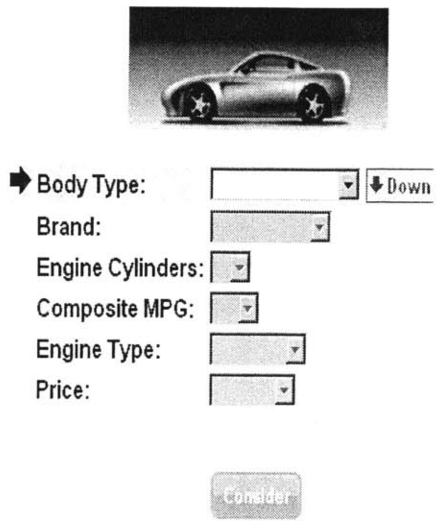
Pontiac	97%	C-rating 5	0%	$17,000	54%
Saturn	95%	C-rating 4	27%	$22,000	46%
Subaru	99%	C-rating 3	27%	$27,000	48%
Toyota	15%			$32,000	61%
VW	86%			$37,000	71%
				$45,000	87%

Figure 1
Taxonomy of Existing Methods to Select Questions to Identify Consumer Decision Rules

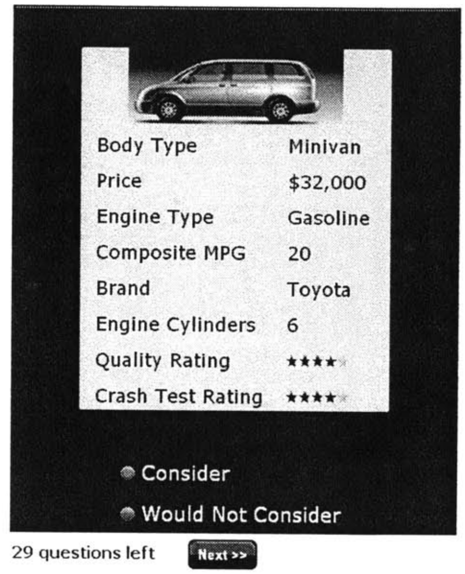


This page intentionally left blank

Figure 2
Example Configurator and Example Queries (Color in Original)



(a) Example Configurator



(b) Example Query

Figure 3
Synthetic-Data Experiments (Base Comparison, No Error)
Percent Uncertainty Explained (U^2) for Alternative Question-Selection Methods

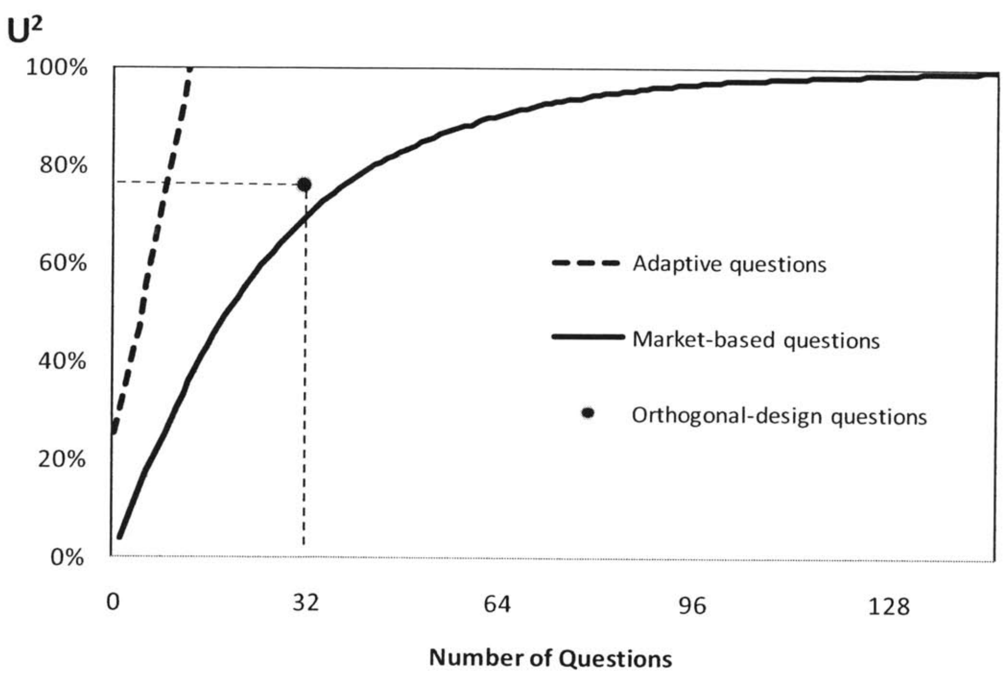
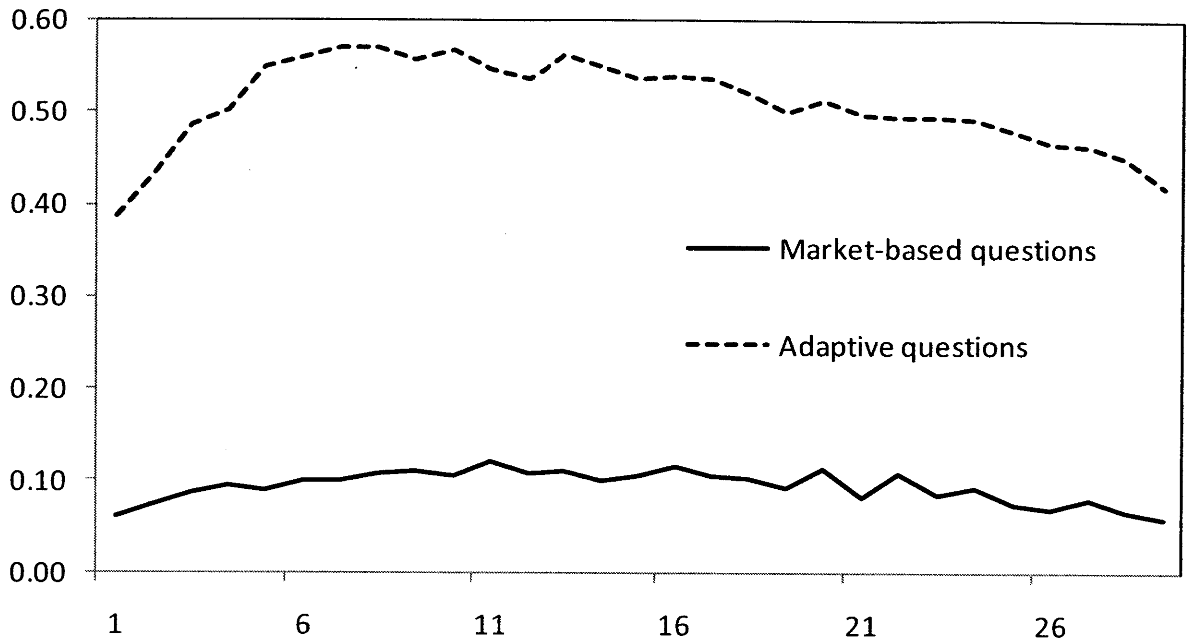


Figure 4
Average Expected Reduction in Entropy up to the 29th Question



Appendix 1. Example Where Uncertainty Sampling Minimizes Posterior Entropy

We choose a simple example with two aspects to demonstrate the intuition. For this formal example, we abstract away from response error by setting $\epsilon_1 = \epsilon_2 = 0$ and we choose uninformative priors such that $p_{i1}^0 = p_{i2}^0 = 0.5$. With two aspects there are four potential queries, $\vec{x}_{i1} = \{0, 0\}$, $\{0, 1\}$, $\{1, 0\}$, and $\{1, 1\}$, and four potential decision rules, $\vec{a}_i = \{-1, -1\}$, $\{-1, +1\}$, $\{+1, -1\}$, and $\{+1, +1\}$, each of which is *a priori* equally likely. But the different $\vec{x}_{i1}$'s provide differential information about the decision rules. For example, if $\vec{x}_{i1} = \{0, 0\}$ and $y_{i1} = 1$ then the decision rule must be $\vec{a}_i = \{-1, -1\}$. At the other extreme, if $\vec{x}_{i1} = \{1, 1\}$ and $y_{i1} = 1$, then all decision rules are consistent. The other two profiles are each consistent with half of the decision rules. We compute $\Pr(y_{i1} = 1 \mid \vec{x}_{i1})$ for the four potential queries as 0.25, 0.50, 0.50, and 1.00, respectively.

We use the formulae in the text for expected posterior entropy, $E[H(\vec{x}_{i1})]$:

Potential Query ($\vec{x}_{i1}$)	$\Pr(y_{i1} = 1 \mid \vec{x}_{i1})$	$E[H(\vec{x}_{i1})]$
{0, 0}	0.25	$-\frac{3}{2}\left(\frac{2}{3}\log_2\frac{2}{3}+\frac{1}{3}\log_2\frac{1}{3}\right) = 1.4$
{0, 1}	0.50	$-\log_2\frac{1}{2} = 1$
{1, 0}	0.50	$-\log_2\frac{1}{2} = 1$
{1, 1}	1.00	$-2\log_2\frac{1}{2} = 2$

Expected posterior entropy is minimized for either of the queries, $\{0, 1\}$ or $\{1, 0\}$, both of which are consistent with uncertainty sampling (choice balance).

Appendix 2. Pseudo-Code for Belief-Propagation Algorithm

Maintain the notation of the text and let $\vec{p}_{iK}$ be the vector of the p_{ijK}'s, let $\vec{p}_{iK,-j}$ be the vector of all but the j^{th} element. Define two index sets, $S_j^+ = \{k \mid x_{ijk} = 1, y_{ik} = 1\}$ and $S_j^- = \{k \mid x_{ijk} = 1, y_{ik} = 0\}$. Let superscript h index an iteration with $h = 0$ indicating a prior. The belief-propagation algorithm uses all of the data, X_K and $\vec{y}_{iK}$, when updating for the K^{th} query. In application, the ϵ's are set by managerial judgment prior to data collection. Our application used $\epsilon_1 = \epsilon_2 = 0.01$ for query selection.

Use the priors to initialize $\vec{p}_{iK}^{\,0}$. Initialize all $\Pr\left(y_{ik} \mid X_{iK}, \vec{p}_{iK,-j}^{\,h-1}, a_{ij} = \pm 1\right)$.

While $\max_j (p_{ijK}^h - p_{ijK}^{h-1}) > 0.001$. [Continue looping until p_{ijK}^h converges.]

For j = 1 to N [Loop over all aspects]

For $k \in S_j^+$ [Use variational distribution to approximate data likelihood]

$\Pr(y_{ik} = 1|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = 1) = (1-\epsilon_2)\prod_{x_{igk}=1, g\neq j} p_{igK}^{h-1} + \epsilon_1(1 - \prod_{x_{igk}=1, g\neq j} p_{igK}^{h-1})$

$\Pr(y_{ik} = 1|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = -1) = \epsilon_1$

end loop $k \in S_j^+$

For $k \in S_j^-$ [Use variational distribution to approximate data likelihood]

$\Pr(y_{ik} = 0|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = 1) = (1-\epsilon_1)(1 - \prod_{x_{igk}=1, g\neq j} p_{igK}^{h-1}) + \epsilon_2 \prod_{x_{igk}=1, g\neq j} p_{igK}^{h-1}$

$\Pr(y_{ik} = 0|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = -1) = (1-\epsilon_1)$

end loop $k \in S_j^-$

$\Pr(\vec{y}_{iK}|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = 1) = \prod_{k=1}^{K} \Pr(y_{ik}|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = 1)$

$\Pr(\vec{y}_{iK}|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = -1) = \prod_{k=1}^{K} \Pr(y_{ik}|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = -1)$ [Compute data likelihoods across all K questions as a product of marginal distributions for each k.]

$\Pr(a_{ij} = 1|X_{iK}, \vec{y}_{iK}, \vec{p}_{iK,-j}^{h-1}) \propto \Pr(\vec{y}_{iK}|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = 1)\Pr(a_{ij} = 1|prior)$

$\Pr(a_{ij} = -1|X_{iK}, \vec{y}_{iK}, \vec{p}_{iK,-j}^{h-1}) \propto \Pr(\vec{y}_{iK}|X_{iK}, \vec{p}_{iK,-j}^{h-1}, a_{ij} = -1)(1 - \Pr(a_{ij} = 1|prior))$

$p_{ijK}^{h} = \Pr(a_{ij} = 1|X_{iK}, \vec{y}_{iK}, \vec{p}_{iK,-j}^{h-1})$ normalized. [Use Bayes Theorem, then normalize]

end loop j [Test for convergence and continue if necessary.]

Appendix 3. Kullback-Leibler Divergence for Empirical Data

The Kullback-Leibler divergence (KL) is an information-theory-based measure of the divergence from one probability distribution to another. In this paper we seek the divergence from the predicted consideration probabilities to those that are observed in the validation data, recognizing the discrete nature of the data (consider or not). For respondent i we predict that profile k is considered with probability, $r_{ik} = \Pr(y_{ik} = 1|\vec{x}_{ik}, model)$. Then the divergence from the true model (the y_{ik}'s) to the model being tested (the r_{ik}'s) is given by Equation A1. With log-based-2, KL has the units of bits.

$$\text{(A1)} \qquad KL = \sum_{k \in validation} \left[y_{ik} \log_2\left(\frac{y_{ik}}{r_{ik}}\right) + (1 - y_{ik}) \log_2\left(\frac{1 - y_{ik}}{1 - r_{ik}}\right) \right]$$

When the r_{ik}'s are themselves discrete we must use the observations of false-positive and false-negative predictions to separate the summation into four components. Let V = the number of profiles in the validation sample, $\hat{C}_v$ = the number of considered validation profiles, F_p = the false-positive predictions, and F_n = the false-negative predictions. Then KL is given by the following equation where $S_{c,c}$ is the set of profiles that are considered in the calibration data and considered in the validation data. The sets $S_{c,nc}$, $S_{nc,c}$, and $S_{nc,nc}$ are defined similarly ($nc \rightarrow$ not considered).

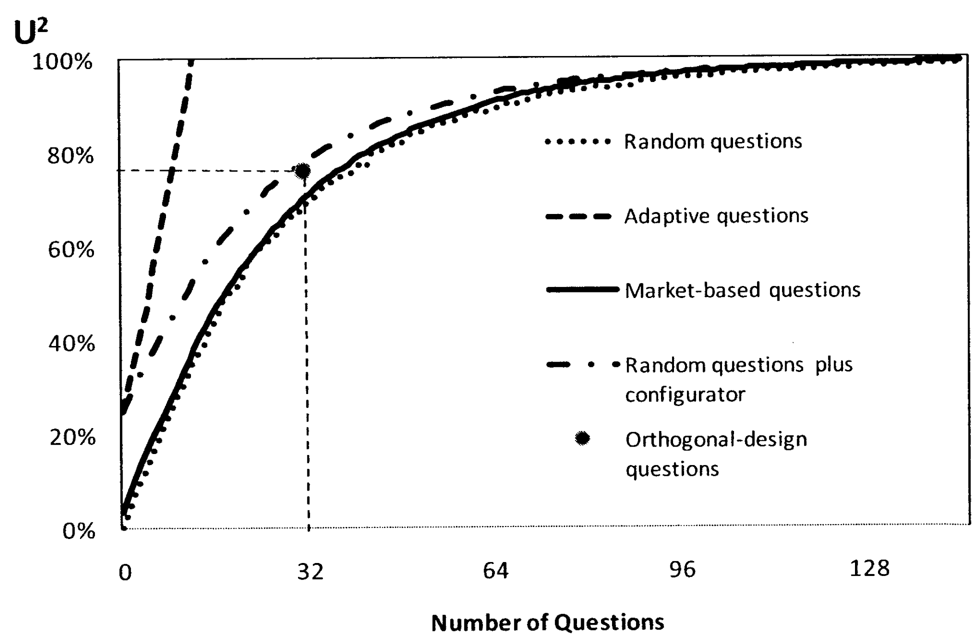
$$KL = \sum_{S_{c,c}} \log_2 \left(\frac{\hat{C}_v}{\hat{C}_v - F_p}\right) + \sum_{S_{c,nc}} \log_2 \left(\frac{V - \hat{C}_v}{F_n}\right) + \sum_{S_{nc,c}} \log_2 \left(\frac{\hat{C}_v}{F_p}\right) + \sum_{S_{c,c}} \log_2 \left(\frac{V - \hat{C}_v}{V - \hat{C}_v - F_n}\right)$$

After algebraic simplification, KL can be written as:

(A2)
$$KL = \hat{C}_v \log_2 \hat{C}_v + (V - \hat{C}_v)\log_2(V - \hat{C}_v) - (\hat{C}_v - F_p)\log_2(\hat{C}_v - F_p)$$
$$- F_n \log_2 F_n - F_p \log_2 F_p - (V - \hat{C}_v - F_n)\log_2(V - \hat{C}_v - F_n)$$

KL is a sum over the set of profiles. Sets with more profiles are harder to fit; if V were twice as large and $\hat{C}_v$, F_p, and F_n were scaled proportionally, then KL would be twice as large. For comparability across respondents with different validation-set sizes, we divide by V to scale KL.

Appendix 4. Percent Uncertainty Explained (U^2) for Other Question-Selection Methods



Appendix 5. Hit Rates for Question-Selection-&-Estimation Combinations (Larger is Better)

	Non-compensatory Heuristics	Compensatory Decision Model

Question-selection method		
Adaptive questions	0.848[*,†,‡,&,$]	0.594[$]
Market-based questions	0.827[†,‡,&,$]	0.806[#,‡,$]
Null Models		
Consider all profiles	0.180	
Consider no profiles	0.820	
Randomly consider profiles	0.732	

[*] Significantly better than market-based questions for non-compensatory heuristics ($p < 0.001$).
[†] Significantly better than compensatory decision model ($p < 0.001$).
[‡] Significantly better than random null model ($p < 0.001$).
[$] Significantly better than consider-all-profiles null model ($p < 0.001$)
[&] Significantly better than consider-no-profiles null model ($p < 0.001$)
[#] Significantly better than adaptive questions for compensatory decision model ($p < 0.001$).

Chapter 3: Product Recommendation Strategy as Consumers Construct Preferences

Abstract

We propose a new, empirically-grounded, model of product search when preferences are constructed during the process of search: consumers learn what they like and dislike as they examine products. Product recommendations, whether made by sales people or online recommendation systems, bring products to the consumer's attention <u>and</u> impact his/her preferences. Changing preferences changes the products the consumer will choose to search; at the same time, the products the consumer chooses to search will determine the future shifts in preferences. Accounting for this two-way relationship between products and preferences is critical in optimizing recommendations. Using simulated consumer preferences and product attributes, we demonstrate that consumers often do not find the best product even if search is costless, and that consumers can benefit from seeing an undesirable product, because it may help them learn their preferences. A survey of realtors confirms model predictions.

Keywords: Search, Constructed Preferences, Product Recommendations

1 Introduction

Consider the following story of a family looking for a house to buy, described in a recent article in MSN Real Estate. The real estate agent who helped Jon Weiner and his wife, Lindsey to find a home began by asking his clients to make a list of what they wanted in a home: what was important and what was simply nice to have. At first, "we were so focused on the interior that we basically said, 'Yard would be good, but we don't really care how big it is,'" Weiner says. But it turned out that the Weiners like their privacy, and "a decent-sized yard became a need. Looking out the window and seeing a house 20 feet away couldn't work," he says. "The postage-stamp property wasn't going to make us happy, regardless of how nice the hardwoods might be." They purchased a home that was smaller than what they originally thought they wanted, but "the slightly smaller house has everything we need and a little bit more," Jon Weiner says (Culler, 2012).

This story is typical. When searching for high-involvement goods such as houses, insurance plans, complex electronics, or automobiles, consumers may start out with a search objective that does not represent his or her true preferences; as preferences shift throughout the search, so does the search objective. The consumer chooses which products to evaluate based on his current preferences; which products the consumer evaluates determines the future shifts in preferences. We model this two-way relationship between preferences and evaluated products, and study its impact on (1) the search process and (2) the strategy for an outside agent making product recommendations.

Understanding the nature of consumer search behavior is critical for influencing brand choice and managing marketing communication. Consumer search behavior has been widely studied in marketing, economics, and operations research. In most models of consumer product search, preferences are assumed to be static and known by the consumer, and the consumer explores the market in search for the utility-maximizing product. Another stream of literature, in consumer behavior, provides evidence that consumer preferences are learned, rather than static. In this paper we combine insights from both streams and explore how dynamic (constructed) preferences that change during the search process.

We demonstrate that without product recommendations, consumers often make suboptimal choices, *even if search is costless*. In existing search models, consumers make suboptimal choices because the cost of search prohibits them from finding the utility-maximizing product; if the cost of searching one extra product was zero or negligibly small, consumers would search all products and find the maximum. When there are constructed preferences, consumers may make sub-optimal choices because they have not yet learned the correct objective function: they do not find the optimal product because they don't know they should be looking for it. This result is particularly relevant in many of today's *high information* shopping environments, such as electronics, real estate, and insurance plans to name a few. In these markets, consumers obtain information on existing products easily and with little cost.

Constructed preferences provide one explanation for why consumers continue to use sales agents and decision aides, even in high information markets, and agents often show buyers potential products that vary widely, including some products that buyers would not be expected to buy. By making recommendations, sales agents bring the recommended products to the consumers' attention and help them learn their own preferences. Constructed preferences introduce a new source of information asymmetry between the buyer and the recommending agent, even if both of them have access to the same information on product attributes. For example, from surveying realtors, we have observed that the best sales people have expertise on consumers' preferences. Having worked with many buyers, they often predict shifts in consumer preferences before they occur. After seeing a home with a terrible view, they anticipate that the buyer is likely to appreciate a good view in another home; after seeing a home with a deck, they anticipate that buyers, who had not thought about having a deck, may start seeking decks as they continue their search.

The ability to forecast preference shifts is critical to sales agents' ability to continue providing value to buyers in high-information markets. For example in the real estate industry, the Multiple Listing Service (MLS) provides a database of all houses and their attributes in an easily accessible format. Before such resources were widely available, consumers needed real estate agents to search for properties for sale. A recent article in the *Realtor Magazine* describes the resulting changes in the role of real estate agents: "There was a time when providing great service in real estate meant [...] teaching [clients] what they didn't already know. Simply providing them with information on homes for sale, pricing, current mortgage rates and data relevant to the transaction gave practitioners a strong leg-up as consumer access to such information was limited. Today, that doesn't happen anymore. Consumers — especially those under 40 — want to find that information on their own." (Summerfield, 2012) If preferences were static and known to the consumer, and consumers could easily explore available properties themselves, there would be little left for a sales agent to do in a high-information market.

Constructed preferences provide an opportunity for product recommendations to improve consumer search, but the recommendation must carefully account for the two-way relationship between the products the consumer views and his/her preferences. For example, some consumers benefit from viewing a bad product, because it may help them fine-tune their preferences, making the future search more efficient.

In this paper, we demonstrate that the theory of constructed preferences on product recommendations and advertising targeting provides one explanation for observed behavior by realtors and their customers. After reviewing the relevant literature, we describe depth interviews with real estate agents that ground our model of clients' interaction with realtors.

We then develop a formal model based on constructed preferences and show it is consistent with the observed behavior. To test some of the model's predictions, we provide a summary of a survey conducted with 140 professional real estate agents. While constructed preferences is consistent with observed behavior it may be only part of the explanation. We explore alternative explanations and suggest critical experiments for future research.

2 Literature Review

The research in this paper builds upon results in three literatures: consumer search, constructed preferences, and recommendation systems. Each of these literatures is extensive. In this section we review the findings that are most relevant to the model developed in this paper.

Search

Marketing scholars have long recognized that consumers invest a substantial amount of time in searching prior to making a purchase, particularly in expensive durable goods categories. This paper differs from this literature by allowing the search objective to change depending on the products the consumer sees as s/he searches. Early research focused on consumer search for products that differ on a single vertical dimension (e.g. Stigler 1961, Weitzman 1979). Consumers engage in sequential search and the problem becomes an optimal stopping problem; consumers continue searching if and only if the expected gain from searching outweighs the costs. This is a variant of the well known "secretary problem," first posed in Gilbert and Mosteller (1966), and since extended and generalized in many different directions (for a review, see Ferguson 1981). More recently, researchers have used structural models (e.g. Erdem and Keane 1996), in which forward-looking consumers trade off consumption with search for information, usually with application to frequently purchased goods. Recent models (e.g. Hauser, Urban, and Weinberg 1993; Kim, Albuquerque and Bronnenberg 2010; Moorthy, Ratchford and Talukdar 1995) explore search on horizontally differentiated products with multiple attributes, but assume consumers are fully aware of their preferences, and only have uncertainty about product attributes. The model developed in this paper adapts the widely used standard sequential search and linear additive utility assumptions, but deviates from the extant literature in that consumers lack knowledge about their own preferences, rather than product attributes.

Constructed Preferences

The proposed theory is based on three ideas that are firmly established in the literature. (1) The purchase decision rule is adaptive and learned. Rather than starting out with well-articulated preferences, consumers form their preferences as they search and evaluate products (e.g., Bettman, Luce, and Payne 1998, 2008; Liechty, Fong, DeSarbo 2010; Feldman and Lynch 1988; Payne, Bettman, and Johnson 1988, 1992; Slovic, Griffin, and Tversky 1990; Ülkümen, Chakravarti, Morwitz 2010). (2) Tasks that force consumers to think deeply about their own preferences cause consumers to change their preferences (Huber et al 1993; Nordgren and Dijksterhuis 2008, Hauser, Dong, and Ding 2011). Articulated preferences change after self-reflection is induced. (3) As consumers learn and evolve from novice to expert consumers (for the category), their preferences change and eventually stabilize (Alba and Hutchinson 1987, 2000; Betsch, et al 2001; Brucks 1985). The model used in this essay assumes that the consumer is *discovering* his or her preferences, and that, eventually, preferences converge to a steady state. However, this steady state assumption is not critical. As long as viewing properties changes the consumer's preferences *in a way that is predictable to the sales agent*, the key results hold.

Recommendation Systems

Recommendation systems play an important role in helping consumers find the best product to fit their needs. Recommendations may come from either online systems or human sales agents, and are used in a variety of product categories ranging from movies to healthcare plans. There is a vast literature that focuses on optimizing which products should be recommended to a specific customer; most techniques focus on recommending products that have a high probability of purchase. Various models have been proposed to infer customers' responses, including the mixture model (Chien and George 1999) and the hierarchical Bayes model (Ansari, Essegaier, and Kohli 2000). Ying, Feinberg, and Wedel (2006) demonstrate how incorporating missing ratings into the model improves the quality of recommendations. Bodapati (2009) proposes recommending the products most sensitive to product recommendation by discounting products of which the consumer is likely already aware. In this paper, we allow that consumers learn their preferences as they evaluate products, so a product recommendation not only brings the product to the consumers attention, but may also change his/her preferences, thus altering the entire future search path.

Since our focus in this paper is to understand the consumer's search behavior, and how it is influenced by product recommendations, we will abstract from agency issues and take the perspective of a recommender system that is incentive aligned to help the consumer make the best choice. While this assumption does not hold in many real markets, such as automobiles, we argue that in other markets it is more realistic. Qualitative evidence presented in Section 3 supports this assumption for some real estate markets, especially when word of mouth and reputation effects are strong.

3 Interviews with Realtors

To build an empirically grounded model, we began with exploratory work, conducting depth interviews with realtors in a densely populated residential neighborhood in a large US city. We chose a market that would help us isolate the effect of constructed preferences from other factors that impact search aided by a sales agent, especially misaligned sales agent incentives. This neighborhood fits well, because the real estate offices are located close to each other, at least 15 of the offices (with multiple agents) along a quarter-mile stretch of a single street. In such a competitive environment, realtors are very aware of the possibility of losing clients to one another and the effects of negative WOM. Realtors in such an environment report that they are more likely to be concerned with helping the consumer find the best home, rather than trying to maximize commissions. Rather than providing detailed quotes from the interviews here, we provide them throughout the paper to motivate assumptions or to interpret results. Anticipating those details, we summarize the key insights. 1) Realtors are aware that consumer's preferences change during the search process, and account for it when choosing homes to show. 2) Many realtors cite preference changes as the reason for consumers' viewing multiple homes, as opposed to just being able to find and show a customer the home that best fits his/her preferences. 3) The examined real estate market is very competitive, and realtors are much more concerned with turnover than commission incentives.[1]

4 Model of Sequential Search with Dynamic Preferences

High Information Market We assume that the consumer has access to and can search all the available products on the market. This setting is particularly justified with the proliferation of online resources for product search. Examples include the Multiple Listing Service (MLS) for real estate and Edmunds.com for cars. Indeed, according to NAR's 2010 *Home Buyer and Seller Profile*, last year more people first identified potential homes through the Internet than through a real estate agent. The ease with which consumers find information about the market is well known to practitioners. For example, one consultant was quoted in a real estate magazine regarding selling to Generation Y consumers, "This is a generation that takes advantage of information. They come in [to a home showing] knowing everything about the property." The real estate agent can help navigate the search, but consumers can access many (not all) product attributes while incurring little or no search cost. In this paper, we will

[1] The goal of this paper is to demonstrate that markets exist where constructed preferences change the behavior of both consumers and sales agents. I do not claim these effects hold in all markets, even all real estate markets.

abstract away from any attribute uncertainty, and assume that consumers obtain all attribute information at no cost.[2] Additionally, we assume that a consumer must visit the home before buying it, despite having no uncertainty about its attributes.

<u>Linear Additive Utility with Evolving Weights</u> We assume consumers have a linear additive utility function, and product features have discrete levels. Each *aspect*, or feature level, is assigned a partworth, and the utility for product k is the sum of the partworths of all its aspects ($\vec{x}_k$). Let $\vec{x}_k$ be a binary (0,1) vector representing product k, such that $x_{kj} = 1$ if product k has aspect j, and 0 otherwise. For example, in the automobile category, Toyota, Chevrolet, and Ford are three different aspects and correspond to different elements of $\vec{x}_k$. Also, let X_k represent the *set* of aspects present in product k, such that $X_k = \{j|x_{kj} = 1\}$. The errors are independent, identical random variables with double exponential distribution. For a given consumer, i, the partworths are contained in the vector $\vec{w}_i$, such that each aspect j has partworth w_{ij}. Let us call $\vec{x}_i^* = \arg\max_k \vec{x}_k\vec{w}_i$, the true utility maximizing product for consumer *I*.

<u>Preference Updating Rule</u> The consumer may not know the value of the complete vector $\vec{w}_i$.

S/he has an estimate of $\vec{w}_i$, which we call $\vec{v}_{it}$, where t indexes time (discrete). For example, s/he may under- or over-estimate the value of some aspects ($w_{ij} > v_{ij}$ or $w_{ij} < v_{ij}$, respectively), or be unaware of some important aspects ($v_{ij} = 0$, $w_{ij} \neq 0$). If the consumer had complete information, we would have $\vec{w}_i = \vec{v}_i$, or perfect knowledge of preferences. We do not require that $\vec{v}_{it}$ converge to $\vec{w}_i$, but in general, evaluating products moves $\vec{v}_{it}$ toward $\vec{w}_i$. After evaluation of a product k, the consumer updates his preferences such that $v_{ij,t+1} = w_{ij}$ for all $j \in X_k$ and the other corresponding values of v_{it} are left unchanged.

[2] Uncertainty about product attributes has been extensively studied in the search literature, and our goal here is to explore constructed preferences as an alternative explanation for observed search behavior. I expect that constructed preferences will have a similar impact on search behavior in the presence of attribute uncertainty, but this research is beyond the scope of this paper

Outside Good The consumer has access to an outside good, with constant utility B, that is known to both the realtor and the consumer. In the case of real estate, the outside good may represent, for example, staying at their current place of residence, choosing to rent instead of buying, or the option-value to searching in a neighborhood outside the realtors' territory. The higher the value of B, the higher the probability that the buyer will choose the outside good, and the realtor will not be able to strike a deal. Buyers who are very set on purchasing something have a low B, while those who are flexible have a high B. The sales agent can infer the value of B based on his or her communication with the potential client.

A graphical representation of the search process is presented in Figure 1. When searching on their own, without product recommendations, consumers iterate through the following steps:

1. **Decide what/whether to search.** Select the perceived utility maximizing product, $\vec{x}_t^* = \arg\max_k \vec{v}_{it}\vec{x}_k$ from the database. Since consumers do not know their true utility $\vec{w}_i$, they use their current perceived utility, $\vec{v}_{it}$ when they search the product space. If the perceived utility maximizing product is one that the consumer has already viewed, i.e. $\vec{x}_t^* \in \{\vec{x}_1^*, \ldots \vec{x}_{t-1}^*\}$, or has not yet been viewed but is not perceived to have higher utility than one of the previously viewed products minus the search cost, the consumer decides it is not worth seeing; consumer stops searching and goes to step 3. Otherwise, goes to step 2.
2. **Reconstruct preferences.** The consumer and realtor both pay a search cost, c_s, and visit the property. The consumer learns his true valuation, w_{ij}, of this property's aspects, thus constructs $\vec{v}_{i(t+1)}$, and returns to step 1.
3. **Decide whether or not to buy.** The property that maximizes the current utility $\vec{v}_{it}$ enters into a competition with the outside good with value B, with some noise introduced at the time of the purchasing decision. This noise is not predictable to

either the consumer or the sales agent, so the choice is probabilistic. The probability the product is purchased (chosen over the outside good) is given by[3]:

$$\Pr(buy) = \frac{e^{\vec{x}_t^* \vec{v}_{it}}}{e^B + e^{\vec{x}_t^* \vec{v}_{it}}}$$

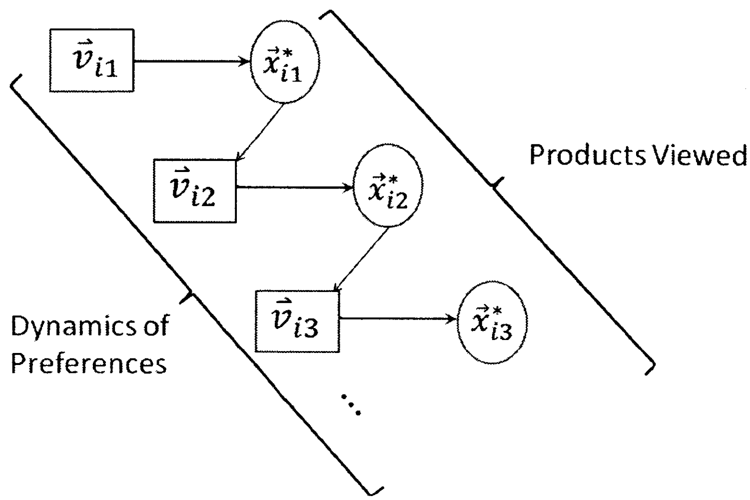


Figure 1

Stopping Rule. We adopt the standard stopping rule in the sequential search literature that is the solution of a cost-benefit tradeoff: consumer stops searching when the expected payoff from finding a higher quality product is lower than the cost of searching more. A few papers explore models with exogenous stopping rules, for example a single take-it-or-leave-it offer (Wernerfelt 1994), or a separate geometric random process, in which an individual continues to search after each step with an independent probability (Johnson et al 2004). In this paper, we stay consistent with the cost-benefit framework of Weitzman 1979, and the stopping rule becomes endogenous to the changing preferences.

5 How Effective is Consumer Search?

In this section, we are interested whether consumer search leads consumers to their true utility maximizing product, and how long it takes. We introduce extra notation. Let $\tilde{t}$ be

[3] We use the logit model specification for the probability of purchase because it is the most commonly used model of probabilistic discrete choice. The results in this paper do not rely on this particular specification; any monotonically increasing function of the utility will generate similar results.

the time when the consumer stops searching, and $\tilde{k}$ be the best product seen up to time $\tilde{t}$. The consumer may then choose to purchase that product, or take the outside good. The net expected payoff from the process for the consumer is given by the expected utility of this choice minus the accumulated search cost up to time $\tilde{t}$:

$$U_i = \frac{\vec{x}_{\tilde{k}}\vec{v}_{i\tilde{t}} \cdot e^{\vec{x}_{\tilde{k}}\vec{v}_{i\tilde{t}}} + B \cdot e^{B}}{e^{B} + e^{\vec{x}_{\tilde{k}}\vec{v}_{i\tilde{t}}}} - \tilde{t}c_s$$

If the consumer searches without outside interruption, the values of $\vec{w}_i$ and $\vec{v}_{i0}$ determine which product the consumer will select at each time t, the corresponding evolution of $\vec{v}_{it}$, how many products the consumer will inspect before purchasing, and how close to the maximum potential payoff the consumer is able to get. We will refer to this deterministic process as the *uninterrupted search process*.

Proposition 1: (Base Case) If $\vec{v}_{i0} = \vec{w}_i$, or the consumer has perfect knowledge of his preferences, he finds the utility maximizing option at $t = 1$, and looks no further after the verification stage. This scenario generates the highest possible payoff for the consumer:

$$U_i^* = \frac{(\vec{x}_i^*\vec{w}_i - c_s) \cdot e^{\vec{x}_i^*\vec{w}_i} + B \cdot e^{B}}{e^{B} + e^{\vec{x}_{\tilde{k}}\vec{v}_{i\tilde{t}}}}$$

Proof: See Appendix 1

This result suggests that realtors who deal with expert consumers, who start out with already well-articulated, stable preferences, will show them only one property, and it is the property that best meets their needs. When available products are easy to search, such a client can communicate his preferences to the sales agent, who is aware of all the products' attributes, and will identify the utility maximizing product immediately. Interviews with real estate agents support this result. The following is a quote from a realtor on the topic of search length:

"It does happen, not very often, that they see just one and take it. Usually with people who have been living (in this neighborhood) for a long time and know the area and know exactly what they want."

Corollary: If a consumer has to see more than one product before purchasing, s/he does not have full knowledge of her/his preferences.

This corollary is the contrapositive[4] of Proposition 1:. That means that, given our model, if a consumer saw multiple products before making a purchase decision, that consumer must not be perfectly aware of his preferences. We find support for this conclusion in our qualitative interview data as well. The realtors reported an average of six visits for a rental, and more than that for a sale. When asked about the reason for the search taking so long, most of the realtors refer to changing preferences throughout the search sequence. The following are sample responses to this question:

- *"Often what people start out thinking they want is not what they end up wanting."*
- *"Let's say they tell you they want three things, like renovated kitchen, pet friendly, and up to $2500. I find them something that has those 3. Then they get there and tell me they hate the view and won't take it because of that."*
- *"People may not think about what common areas look like, but once they actually go out and see it, they realize that they will be affected by it."*
- *"Let's say someone is looking for a 1 bedroom with a good layout. I show them one, and then I walk in the bedroom and open up a French door to a private deck. They love that, and want me to look for more apartments with a deck."*

The exploratory data provide qualitative support for the theory and suggest that dynamic preferences can explain why consumers often examine many products before purchasing. Note that in real markets there are other factors that contribute to the challenge of search, and for any given market it is an empirical question which factors dominate. We discuss alternative explanations and suggest critical experiments in Section 8.

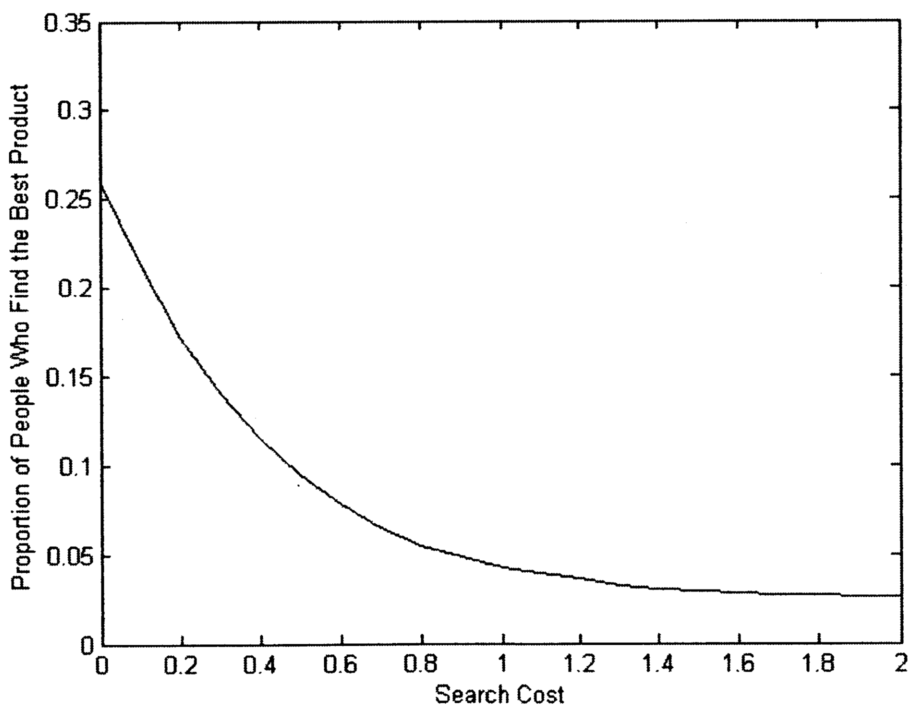
---

[4] A contrapositive of a proposition "if P then Q" is "if not Q then not P". The contrapositive of "if consumer has full knowledge of preferences, then searches only once" is "if consumer searches more than once, then does not have full knowledge of preferences".

Proposition 2: (Local maxima rather than global maximum) There exist values of $\vec{w}_i$ and $\vec{v}_{i0}$, for which an uninterrupted search process will stop at some product before finding $\vec{x}_i^*$.

Proof: See Appendix 1.

Searching on his or her own, the consumer may get "stuck" in a local maximum, and never find the global maximum. To demonstrate this result, we simulated the search process for 5,000 hypothetical consumers. Details of the simulation are provided in Appendix 2. Note that even when search cost is zero, only 26% of consumers find the global maximum. A plot of the relationship between the search cost and the proportion of consumers who find the global maximum before they stop searching is provided in Figure 1. As expected, the higher the search cost, the fewer consumers find the global maximum.



6 What is the Best Product Recommendation?

We now explore how product recommendations influence the consumer, causing the consumer to deviate from the uninterrupted search process. A product recommendation is a way for the sales agent to get the consumer to evaluate a product. For the recommendation to have any impact on the search path, it would not be the product the consumer would have chosen to evaluate at that point; that is, not the current perceived utility maximizing product, $\arg\max_k \vec{v}_{it}\vec{x}_k$. Because the search objective gets updated after each evaluated product, the time series of $\vec{v}_{it}$ may be different from the times series in the corresponding uninterrupted process. This way, a product recommendation can change the entire search path that follows it, not merely bring the recommended product to the consumer's attention. A carefully chosen product recommendation can help alleviate the local maximum problem above, and ultimately lead the consumer to find the global maximum. To explore the effect of product recommendation, we introduce new notation. Let the ordered pair $(\vec{x}_r, t_r)$ represent the recommended profile, and the time of the recommendation, respectively.

First, let us analyze what the best recommendation would be if preferences were static, and the consumer had uncertainty about product attributes. Assume that the sales agent is aware of all products' attributes and tries to maximize the consumers' utility.

> <u>With Static Preferences, Optimal to Recommend Best Product</u> *If the consumer has perfect knowledge of* $\overrightarrow{w_i}$, *but may have uncertainty about product attributes, and the recommending agent knows both* $\overrightarrow{w_i}$ *and the attributes of all product, the sales agent recommends* $\vec{x}_r = \vec{x}_i^*$ *at time* $t_r = 1$ *and the consumer accepts it without searching more.*

Recall that we are taking the position of a recommending agent that gets a binary payoff: 1 if the consumer purchases and 0 if the consumer chooses the outside good, and that both the consumer and agent pay search costs. In this case, the agent wants to maximize the probability that the consumer will purchase, which is monotonically increasing in the utility of the chosen product. Thus, neither the agent nor the consumer have incentives to deviate from the above scenario.

If preferences don't change (and the other assumptions hold), examining an undesirable product has no value, because only the best product viewed will be purchased. Its probability of being chosen over the outside good is directly related to its utility. If the model assumes that preferences are static, the best recommendation is the utility-maximizing product. This result is consistent with a large volume of literature on recommendation systems (for a review, see Adomavicius and Tuzhilin 2005). Most algorithms seek the utility maximizing product to recommend.

If preferences are constructed, optimizing the recommendation is much more difficult. Suppose the consumer accurately communicates his current utility to the sales agent. The sales agent, acting in the best interests of the client, will sometimes recommend a product that does not optimize what the buyer said he wanted: $\vec{x}_t^r \neq \arg\max_k \vec{x}_k \vec{v}_{it}$. Section 7 provides survey results that support this prediction: 82% of professional real estate agents indicate that they sometimes recommend homes that did not optimally match their clients' stated needs. The recommended product has to be different from the product the consumer would have chosen on his/her own at the same time in order to have any impact on the search process (and outcome). Otherwise, the resulting preferences after evaluating the product will be the same as in the uninterrupted process and the following search processes will be identical. If the sales agent can anticipate how the preferences will change, one hypothesis might be to recommend the product the maximizes the client's long-term utility, $\vec{w}_i$. However, we demonstrate that this is not always the case.

The Most Beneficial Recommendation May be an Undesirable Product *There exist values of* $\vec{w}_i$ *and* $\vec{v}_{it}$ *for which it is optimal to recommend the buyer a product that* *s/he is unlikely to buy, even if there exist products that s/he is more likely to buy:* $\vec{x}_t^r \neq \vec{x}_i^*$.

Evaluating a bad product can help the consumer learn his preferences and be more efficient in his/her future search. For example, learning about one's preferences for several undesirable attributes will save the consumer time in the future that he may have otherwise spent looking at products with those attributes. Showing an apartment that is on a garden level and in a bad neighborhood will save the consumer time because the consumer will learn that he or she need not see additional homes in that neighborhood, or more garden level apartments.

To demonstrate this result in simulation, we allow an outside agent to recommend the consumer exactly one of the existing products at the beginning of the search. We perform an exhaustive search, to find the exact optimum for each customer: the product that maximizes the consumer's final choice utility minus search cost. For the parameters used (see Appendix for details), 14% of the optimal product recommendations are not the utility-maximizing product. These consumers' average utility of their respective best products, $\vec{x}_i^*$, is 2.91. The utility of the product they end up choosing when they search on their own *without* a recommendation is 2.55, or 0.36 below the optimum product. When they receive the recommendation, which is a suboptimal product, the utility of the product they end up choosing is 2.91, the same as the utility of $\vec{x}_i^*$: they all choose the optimal product. Survey results (Section 7) demonstrate that real estate agents adopt this strategy.

7 Sales Agents: Survey Methodology and Results

To assess external validity of the model's predictions, an online questionnaire was conducted among 140 US real estate agents working for a large international realty firm. Respondents were recruited through a call for respondents published in a research report that the agents had opted into receiving. The survey took about 10 minutes to complete. In exchange for completing, the first 75 respondents could chose between getting money donated to a charity (their choice among five charities) and being entered in a lottery to win an Amazon gift card. In responding to the survey, respondents were asked to answer with respect to their *last* client who was serious about purchasing a home, whether or not that client ended up purchasing.

We now summarize the key findings from the survey.

Some buyers change their preferences during the search. This construct was measured using a Likert scale. The subsection heading stated: "Please describe the following characteristics of your interaction with your client," and the Likert scale item stated "The client changed their mind about what they wanted during the search process." The respondent rated the statement on a scale from 1 (Strongly Agree) to 5 (Strongly Disagree). Note that the higher the number the *less* the client changed his/her preferences. The average response was 2.77; only 34% of the respondents indicated "Disagree" or "Strongly Disagree".

How much buyers change their mind during search is positively correlated with the number of homes seen. The first construct, how much the client changed their mind during the search, was measured as described above. The number of properties the client visited was measured using a slider, labeled "How many properties, in total, did you take the client to see?" The average response was 9.88. The correlation between the two measures was -0.286, $p<.001$. This significant correlation means that how much a client's preferences change influences the length of their search, which is consistent with Proposition 1 and its corollary, and the fundamental premise of this paper.

Realtors can anticipate if the client is going to change his/her preferences before visiting any homes with the client. The first construct, whether the realtor anticipated changes in the client's preferences as the client visited homes, was measured in Section 1, in which the respondent is asked about "first time you spoke with the client about purchasing a home, BEFORE you and the client went out to look at potential properties". The subsection was labeled "Please describe your impression of the client's description of what s/he/they wanted in a potential property." The Likert scale item was labeled "I could anticipate that the client would change his/her/their minds about what they were looking for after seeing some homes", and the respondent was asked to rate the item from 1 (Strongly Agree) to 5 (Strongly Disagree). The average response was 2.46. The measure of the second construct, how much the client changed their mind during the search, is described in the previous paragraph. The correlation between these two constructs was 0.472, $p<.0005$. This suggests that realtors, who are experts at helping guide people's search process, are good at anticipating which clients are likely to change their preferences, and which clients are not.

Realtors take preference dynamics into account when recommending properties. This is the result of two constructs: whether preference dynamics were present, and whether the realtors' recommendations accounted for the preference dynamics. The first construct, whether the realtor anticipated changes in the client's preferences as the client visited homes, was measured as described above. The second construct was measured using two questions, both on Likert Scales in Section 2. The subsection heading stated: "Please describe the following characteristics of your interaction with your client," and the Likert scale items stated "I only showed the client properties that matched their stated needs," and "I chose some properties to help the client learn what they wanted". The respondents were asked to rate both items from 1 (Strongly Agree) to 5 (Strongly Disagree). The average responses were, respectively, 2.29, and 2.31. The correlations with the client's changing preferences were, respectively, -0.276 ($p<.001$), and 0.290 ($p<.001$). The more the client changed his/her preferences, the less likely was the realtor to only recommend properties that matched their stated needs, and the more likely to include some properties to help them learn what they wanted. As shown in Proposition 3, if preferences were static, the best recommendation would be the product that meets the stated needs.

8 Model Extensions

In sections 5 and 6 we demonstrated that constructed preferences explain consumer and sales agent behavior observed in real markets, and have important implications for selling. In the interest of isolating the effect of constructed preferences on search, we abstracted from many complications that exist in real markets. Behavior in any real market is driven by many phenomena, beyond those isolated in this study. Which phenomena are more prevalent in which markets is an empirical question. In this section we outline several alternative explanations for the observed behavior, and propose critical experiments to understand which behaviors are prevalent in specific markets.

Consumer Uncertainty about Attributes We assume a high information environment, in which consumers can obtain information about product attributes at no cost, and the only information they gain as they search is better knowledge of their own preferences. In most real product search environments, both types of learning goes on: by evaluating a product consumers gain information about some of its attributes. The degree to which this assumption dominates depends on the product category: attributes of automobiles or electronics are relatively easy to find without examining actual products; attributes of clothing are more difficult to identify. A potential way to establish that preference construction is present in addition to attribute learning is to measure preferences before and after seeing a product that is, for example, poor on an attribute that the subject did not indicate was undesirable. If consumers are initially presented with all the attributes, and asked to state their preferences, then a change in preferences after evaluating the product should be attributed to preference construction, rather than just learning attribute information.

Salesforce Incentives The proposed model assumes that sales people are trying to maximize consumers' long term satisfaction with the purchase. This approximates reality in some selling environments, such as a very competitive real estate market, or recommender systems for subscription-based services. Consider, for example, Netflix, a subscription-based online movie provider. To maximize profits, Netflix needs its customers to consume many movies, and enjoy them, such that they do not cancel the subscription. Its profits are not affected by which movies the customers watch, and thus incentives are aligned. However, many sales environments, particularly those in which sales people are employed by the manufacturer, violate this assumption. For example, car salesmen are only interested in consumers buying their manufacturers' cars. We expect that constructed preferences will affect sales agent behavior even if incentives are not aligned. For example, misaligned incentives may drive sales agents to push buyers to purchase products that generate a higher commission, and thus to recommend suboptimal properties to clients, as we observe. The clients, in turn, may infer that this is going on, therefore complicating the interaction even more. When the realtor market is very competitive and it is easy for an unhappy client to go to a different realtor, the effect of realtor commissions is diminished. I chose to collect qualitative data in a market that is very competitive for realtors in order to minimize this effect. It is possible to fully isolate the effect of constructed preferences on buyers' interaction with sales people from the sales people's incentives in an experiment. For example, an experiment might present a subject with a task to sell something to another subject, and compensate the seller proportionally to the buyer's satisfaction.

Realtor Needs to Learn Preferences The model assumes that consumers are able to articulate their perceived preferences to the sales agent, and the sales agent can anticipate the changes in preferences in response to visiting homes. Of course in most real markets, there is some uncertainty on the sales agent's part as well, about both the perceived and true preferences, which will affect the recommendations. Uncertainty in perceived preferences might arise due to a communication problem: the buyer may not be able to accurately communicate the preferences to the realtor. However, since visiting homes is costly for both the buyer and realtor, it is in both of their best interests for the realtor to understand as much about the buyer's preferences as feasible. Qualitative interviews support this assumption. One realtor explained, "I want to know as much as I can before we even leave my office. Because otherwise we are walking, walking, walking, and we are just discovering what the person wants." Uncertainty about true preferences is probably more commonplace, as the sales agent may not be able to perfectly infer true preferences from the buyer's statements and other known consumer characteristics. The realtor is then solving a dynamic exploration/exploitation problem, which is an interesting extension to the current research.

9 Contribution and Placement

The goal of this work is to offer an explanation for why consumer product search involves viewing multiple products, even in the presence of sales agents. Our explanation is based on the dynamic nature of consumer preferences. As they search for a home, many people learn not only what products are available on the market, but also what their own preferences are. This work extends the consumer search literature by introducing sales agents and preference dynamics. We focus on the real estate industry to motivate the analysis, but the model applies to other high involvement consumer products, such as automobiles, and a broad spectrum of B2B sales, such as IT solutions and advertising tools. We demonstrate how dynamic preferences can result in a long search process even when all product attributes are easily searched. To the best of our knowledge, this is the first attempt to model dynamic preferences as they relate to a sales agent's recommendations. Behavioral researchers have studied the nature of preference construction extensively. Practitioners are aware of it, and agree that it is important and relevant to their strategy. This study is a step towards formalizing the process of preference evolution during the search process and exploring its implications.

References

Adomavicius, Gedimanas, and Alexander Tuzhilin (2005), "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions." *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6

Star, Steven (1989), "Marketing and Its Discontents," Harvard Business Review (November-December), 148 -54.

Brucks, Merrie (1985), "The Effects of Product Class Knowledge on Information Search Behavior," *Journal of Consumer Research*, 12 (June), pp. 1-16.

Branco, Fernando, Monic Sun, J. Miguel Villas-Boas (forthcoming), "Optimal Search for Product Information," *Management Science*, forthcoming

Coughlan, Anne T. and Subrata K. Sen (1989) "Salesforce Compensation: Theory and Managerial Implications," *Marketing Science*, vol. 8, no. 4 (Fall), pp. 324-342.

Culler, Leah L. "Forget A Dream Home – Aim For What's Practical." *MSN Real Estate*. 15 Mar 2012. <http://realestate.msn.com/forget-a-dream-home-aim-for-what-is-practical>

Chu, Wujin, Eitan Gerstner and James D. Hess (1995) "Costs and Benefits of a Hard-Sell

Feldman, Jack M. and John G. Lynch Jr. (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," Journal of Applied Psychology, 73 (3), 421–35

Ferguson Thomas S. (1989), "Who solved the secretary problem?" *Statistical Science*, Volume 4, Number 3, pp 282–296.

Hauser, John R., Glen L. Urban, and Bruce Weinberg (1993), "How Consumers Allocate their Time When Searching for Information," *Journal of Marketing Research,*30, 4, (November), 452-466.

Slovic, P., Griffin,D., and Tversky, A. 1990. Compatibility effects in judgment and choice. In Hogarth, R. (ed.) Insights in decision making: Theory and applications. Chicago: University
of Chicago Press.

Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, Gerald L. Lohse (2004), "On the Depth and Dynamics of Online Search Behavior," Management Science Vol. 50, No. 3, March 2004, pp. 299–308

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2010), "Online Demand under Limited Consumer Search," Marketing Science, Vol. 29, No. 6, November-December, pp. 1001-1023.

Moorthy, Sridhar , Ratchford, Brian T. and Talukdar, Debabrata (1995), "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, Vol. 23, No. 4, pp. 263-277.

Morwitz, Vicki, Eric J. Johnson and David Schmittlein (1993), "Does Measuring Intent Change Behavior?" Journal of Consumer Research, Vol. 20, pp. 46-66.

Nordgren, Loran F and Ap Dijksterhuis (2009), "**The Devil Is in the Deliberation: Thinking Too Much Reduces Preference Consistency**", *Journal of Consumer Research* Vol. 36, No. 1 (June 2009), pp. 39-46.

Sharma, Arun (1997), "Customer Satisfaction-Based Incentive Systems: Some Managerial and Salesperson Considerations," *Journal of Personal Selling and Sales Management*, 17, 2 (Spring), 61-70.

Richins, Marsha L. (1983), "Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study," *The Journal of Marketing*, Vol. 47, No. 1 (Winter), pp. 68-78.

Summerfield, Brian. "What Customer Service Means Today." *Realtor Magazine*. 03 Feb. 2012. < http://realtormag.realtor.org/daily-news/2012/02/03/what-customer-service-means-today&WT.cg_n=RMO&WT.cg_s=RSSDaily>

Ying, Yuanping, Fred Feinberg, Michel Wedel (2006), "Improving Online Product Recommendations by Including Nonrated Items," *Journal of Marketing Research*, 43 (August), 355-365

APPENDIX 1 – PROOFS OF PROPOSITIONS 1 & 2 – UNINTERRUPTED SEARCH

Proposition 1: If $\vec{q}_{i0} = \vec{p}_i$, or the consumer has perfect knowledge of his preferences, he finds the utility maximizing option at $t = 1$, and looks no further after the verification stage. This scenario generates the highest possible payoff for the consumer:

$$U_i^* = \frac{(\vec{x}_i^* \vec{p}_i - c_s) \cdot e^{\vec{x}_i^* \vec{p}_i} + B \cdot e^B}{e^B + e^{\vec{x}_{\bar{k}} \vec{q}_{i\bar{t}}}}$$

Proof: At $t = 0$, the consumer chooses $\vec{x}_0^* = \arg\max_k \vec{x}_k \vec{q}_{i0} = \arg\max_k \vec{x}_k \vec{p}_i = \vec{x}_i^*$. After examining it, the preferences do not change, so $\vec{q}_{i1} = \vec{q}_{i0} = \vec{p}_i$, and

$$\vec{x}_1^* = \arg\max_k \vec{x}_k \vec{q}_{i1} = \arg\max_k \vec{x}_k \vec{q}_{i0} = \vec{x}_0^*.$$

The consumer stops searching, and purchases the product with probability $\frac{e^{\vec{x}_i^* \vec{p}_i}}{B + e^{\vec{x}_{\bar{k}} \vec{q}_{i\bar{t}}}}$, otherwise takes the outside good.

Proposition 2: (Local maxima rather than global maximum) There exist values of $\vec{p}_i$ and $\vec{q}_{i0}$, for which an uninterrupted search process will stop at some product before finding $\vec{x}_i^*$.

Proof: Suppose there are two attributes (brand and car body type), two levels each (Toyota, Subaru; Sedan, SUV). Suppose the consumer's true preferences and perceived preferences are:

	Toyota	Subaru	Sedan	SUV
$\vec{w}_i$	3	5	2	3
$\vec{v}_{i0}$	3	0	2	3

Then, the consumer chooses $\vec{x}_{i1}^* = [1\ 0\ 0\ 1]$, the Toyota SUV at time $t = 1$. After inspecting it, preferences do not change, since $\vec{v}_{i0}$(Toyota, Sedan) = $\vec{w}_i$(Toyota, Sedan). The consumer was unaware of the Subaru brand, and remains unaware. The new utility maximizing product is the same as $\vec{x}_{i1}^*$, so the consumer stops searching, even though $\vec{x}_{i1}^* \neq \vec{x}_i^*$.

APPENDIX 2 – SIMULATIONS

Generating Data

Product Attributes. We used a product space with 4 attributes, 3 levels each, for a total of 12 aspects. First, we generate a full factorial design of all 81 possible combinations. To account for the fact that in most real markets some aspects are anti-correlated (e.g. homes in a certain neighborhood are old, cars with large engines have low fuel efficiency, Porsche does not make pickup trucks, etc.) we create an "anti-correlation" matrix, C. For each pair of aspects (i, j) a product that contains both attributes does not exist (gets deleted from the full factorial matrix). Thus, the probability that a given product gets deleted is the product of $C(i, j)$ for all pairs of aspects in it. C is a symmetric matrix, with values drawn from *Beta*(0.1, 0.8). This distribution is U-shaped, meaning most pairs of attributes are unlikely to lead to deletion, but some pairs lead to deletion with a very high probability.

Consumers. For each of 5,000 simulated consumers, partworths of both the true and perceived utilities are drawn from a uniform [0,1] distribution. Search cost was varied between 0 and 3; the value of the outside good is set at $B = 2$.

Simulating Search

Basic uninterrupted search is simulated as described in Section 4.

Recommendations can be made to the uninterrupted search at time $t = 1$. We find the truly optimal recommendation by solving the forward-looking problem that accounts for all future changes to the perceived utility and search costs. For every possible product recommendation (all existing products), we compute the resulting search path and the net payoff. Then we select the recommendation that maximizes the net payoff at the end of the search

Summary of Results

The average utility of the best product, $\vec{x}_i^*$, is 2.85

With search cost of 0.1, 24% find $\vec{x}_i^*$ on their own; 99% find $\vec{x}_i^*$ with a product recommendation

The average net payoff, including search cost, is 1.9 when consumer searches on his own; 2.42 when a recommendation is made at time $t = 1$.

For 14% of people, the optimal recommendation is NOT $\vec{x}_i^*$