Disjunctions of Conjunctions, Cognitive Simplicity and Consideration Sets

by

John R. Hauser
Olivier Toubia
Theodoros Evgeniou
Rene Befurt
Daria Silinskaia

March 2009

Forthcoming, Journal of Marketing Research

John R. Hauser is the Kirin Professor of Marketing, MIT Sloan School of Management, Massachusetts Institute of Technology, E40-179, One Amherst Street, Cambridge, MA 02142, (617) 253-2929, fax (617) 253-7597, jhauser@mit.edu.

Olivier Toubia is the David W. Zalaznick Associate Professor of Business, Columbia Business School, Columbia University, 522 Uris Hall, 3022 Broadway, New York, NY, 10027, (212) 854-8243, ot2107@columbia.edu.

Theodoros Evgeniou is an Associate Professor of Decision Sciences and Technology Management, INSEAD, Boulevard de Constance 77300, Fontainebleau, FR, (33) 1 60 72 45 46, theodoros.evgeniou@insead.edu.

Rene Befurt is an Associate at the Analysis Group, 111 Huntington Avenue, Tenth Floor, Boston, MA 02199, 617-425-8283, RBefurt@analysisgroup.com.

Daria Silinskaia is a doctoral student at the MIT Sloan School of Management, Massachusetts Institute of Technology, E40-170, One Amherst Street, Cambridge, MA 02142, (617) 253-2268, dariasil@mit.edu.

We would like to thank Daniel Bailiff (AMS), Simon Blanchard (PSU), Robert Bordley (GM), Anja Dieckmann (GfK), Holger Dietrich (GfK), Min Ding (PSU), Steven Gaskin (AMS), Patricia Hawkins (GM), Phillip Keenan (GM), Clarence Lee (MIT), Carl Mela (Duke), Andy Norton (GM), Daniel Roesch (GM), Matt Selove (MIT), Glen Urban (MIT), Limor Weisberg and Kaifu Zhang (INSEAD) for their insights, inspiration, and help on this project. This paper has benefited from presentations at the Analysis Group Boston, the Columbia Business School, Digital Business Conference at MIT, Duke University, General Motors, the London Business School, Northeastern University, the Marketing Science Conference in Vancouver, B.C., and the Seventh Triennial Choice Symposium at the University of Pennsylvania. Upon publication data are available from the authors.
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, $\ell$ 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., $LF$). Our models and estimation can (and do) handle such dependency, but the notation is cumbersome. Let $x_{j\ell} = 1$ if profile $j$ has feature $f$ at level $\ell$. Otherwise $x_{j\ell} = 0$. Let $\tilde{x}_j$ 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 $\tilde{y}_h$ 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 fea-
tures is at an “acceptable” (or satisfactory) level. Let $a_{h\ell} = 1$ if level $\ell$ of feature $f$ is acceptable to respondent $h$. Otherwise, $a_{h\ell} = 0$. Let $\tilde{a}_h$ be the binary vector of acceptabilities for respondent $h$. A disjunctive rule states that respondent $h$ considers profile $j$ if $\tilde{x}_j^t \tilde{a}_h \geq 1$.

**Conjunctive rules.** In a conjunctive rule, a profile is considered if all 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 $\tilde{x}_j^t \tilde{a}_h = F$.

**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 $\tilde{x}_j^t \tilde{a}_h \geq S$. Clearly, a disjunctive rule is a special case where $S = 1$ and, because $\tilde{x}_j^t \tilde{a}_h$ 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 $\tilde{\beta}_h$ is the vector of partworths for respondent $h$, then $h$ considers profile $j$ if $\tilde{x}_j^t \tilde{\beta}_h \geq T_h$. 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 spec-
ify 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 either 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: \((\text{B&W screen} \land \text{small size} \land \text{high resolution}) \lor (\text{color screen} \land \text{large size})\), where \(\land\) is the logical “and” and \(\lor\) 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: \((\text{B&W screen} \land \neg \$399)\).

**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 \((\text{B&W screen} \land \neg \$399)\). We define the size, \(s\), of a pattern as the number of aspects in the pattern. For example, \((\text{B&W screen} \land \neg \$399)\) 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 \(S\)).\(^1\)

\(^1\) 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 \(S\).
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 \(\tilde{w}_h\) and \(\tilde{m}_j\) 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 \(\tilde{m}_j' \tilde{w}_h \geq 1\).

**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 \(\text{DOC}(S)\) 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 \((y_{jh})\). 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 \(\tilde{m}_j' \tilde{w}_h \geq 1\) and not considered if \(\tilde{m}_j' \tilde{w}_h = 0\). (Recall \(\tilde{m}_j\) and \(\tilde{w}_h\) 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, \( \tilde{w}_h \), 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, $\tilde{w}_h$, 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, $FP_h(\tilde{w}_h)$, as the number of profiles predicted to be considered but observed as not considered and we define false negatives, $FN_h(\tilde{w}_h)$, 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 $\tilde{w}_h$ 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 propor-
tional to the sum of false positives and false negatives in the sample (sum over all respondents).

Formally, our objective function is:

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

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 \( \tilde{w}_h \) to be continuous on \([0, 1]\), then round up any positive \( w_{hi} \) that is above a threshold (Hastie, Tibshirani, 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: \( \gamma_M \) tells us how much to penalize lack of sample-level fit, \( \gamma_c \) 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 \( \tilde{w}_h \), 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 $\gamma_M$ that is an arbitrarily small number such that sample-level consideration is used only to break ties among patterns. For $\gamma_c$, cross-validation (and predictive tests) vary little in the range $\gamma_c \in [1, 4.5]$. 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 patterns, 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. In addition to detailed formulations, the Web Appendix also contains simulations which compare some of the benchmarks to DOC-based methods on synthetic data.

**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.]

Using the sixteen features we generated an orthogonal design of 32 GPS profiles. 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

---

2 The basic relative comparisons with DOC-based models are similar for $S \sim 1, 2, 3, \text{ or } 4$.
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.
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.
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’ 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 incen-
tive 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, and 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). 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 and 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 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 among the benchmarks themselves. Our data are available for readers who wish to explore machine learning, HB, or other methods for the benchmark rules.

**Empirical Evidence is Consistent with Cognitive Simplicity**

Although DOCMP and LAD-DOC are designed to favor cognitive simplicity, uncon-

---

\(^6\) 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.
strained estimation could conceivably predict better. We re-estimated DOCMP with the $\gamma$s 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 $\gamma$s; $p < 0.01$ 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 patterns ($P$). 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 re-
jecting 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. 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

---

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

**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 represen-
tations 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


Ding, Min (2007), “An Incentive-Aligned Mechanism for Conjoint Analysis,” *Journal of Mar-


FIGURE 1
FEATURES OF HANDHELD GPSS

- Screen Color (color/monochrome)
- Screen Size (big/small)
- Brand (Garmin/Magellan)
- Weight (4 ounces/7 ounces)
- Display Brightness (normal/extra bright)
- Display Resolution (high/low)
- Acquisition Time (2 sec/10 sec)
- Battery Life (30 hours/12 hours)
- Reception (average/under trees)
- Accuracy (50 feet/a few feet)
- Track Log (retraces route) (yes/no)
- Mini-USB port (yes/no)
- Floats on Water (yes/no)

Price: $299
Size of GPS (big/small)

FIGURE 2
CONSIDERATION TASK IN ONE OF THE FORMATS (ENGLISH AND GERMAN)
<table>
<thead>
<tr>
<th>Estimation method</th>
<th>Overall hit rate (%</th>
<th>Relative hit-rate improvement (%)</th>
<th>K-L divergence percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical Bayes Benchmarks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conjunctive (S = 16)</td>
<td>77.7</td>
<td>35.6</td>
<td>6.2</td>
</tr>
<tr>
<td>Disjunctive (S = 1)</td>
<td>66.7</td>
<td>3.8</td>
<td>17.8</td>
</tr>
<tr>
<td>Subset Conjunctive (S = 4)</td>
<td>75.4</td>
<td>29.0</td>
<td>24.7</td>
</tr>
<tr>
<td>(q)-Compensatory</td>
<td>73.4</td>
<td>37.6</td>
<td>14.6</td>
</tr>
<tr>
<td>Additive</td>
<td>78.5</td>
<td>38.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Machine-Learning Benchmarks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conjunctive (S = 16)</td>
<td>52.6</td>
<td>-36.8</td>
<td>13.3</td>
</tr>
<tr>
<td>Disjunctive (S = 1)</td>
<td>77.5</td>
<td>35.6</td>
<td>8.1</td>
</tr>
<tr>
<td>Subset Conjunctive (S = 4)</td>
<td>73.7</td>
<td>24.3</td>
<td>6.3</td>
</tr>
<tr>
<td>(q)-Compensatory</td>
<td>76.2</td>
<td>31.3</td>
<td>6.3</td>
</tr>
<tr>
<td>Additive</td>
<td>80.6</td>
<td>44.0</td>
<td>23.0</td>
</tr>
<tr>
<td>DOC-Based Estimation Methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOCMP (S = 4)</td>
<td>81.9*</td>
<td>47.8*</td>
<td>32.0*</td>
</tr>
<tr>
<td>LAD-DOC (S = 4, (P = 2))</td>
<td>82.2*</td>
<td>48.6*</td>
<td>34.6*</td>
</tr>
</tbody>
</table>

\(\text{a Hit rate is the number of profiles predicted correctly, divided by 32.}\)

\(\text{* Best or not significantly different than the best at the 0.05 level.}\)
THE FOLLOWING SUPPLEMENTS ARE AVAILABLE IN A WEB APPENDIX

Theme 1. Summary of Notation and Acronyms.

Theme 2. Proofs to the Results that DOC Decision Rules Nest Conjunctive, Disjunctive, and Subset Conjunctive Decision Rules.

Theme 3. HB Estimation of the Conjunctive, Disjunctive, Subset Conjunctive, Additive, and $q$-Compensatory Models.

Theme 4. Integer Programming Estimation of the DOC, Conjunctive, Disjunctive, Subset Conjunctive, Additive, and $q$-Compensatory Models.

Theme 5. Kullback-Leibler Divergence Formulae for Consideration Sets.


Theme 7. Synthetic-Data Experiments.

Theme 8. Results for Alternative Formats.

Theme 9. Results US Sample “Evaluate all Profiles”

Theme 10. Results for Text-only Format.


Theme 12. Consider-only, Reject-only, No-browsing, Text-only, Example Feature-Introduction, and Instruction Screenshots.