Disjunctions of Conjunctions, Cognitive Simplicity and Consideration Sets

Web Appendix

by

John R. Hauser

Olivier Toubia

Theodoros Evgeniou

Rene Befurt

Daria Silinskaia

March 2009

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 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.

THEME 1: SUMMARY OF NOTATION AND ACRONYMS
(Some Notation is Used Only in This Web Appendix)

$a_{hf\ell}$ binary indicator of whether level ℓ of feature f is acceptable to respondent h (disjunctive, conjunctive, or subset conjunctive models, use varies by model)

$\vec{a}_h$ binary vector of acceptabilities for respondent h

b_1, b_2 parameters of the HB subset conjunctive model, respectively, the probability that a profile is considered if $\vec{x}_j'\vec{a}_h \geq S$ and the probability it is not considered if $\vec{x}_j'\vec{a}_h < S$

$\vec{e}$ a vector of 1's of length equal to the number of potential patterns

D covariance matrix used in estimation HB compensatory

f indexes features, F is the total number of features

h indexes respondents (mnemonic to households), H is the total number of respondents

I the identity matrix of size equal to the total number of aspects

j indexes profiles, J is the total number of profiles

ℓ indexes levels within features, L is the total number of levels

m_{jp} binary indicator of whether profile j matches pattern p

$\vec{m}_j$ binary vector describing profile j by the patterns it matches

n size of the consideration set

M_j percent of respondents in the sample ("market") that consider profile j

p indexes patterns; also used for significance level in t-tests when clear in context

P maximum number of patterns [LAD-DOC(P, S) estimation]

Q number of partworths (compensatory model)

s size of a pattern (number of aspects in a conjunction)

S maximum subset size [Subset(S) model] or maximum number of aspects in a conjunctive pattern [DOC(S) model, LAD-DOC(P, S) estimation]

T_h threshold for respondent h in compensatory model

w_{hp} binary indicator of whether respondent h considers profiles with pattern p

$\vec{w}_h$ binary vector indicating the patterns used by respondent h

$x_{jf\ell}$ binary indicator of whether profile j has feature f at level ℓ

$\vec{x}_j$ binary vector describing profile j

y_{hj} binary indicator of whether respondent h considers profile j

$\vec{y}_h$ binary vector describing respondent h's consideration decisions

$\vec{\beta}_h$ vector of partworths (compensatory model) for respondent h

ε_{hj} extreme value error in compensatory model

γ_C, γ_M parameters penalizing, respectively, complexity and deviation from the "market"

ξ_{hj}^{+} non-negative integer that indicates a model predicts consideration if $\xi_{hj}^{+} \geq 1$

ξ_{hj}^{-} non-negative integer that indicates a model predicts non-consideration if $\xi_{hj}^{-} \geq 1$

DOC(S) set of disjunctions of conjunctions models. S, when indicated, is the maximum size of the patterns.

DOCMP combinatorial optimization estimation for DOC models

LAD-DOC alternative estimation method for DOC models in which we limit both the number of patterns, P, and the size of the patterns, S

Subset(S) set of subset conjunctive models with maximum subset size of S

THEME 2: PROOFS TO FORMAL RESULTS THAT DISJUNCTION-OF-CONJUNCTION DECISION RULES NEST OTHER NON-COMPENSATORY DECISION RULES

Result 1. The following sets of rules are equivalent (a) disjunctive rules, (b) Subset(1)rules, and (c) DOC(1) rules.

Proof. A disjunctive rule requires $\vec{x}_j'\vec{a}_h \geq 1$; a Subset($S$) rule requires $\vec{x}_j'\vec{a}_h \geq S$; a DOC($S$) rule requires $\vec{m}_j'\vec{w}_h \geq 1$. Clearly the first two rules are equivalent with $S = 1$. For DOC(1) recognize that all patterns are single aspects hence $\vec{m}_j$ and $\vec{w}_h$ correspond one-to-one with aspects and $\vec{m}_j$ can be recoded to match $\vec{x}_j$ and $\vec{w}_h$ can be recoded to match $\vec{a}_h$.

Result 2. Conjunctive rules are equivalent to Subset(F) rules which, in turn, are a subset of the DOC(F) rules, where F is the number of features.

Proof. A conjunctive rule requires $\vec{x}_j'\vec{a}_h = F$ Setting $S = F$ establishes the first statement. The second statement follows directly from Result 3 with $S = F$.

Result 3. A Subset(S) rule can be written as a DOC(S) rule, but not all DOC(S) rules can be written as a Subset(S) rule.

Proof. $\vec{x}_j'\vec{a}_h \geq S$ holds if any S aspects are acceptable. Therefore $\vec{x}_j$ must match at least one pattern of length S. Let Σ_S be the set of such patterns, then $\vec{x}_j$ matches at least one element of Σ_S. Consider the DOC(S) rule defined by $w_{hj} = 1$ for any pattern in Σ_S. The inequality $\vec{x}_j'\vec{a}_h \geq S$ holds if and only if $\vec{m}_j'\vec{w}_h \geq 1$, establishing that Subset(S) can be written as a DOC(S) rule. By definition, a DOC(S) rule also includes patterns of size less than S, hence, $\vec{x}_j'\vec{a}_h < S$ for some DOC(S) rules. This establishes the second statement.

Result 4. Any set of considered profiles can be fit perfectly with at least one DOC rule. Moreover, the DOC rule need not be unique.

Proof. For each considered profile, create a pattern of size *F* that matches that profile. This pattern will not match any other profile because *F* aspects establishes a profile uniquely. Create $\vec{w}_h$ such that $w_{hj} = 1$ for all such profiles and $w_{hj} = 0$ otherwise. Then $\vec{m}_j'\vec{w}_h = 1$ if profile *j* is considered and $\vec{m}_j'\vec{w}_h = 0$ otherwise. The second half of the proof is established by the examples in the text which establishes the existence of non-unique DOC rules.

THEME 3: HB ESTIMATION OF THE SUBSET CONJUNCTIVE, ADDITIVE, AND *q*-COMPENSATORY MODELS

Subset Conjunctive Model (includes Disjunctive and Conjunctive)

All posterior distributions are known, hence we use Monte Carlo Markov chains (MCMC) with Gibbs sampling. Recall that *S* is fixed.

$\Pr(a_{hf\ell} \mid y_{hj}'s, other\ \vec{a}_h's, \theta_{f\ell}'s, b_1, b_2)$. We follow Gilbride and Allenby (2004, p. 404) and use a "Griddy Gibbs" algorithm. For each *h* we update the acceptabilities, $a_{hf\ell}$, aspect by aspect. For each candidate set of acceptabilities we compute the likelihood as if we kept all other acceptabilities constant replacing only the candidate $a_{hf\ell}^c$. The likelihood is based on Equation 5 and the prior on the $\theta_{f\ell}$'s. The probability of drawing $a_{hf\ell}^c$ is then proportional to the likelihood times the prior summed over the set of possible candidates.

$\Pr(\theta_{fl} \mid y_{hj}'s, \vec{a}_h's, b_1, b_2)$. The $\theta_{f\ell}$'s are drawn successively, hence we require the marginal of the Dirichlet distribution – the beta distribution. Because the beta distribution is conjugate to the binomial likelihood, we draw $\theta_{hf\ell}$ from $Beta[6 + \sum_h a_{hf\ell}, 6 + \sum_h (1 - a_{hf\ell})]$.

$\Pr(b_1, b_2 \mid y_{hj}'s, \vec{a}_h's, \theta_{f\ell}'s)$. Because the beta distribution is conjugate to the binomial likelihood, we draw b_1 from $Beta[1 + \sum_{h,j} y_{hj}\delta(\vec{x}_j'\vec{a}_h \geq S), \sum_{h,j}(1 - y_{hj})\delta(\vec{x}_j'\vec{a}_h \geq S)]$ and we draw b_2 from $Beta[1 + \sum_{h,j} y_{hj}\delta(\vec{x}_j'\vec{a}_h < S), \sum_{h,j}(1 - y_{hj})\delta(\vec{x}_j'\vec{a}_h < S)]$, where $\delta(\bullet)$ is the indicator function.

For the disjunctive model we set $S = 1$; for the fully conjunctive model we set $S = 16$, and for the subset conjunctive model we set $S = 4$.

Additive Model

Respondent *h* considers profile *j* if $\vec{x}_j'\vec{\beta}_h + \varepsilon_{hj}$ is above a threshold. Subsuming the threshold in the partworths, we get a standard logit likelihood function:

$$\Pr(y_{hj} = 1 \mid \vec{x}_j, \vec{\beta}_h) = \frac{e^{\vec{x}_j'\vec{\beta}_h}}{1 + e^{\vec{x}_j'\vec{\beta}_h}}$$

$\Pr(y_{hj} = 0 \mid \vec{x}_j, \vec{\beta}_h) = 1 - \Pr(y_{hj} = 1 \mid \vec{x}_j, \vec{\beta}_h)$. We impose a first-stage prior on $\vec{\beta}_h$ that is normally distributed with mean $\vec{\beta}_0$ and covariance *D*. The second stage prior on *D* is inverse-Wishart with parameters equal to $I/(Q+3)$ and $Q+3$, where *Q* is the number of parameters to be estimated and *I* is an identity matrix. We use diffuse priors on $\vec{\beta}_0$. Inference is based on a Monte Carlo Markov chain with 20,000 iterations, the first 10,000 of which are used for burn-in.

q-Compensatory Model

Estimation is the same as in the additive model except we use rejection sampling to enforce the constraint that the importance on any feature is no more than *q* times as large as any other feature.

References for Theme 3

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

THEME 4: INTEGER PROGRAMMING ESTIMATION OF THE DOC, SUBSET CONJUNCTIVE, ADDITIVE, AND Q-COMPENSATORY MODELS

All mathematical programs were formulated to be as similar as feasible to DOCMP. CompMP and SubsetMP can be simplified with algebraic substitutions. We subsume the threshold in the partworths estimated by CompMP. We set K to a number that is large relative to T_h. For comparability and to be conservative, we set $S = 4$ in SubsetMP. For disjunctive we set $S = 1$ and for fully conjunctive we set $S = 16$.

DOCMP: $$\min_{\{\vec{w}_h, \vec{\xi}_h\}} \sum_{j=1}^{J} [y_{hj}\xi_{hj}^{-} + (1 - y_{hj})\xi_{hj}^{+}] + \gamma_M \sum_{j=1}^{J} [M_j \xi_{hk}^{-} + (1 - M_j)\xi_{hj}^{+}] + \gamma_c \vec{e}'\vec{w}_h$$

Subject to: $\vec{m}_j'\vec{w}_h \le \xi_{hj}^{+}$ for all $j = 1$ to J

$\vec{m}_j'\vec{w}_h \ge 1 - \xi_{hj}^{-}$ for all $j = 1$ to J

$\xi_{hj}^{+},\ \xi_{hj}^{-} \ge 0$, $\vec{w}_h$ a binary vector

Allowable patterns have length at most S.

SubsetMP: $$\min_{\{\vec{a}_h, \vec{\xi}_h, S\}} \sum_{j=1}^{J} [y_{hj}\xi_{hj}^{-} + (1 - y_{hj})\xi_{hj}^{+}] + \gamma_M \sum_{j=1}^{J} [M_j \xi_{hk}^{-} + (1 - M_j)\xi_{hj}^{+}] + \gamma_c S$$

Subject to: $\vec{x}_j'\vec{a}_h \le S\xi_{hj}^{+}$ for all $j = 1$ to J

$\vec{x}_j'\vec{a}_h \ge S(1 - \xi_{hj}^{-})$ for all $j = 1$ to J

$\xi_{hj}^{+},\ \xi_{hj}^{-} \ge 0$, $\vec{a}_h$ a binary vector, $S > 0$, integer

CompMP: $\min_{\{\vec{\beta}_h, \vec{\xi}_h\}} \sum_{j=1}^{J}[y_{hj}\xi_{hj}^{-} + (1-y_{hj})\xi_{hj}^{+}] + \gamma_M \sum_{j=1}^{J}[M_j\xi_{hk}^{-} + (1-M_j)\xi_{hj}^{+}] + \gamma_c \vec{e}'\vec{\beta}_h$

Subject to: $\vec{x}_j'\vec{\beta}_h \le T_h + K\xi_{hj}^{+}$ for all j = 1 to J

$\vec{x}_j'\vec{\beta}_h \ge T_h(1-\xi_{hj}^{-})$ for all j = 1 to J

$\xi_{hj}^{+},\ \xi_{hj}^{-} \ge 0, \quad \vec{\beta}_h \ge 0$

CompMP(q): $\min_{\{\vec{\beta}_h, \vec{\xi}_h\}} \sum_{j=1}^{J}[y_{hj}\xi_{hj}^{-} + (1-y_{hj})\xi_{hj}^{+}] + \gamma_M \sum_{j=1}^{J}[M_j\xi_{hk}^{-} + (1-M_j)\xi_{hj}^{+}] + \gamma_c \vec{e}'\vec{\beta}_h$

Subject to: $\vec{x}_j'\vec{\beta}_h \le T_h + K\xi_{hj}^{+}$ for all j = 1 to J

$\vec{x}_j'\vec{\beta}_h \ge T_h(1-\xi_{hj}^{-})$ for all j = 1 to J

$\max_\ell\{\beta_{f\ell}\} - \min_\ell\{\beta_{f\ell}\} \le q[\max_\ell\{\beta_{n\ell}\} - \min_\ell\{\beta_{n\ell}\}]$ for all f, n

$\xi_{hj}^{+},\ \xi_{hj}^{-} \ge 0, \quad \vec{\beta}_h \ge 0$

THEME 5: KULLBACK-LEIBLER DIVERGENCE FOR CONSIDERATION DATA

To describe this statistic, we introduce additional notation. Let q_j be the null probability that profile j is considered and let r_j be the probability that profile j is considered based on the model and the observations. The K-L divergence for respondent h is $\sum_j \{r_j \ln[r_j / q_j] + (1-r_j)\ln[(1-r_j)/(1-q_j)]\}$. To use the K-L divergence for discrete predictions we let z_{hj} and $\hat{z}_{hj}$ be the indicator variables for validation consideration, that is, z_{hj} =1 if respondent h considers profile j and $\hat{z}_{hj}$ = 1 if respondent h is predicted to consider profile j. They are zero otherwise. Let $C_e = \sum_j y_{hj}$ be the number of profiles considered in the estimation task. Let $C_v = \sum_j z_{hj}$ and $\hat{C}_v = \sum_j \hat{z}_{hj}$ be corresponding observed and predicted numbers for the validation task. Let $F_n = \sum_j z_{hj}(1-\hat{z}_{hj})$ be the number of false negatives (observed as considered but predicted as not considered) and $F_p = \sum_j (1-z_{hj})\hat{z}_{hj}$ be the number of false positives (observed as not considered but predicted as considered). (F_n and F_p are not to be confused with F, the number of features as used in the text.) Substituting, we obtain the K-L divergence for a model being evaluated. The second expression expands the summations and simplifies the fractions.

$$\text{K-L divergence} = \sum_{j:\hat{z}_{hj}=1}\left[\frac{\hat{C}_v - F_p}{\hat{C}_v}\ln\frac{\frac{\hat{C}_v - F_p}{\hat{C}_v}}{C_e/J} + \frac{F_p}{\hat{C}_v}\ln\frac{\frac{F_p}{\hat{C}_v}}{(J-C_e)/J}\right] + \sum_{j:\hat{z}_{hj}=0}\left[\frac{F_n}{J-\hat{C}_v}\ln\frac{\frac{F_n}{J-\hat{C}_v}}{C_e/J} + \frac{J-\hat{C}_v-F_n}{J-\hat{C}_v}\ln\frac{\frac{J-\hat{C}_v-F_n}{J-\hat{C}_v}}{(J-C_e)/J}\right]$$

$$= (\hat{C}_v - F_p)\ln\frac{J(\hat{C}_v - F_p)}{\hat{C}_v C_e} + F_p\ln\frac{JF_p}{\hat{C}_v(J-C_e)} + F_n\ln\frac{JF_n}{(J-\hat{C}_v)C_e} + (J-\hat{C}_v-F_n)\ln\frac{J(J-\hat{C}_v-F_n)}{(J-\hat{C}_v)(J-C_e)}$$

The perfect-prediction benchmark sets $z_{hj} = \hat{z}_{hj}$, hence $F_n = F_p = 0$ and $\hat{C}_v = C_v$. The relative K-L divergence is the K-L divergence for the model versus the null model, divided by the K-L divergence for perfect prediction versus the null model.

THEME 6: GENERATION OF SYNTHETIC DATA FOR SIMULATION EXPERIMENTS

Compensatory Rules (First Simulations, Data Chosen To Favor HB Estimation)

We drew partworths from a normal distribution that was zero-mean except for the intercept. The covariance matrix was *I/2*. We adjusted the value of the intercept (to 1.5) such that respondents considered, on average, approximately 8 profiles. Profiles were identified as considered with Bernoulli sampling from logit probabilities.

Compensatory Rules (Second Simulations, Data Chosen To Favor Machine Learning Estimation)

Following Toubia, et. al. (2003) we drew partworths from a normal distribution with mean 50 and standard deviation 30, truncated to the range of [0, 100]. We adjusted the value of the intercept such that respondents considered, on average, approximately 8 profiles. Profiles were identified as considered if they passed the threshold with probability $b_1 = 0.99$. If they did not pass the threshold, they were considered with probability $b_2 = 0.01$. We enforced the *q*-compensatory constraint by rejection sampling.

Subset Conjunctive Rules

We drew each acceptability parameter from a binomial distribution with the same parameters for all features and levels. We adjusted the binomial probabilities such that respondents considered, on average, approximately 8 profiles. This gave us S 4. We set $b_1 = 0.95$ and $b_2 = 0.05$ in the first set of simulations and $b_1 = 0.99$ and $b_2 = 0.01$ in the second set of simulations.

Disjunctions of Conjunctions Rules

We drew binary pattern weights from a Dirichlet distribution adjusting the marginal binomial probabilities such that respondents considered, on average, approximately 8 profiles. This gave us 0.025, 0.018, and 0.017 for $S = 2$ to 4. We simulate consideration decisions such

that the probability of considering a profile with a matching pattern and the probability of considering a profile without a matching pattern is the same in the DOC rules as in the compensatory and subset conjunctive rules. In the first set of simulations we generated DOC rules for $S \sim$ 1, 2, 3, and 4 where $S = 1$ corresponds to disjunctive rules and $S = 4$ is similar to, but not identical to conjunctive rules. In the second set of simulations we focused on $S = 3$ for simplicity.

THEME 7: SYNTHETIC DATA EXPERIMENTS

The focus of our paper is on the predictive ability of DOC-based models for the empirical GPS data. One can also create synthetic respondents such that an estimation method that assumes a particular decision rule does well when data are generated by that decision rule. Because the synthetic data experiments are computational intensive we focus on key comparisons to provide initial perspectives. Our simulations are not designed to explore every one of the ten benchmarks in the paper. We encourage readers to explore synthetic-data experiments further.

Simulations 1. Synthetic Data Chosen to Favor Hierarchical Bayes Specifications

TABLE W1

OUT-OF-SAMPLE HIT RATE IN FIRST SIMULATIONS

(Each Estimation Method and Each Data-Generation Decision Rule)

	Hit Rate for Indicated Estimation Method (%)					
Data Generation Decision Rule	**HB Compensatory**	**HB Subset(1)**	**HB Subset(2)**	**HB Subset(3)**	**HB Subset(4)**	**DOCMP**
Compensatory	**74.6***	45.2	59.3	66.7	72.4	72.8
Subset(2)	78.5	71.1	**88.0***	85.4	80.3	84.5
Subset(3)	78.6	61.3	81.9	**87.2***	80.9	83.8
Conjunctive [Subset(4)]	78.7	60.3	80.7	87.1	**89.0***	**89.2***
Disjunctive [DOC(1), Subset(1)]	84.4	85.6	86.4	86.1	83.7	**90.8***
DOC(2)	77.6	70.6	76.1	78.6	78.8	**87.0***
DOC(3)	76.3	51.0	65.4	76.4	77.8	**83.3***
DOC(4)	74.8	53.7	65.8	75.0	76.9	**82.9***

*Best predictive hit rate, or not significantly different than the best at the 0.05 level, for that decision rule (row).

Simulations 2. Synthetic Data Chosen to Favor Machine Learning Specifications

TABLE W2

OUT-OF-SAMPLE HIT RATE IN SECOND SIMULATIONS

(Each Estimation Method and Each Data-Generation Decision Rule)

	Hit Rate for Indicated Estimation Method (%)						
Data Generation Decision Rule	**HB Add-itive**	**HB Conj-unctive**	**HB Disj-unctive**	**Comp-MP (*q* = *4*)**	**DOC-MP (*S* = 4)**	**LAD-DOC (∞,∞)**	**LAD-DOC (2, 4)**
Additive (*q* = 4)	80.6	83.3	74.5	81.5	**87.8**	83.8	79.2
Conjunctive	79.9	**87.9**	59.6	76.7	**88.6**	**87.1**	**87.3**
Disjunctive	79.1	58.8	**86.0**	78.7	82.1	83.5	80.3
DOC	82.1	85.5	69.2	82.1	**89.8**	**89.4**	**89.3**

TABLE W3

OUT-OF-SAMPLE K-L DIVERGENCE PERCENTAGE IN SECOND SIMULATIONS

(Each Estimation Method and Each Data-Generation Decision Rule)

	K-L Divergence Percentage for Indicated Estimation Method (%)						
Data Generation Decision Rule	**HB Add-itive**	**HB Conj-unctive**	**HB Disj-unctive**	**Comp-MP (*q* = *4*)**	**DOC-MP (*S* = 4)**	**LAD-DOC (∞,∞)**	**LAD-DOC (2, 4)**
Additive (*q* = 4)	6.9	24.2	22.1	20.0	**39.9**	29.7	25.4
Conjunctive	8.1	**40.5**	16.1	14.2	**42.5**	**39.9**	**39.9**
Disjunctive	6.5	14.9	**37.4**	21.8	29.0	29.5	26.7
DOC	8.3	28.9	21.8	21.5	**45.3**	**46.1**	**48.3**

Discussion of the Second Set of Simulations

DOC-based estimation methods tend to predict best for data generated with DOC rules. Because DOC rules nest conjunctive and disjunctive rules, DOC-based estimation also predicts well for data generated by conjunctive and disjunctive rules. The strong showing of DOCMP for

additive rules ($q = 4$) is a topic worth further exploration. One untested hypothesis is that the strong showing might be due to the fact that the synthetic data is based on 4 features with $S = 4$, in contrast to the empirical GPS data which have 12 features with $S = 4$.

We also did some preliminary exploration comparing CompMP ($q = \infty$) to CompMP ($q = 4$). As expected, the more-general unconstrained model, which nests some non-compensatory rules, tends to predict better than CompMP ($q = 4$) for the non-compensatory rules. The mixed model, CompMP ($q = \infty$) also does well when we simulate a DOC ($S = 3$) model which has complex conjunctions. As the generating model becomes more complex, the unconstrained models do well as paramorphic models. These results are tangential to our focus on DOC-based estimation, but worth further exploration by readers wishing to explore variations among the benchmarks.

In the empirical data most respondents were fit with DOC-based models that included only one conjunction. For example, for DOCMP, 7.1% of the "evaluate-all-profiles" respondents were fit with two conjunctions. This is moderately close to our synthetic-data condition of conjunctive respondents. (In contrast, synthetic respondents generated with the DOC rule were allowed up to three conjunctions.) In the conjunctive domain, the DOC-based models predict best. The conjunctive model predicts better in the conjunctive domain than in our data because no synthetic respondents have more than one conjunction. For this domain, CompMP ($q = \infty$) achieves a predictive hit rate of 84.4% and a K-L percentage of 31.1%, which are good, but less than DOCMP, LAD, or conjunctive estimation.

Finally, as in most synthetic-data experiments, it would be interesting to explore whether the results vary based on the various parameters used to generate synthetic respondents. Such explorations are beyond the scope of this Web Appendix.

THEME 8. RESULTS FOR ALTERNATIVE FORMATS

Some of the models which performed poorly on the primary format are not included in Appendices 8, 9, and 10. Although we have not yet run these benchmark models due to computational constraints, we expect no surprises from these models relative to those that have already been run. Data are available should the reader wish to investigate these benchmark models further. In all cases at least one DOC-based model is best or not significantly different than best on both metrics. The additive machine-learning model is significantly worse on K-L percentages, but, on one format, matches the DOC-based models on hit rate. These results are consistent with the basic qualitative directions discussed in the text. The sample sizes are evaluate-all-profiles (93), consider-only (135), reject-only (94), no-browsing (123), and text-only-evaluate-all-profiles (135).

TABLE W4
EMPIRICAL COMPARISON OF ESTIMATION METHODS
NO BROWSING FORMAT
(Representative German Sample, Task Format in Theme 12)

Estimation method	Overall hit rate (%)†	K-L divergence percentage (%)
Hierarchical Bayes Benchmarks		
Disjunctive	54.5	17.8
Subset Conjunctive	73.0	25.7
Additive	77.3	17.6
Machine-Learning Benchmarks		
Additive	78.8	26.1
DOC-Based Estimation Methods		
DOCMP	**81.5***	**34.1***
LAD-DOC	**80.7***	**32.6***

† Number of profiles predicted correctly, divided by 32. * Best or not significantly different than best at the 0.05 level.

TABLE W5
EMPIRICAL COMPARISON OF ESTIMATION METHODS
CONSIDER-ONLY FORMAT
(Representative German Sample, Task Format in Theme 12)

Estimation method	Overall hit rate (%)†	K-L divergence percentage (%)
Hierarchical Bayes Benchmarks		
Disjunctive	50.5	9.3
Subset Conjunctive	78.4	15.5
Additive	87.1	5.7
Machine-Learning Benchmarks		
Additive	**88.6***	16.9
DOC-Based Estimation Methods		
DOCMP	**88.6***	**29.4***
LAD-DOC	**88.4***	**29.4***

† Number of profiles predicted correctly, divided by 32. * Best or not significantly different than best at the 0.05 level.

TABLE W6

EMPIRICAL COMPARISON OF ESTIMATION METHODS – REJECT-ONLY FORMAT

(Representative German Sample, Task Format in Theme 12)

Estimation method	Overall hit rate (%)†	K-L divergence percentage (%)
Hierarchical Bayes Benchmarks		
Disjunctive	74.1	20.5
Subset Conjunctive	78.4	27.9
Additive	76.8	14.6
Machine-Learning Benchmarks		
Additive	81.5	31.7
DOC-Based Estimation Methods		
DOCMP	**83.7***	**42.1***
LAD-DOC	81.9	**39.1***

† Number of profiles predicted correctly, divided by 32. * Best or not significantly different than best at the 0.05 level.

THEME 9. RESULTS FOR THE US SAMPLE

We present here the results for the evaluate-all-profiles format. Limited testing on the other formats (HB benchmarks only) are consistent with the results for the German sample and with the US results for the evaluate-all-profiles format.

The basic results from the US sample evaluate-all-profiles format are consistent with those from the German sample. The biggest difference is that the US sample is based on a smaller sample size (38 respondents) and, hence, it is more difficult to establish statistical significance. DOC-based methods are significantly different than the additive machine-learning benchmark on the K-L percentage, but the additive machine-learning benchmark is not significantly different than the DOC-based methods on hit rates. The comparisons to the HB benchmarks remain consistent.

TABLE W7
EMPIRICAL COMPARISON OF ESTIMATION METHODS US EVALUATE-ALL-PROFILES FORMAT
(Representative German Sample, Task Format in Theme 12)

Estimation method	Overall hit rate (%)†	K-L divergence percentage (%)
Hierarchical Bayes Benchmarks		
Disjunctive	61.2	20.6
Subset Conjunctive	72.7	26.6
Additive	78.9	19.4
Machine-Learning Benchmarks		
Additive	**82.7***	30.0
DOC-Based Estimation Methods		
DOCMP	**82.3***	**36.5***
LAD-DOC	**82.7***	**36.0***

† Number of profiles predicted correctly, divided by 32. * Best or not significantly different than best at the 0.05 level.

THEME 10: RESULTS FOR TEXT-ONLY FORMAT

TABLE W8
EMPIRICAL COMPARISON OF ESTIMATION METHODS
TEXT-ONLY FORMAT
(Representative German Sample, Task Format in Theme 12)

Estimation method	Overall hit rate (%)†	K-L divergence percentage (%)
Hierarchical Bayes Benchmarks		
Disjunctive	43.7	11.2
Subset Conjunctive	72.4	21.3
Additive	78.4	13.9
Machine-Learning Benchmarks		
Additive	**81.4***	26.9
DOC-Based Estimation Methods		
DOCMP	**81.5***	**30.5***
LAD-DOC	80.7	**30.6***

† Number of profiles predicted correctly, divided by 32. * Best or not significantly different than best at the 0.05 level.

THEME 11: DECISION TREES AND CONTINUOUSLY-SPECIFIED MODELS

Decision Trees

Decision trees, as proposed by Currim, Meyer and Le (1988) for modeling consumer choice, are compatible with DOC rules for classification data (consider vs. not consider). In the growth phase, decision trees select the aspect that best splits profiles into considered vs. not considered. Subsequent splits are conditioned on prior splits. For example, we might split first on "B&W" vs. "color," then split "B&W" based on screen size and split "color" based on resolution. With enough levels, decision trees fit estimation data perfectly (similar to Result 4 in Theme 2), hence researchers either prune the tree with a defined criterion (usually a minimum threshold on increased fit) or grow the tree subject to a stopping criterion on the tree's growth (e.g., Breiman, et. al. 1984).

Each node in a decision tree is a conjunction, hence the set of all "positive" nodes is a DOC rule. However, because the logical structure is limited to a tree-structure, a decision tree often takes more than *S* levels to represent a DOC(*S*) model. For example, suppose we generate errorless data with the DOC(2) rule: $(a \wedge b) \vee (c \wedge d)$. To represent these data, a decision tree would require up to 4 levels and produce either $(a \wedge b) \vee (a \wedge \neg b \wedge c \wedge d) \vee (\neg a \wedge c \wedge d)$ or equivalent reflections. Depending on the incidence of profiles, the decision tree might also produce $(c \wedge d) \vee (c \wedge \neg d \wedge a \wedge b) \vee (\neg c \wedge a \wedge b)$, which is also logically equivalent to $(a \wedge b) \vee (c \wedge d)$. Other logically equivalent patterns are also feasible. This DOC(3) rule is logically equivalent to $(a \wedge b) \vee (c \wedge d)$, but more complex in both the number of patterns and pattern lengths. To impose cognitive simplicity we would have to address these representation and equivalence issues.

As a test, we applied the Currim, Meyer and Le (1988) decision tree to the data in Table

2. We achieved a relative hit rate of 38.5% and a K-L divergence of 28.4%, both excellent, but not as good as those obtained with DOCMP and LAD-DOC estimation. LAD-DOC ($p = 0.002$) and DOCMP ($p = 0.01$) are significantly better on relative hit rate. LAD-DOC ($p = 0.002$) is significantly better and DOCMP is better ($p = 0.06$) on information percentage. While many unresolved theoretical and practice issues remain in order to best incorporate cognitive simplicity and market commonalities into decision trees, we have no reason to doubt that once these issues are resolved, decision trees can be developed to estimate cognitively-simple DOC rules.

Continuously-Specified Models

Conjunctions are analogous to interactions in a multilinear model; DOC decision rules are analogous to a limited set of interactions (Bordley and Kirkwood 2004; Mela and Lehmann 1995). Thus, in principle, we might use continuous estimation to identify DOC decision rules. For example, Mela and Lehmann (1995) use finite-mixture methods to estimate interactions in a two-feature model. In addition, continuous models can be extended to estimate "weight" parameters for the interactions and thresholds on continuous features.

We do not wish to minimize either the practical or theoretical challenges of scaling continuous models from a few features to many features. For example, without enforcing cognitive simplicity there are over 130,000 interactions to be estimated for our GPS application. Cognitive simplicity constrains the number of parameters and, potentially, improves predictive ability, but would still require over 30,000 interactions to be estimated. Nonetheless, with sufficient creativity and experimentation researchers might extend either finite-mixture, Bayesian, simulated-maximum-likelihood, or kernel estimators to find feasible and practical methods to estimate continuously-specified DOC rules (Evgeniou, Boussios, and Zacharia 2005; Mela and Lehmann 1995; Rossi and Allenby 2003; Swait and Erdem 2007).

References for Theme 11

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

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

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.

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

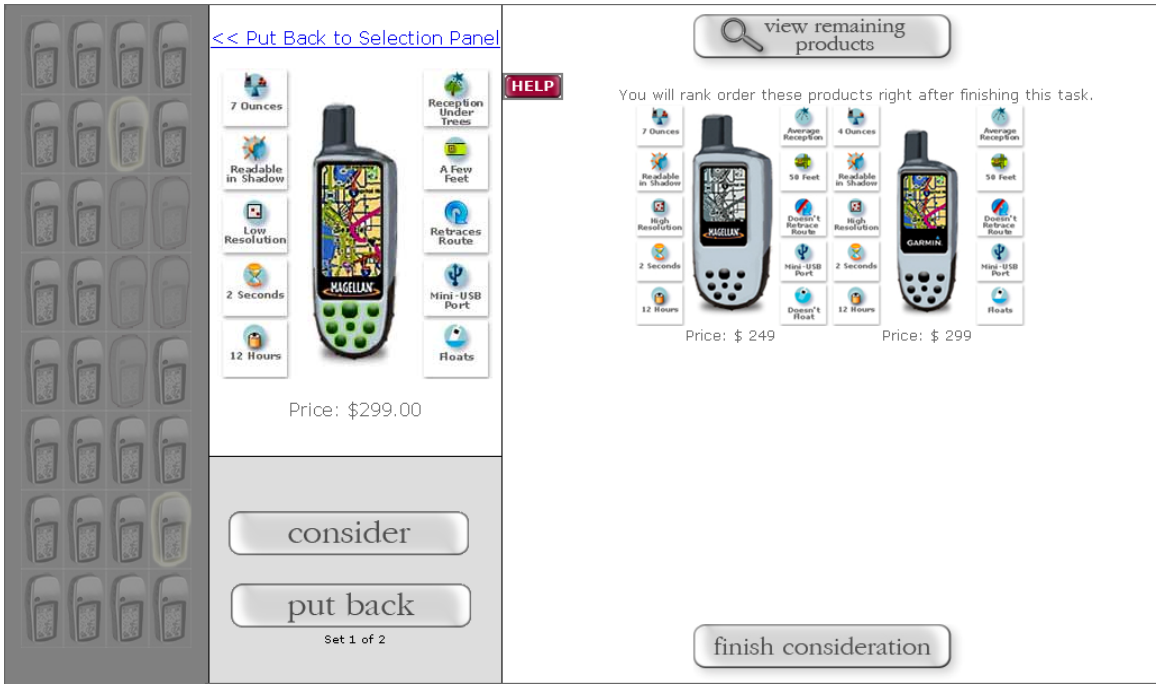
Mela, Carl F. and Donald R. Lehmann (1995), "Using Fuzzy Set Theoretic Techniques to Identify Preference Rules From Interactions in the Linear Model: An Empirical Study," *Fuzzy Sets and Systems*, 71, 165-181.

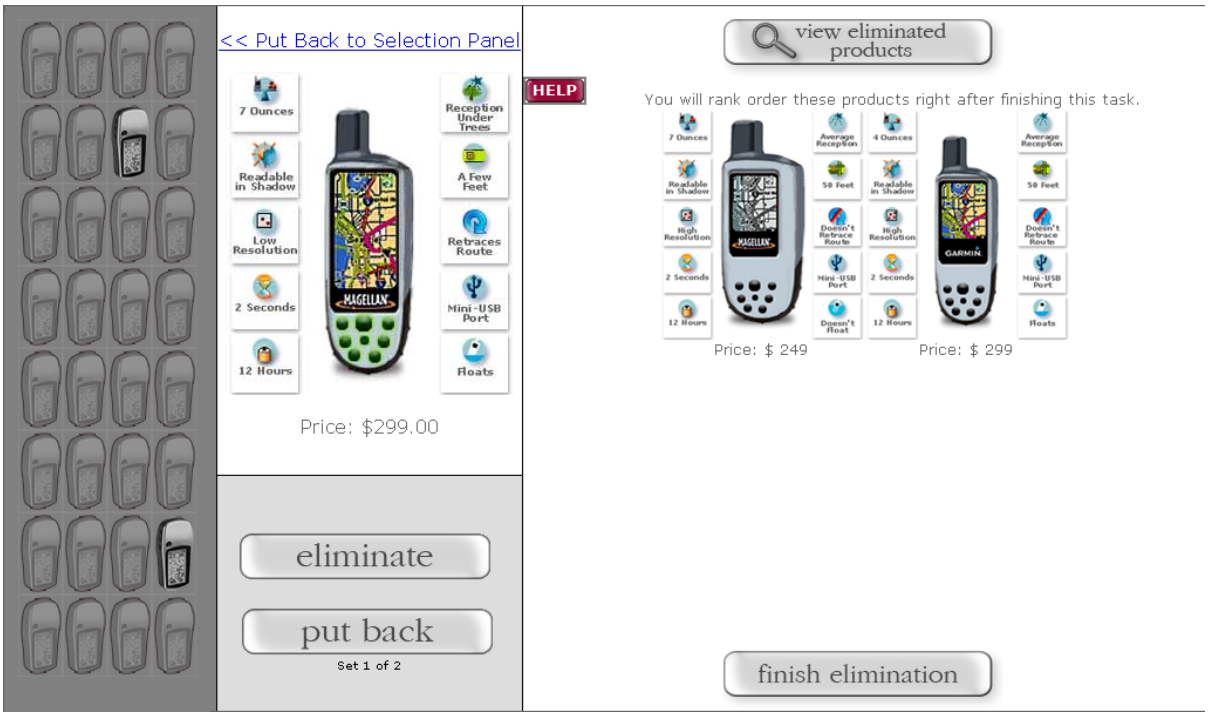
Rossi, Peter E., Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22(3), p. 304-328.

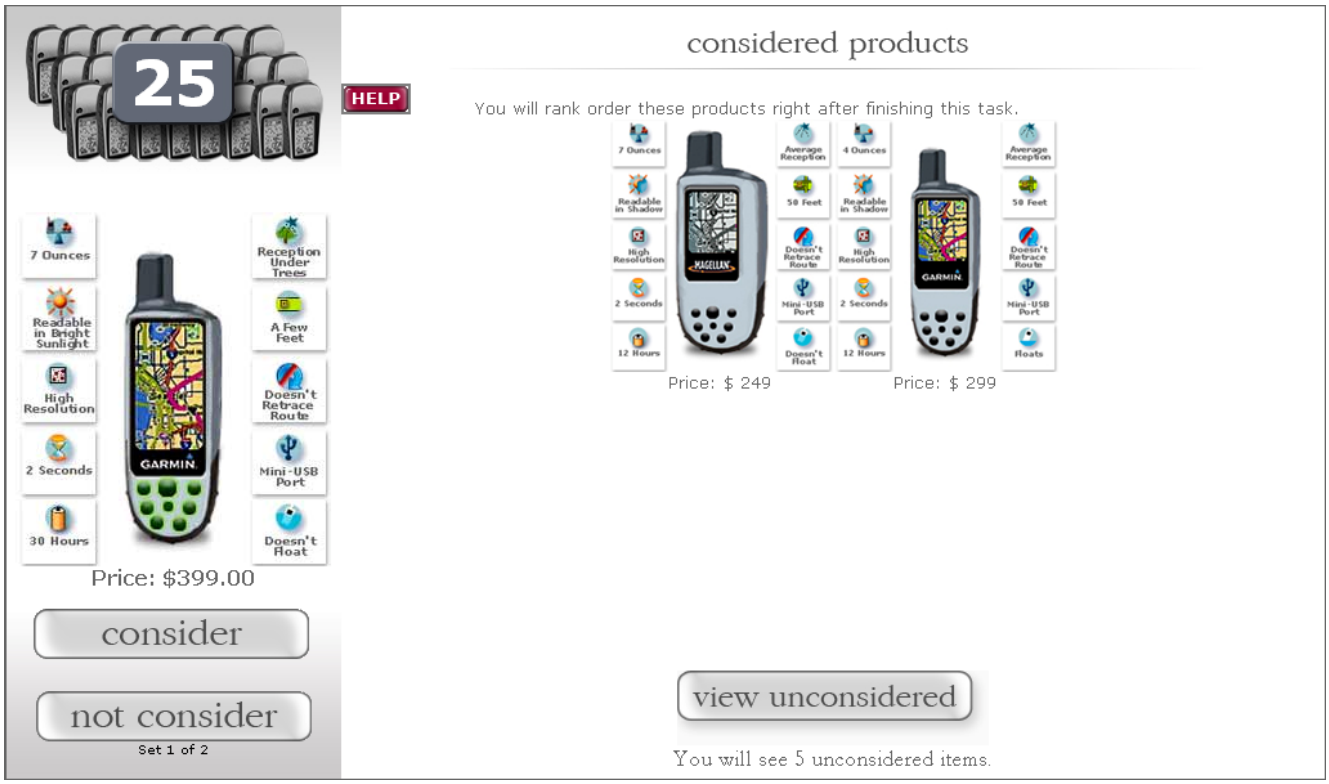
Swait, Joffre and Tülin Erdem (2007), "Brand Effects on Choice and Choice Set Formation Under Uncertainty," *Marketing Science* 26, 5, (September-October), 679-697.

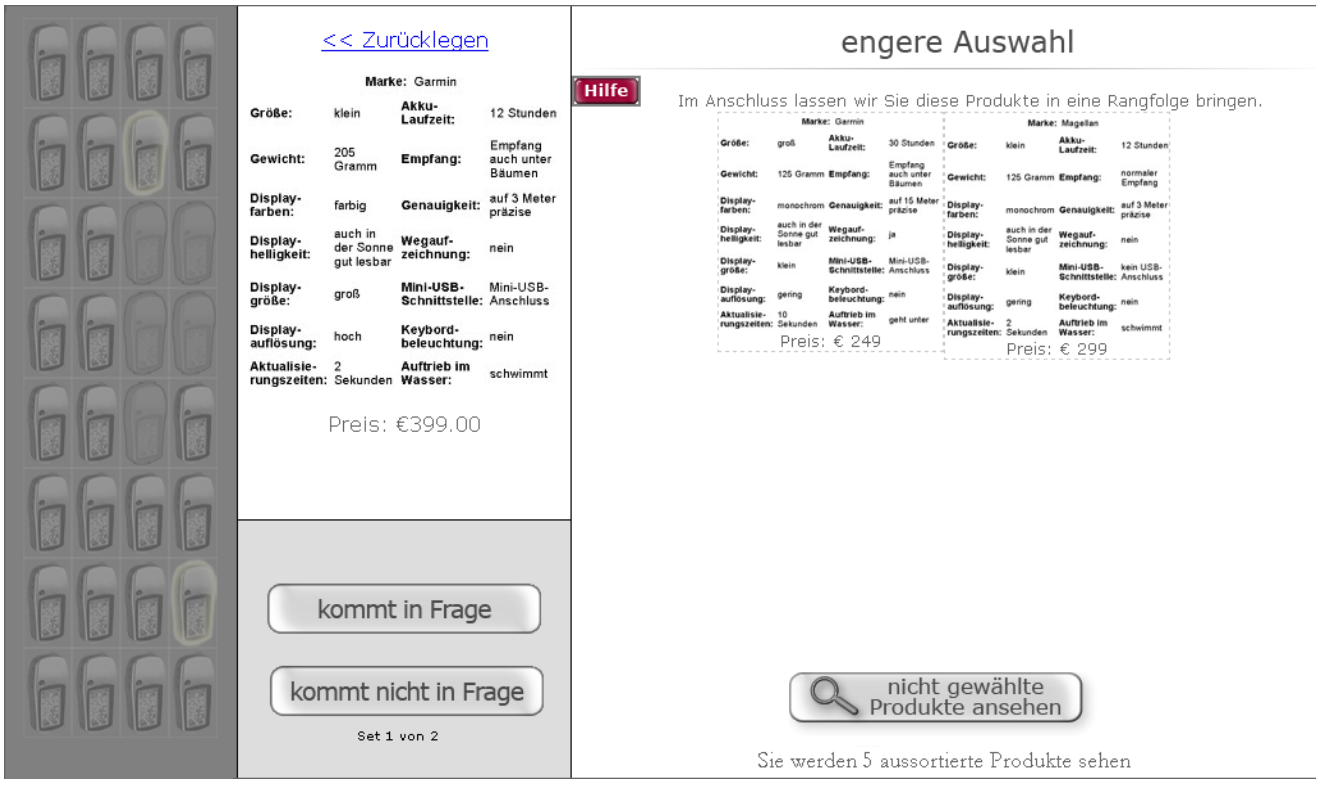
THEME 12: CONSIDER-ONLY, REJECT-ONLY, NO-BROWSING, TEXT-ONLY, EXAMPLE FEATURE-INTRODUCTION, AND INSTRUCTION SCREENSHOTS

Screenshots are shown in English, except for the text-only format. German versions, and other screenshots from the surveys, are available from the authors.






25
HELP
considered products
You will rank order these products right after finishing this task.
7 Ounces
Readable in Bright Sunlight
High Resolution
2 Seconds
30 Hours
Reception Under Trees
A Few Feet
Doesn't Retrace Route
Mini-USB Port
Doesn't Float
GARMIN
Price: $399.00
consider
not consider
Set 1 of 2
Average Reception
50 Feet
Readable in Shadow
Doesn't Retrace Route
Mini-USB Port
Floats
4 Ounces
12 Hours
MAGELLAN
Price: $ 249
Price: $ 299
view unconsidered
You will see 5 unconsidered items.


<< Zurücklegen
Marke: Garmin
Größe: klein
Akku-Laufzeit: 12 Stunden
Gewicht: 205 Gramm
Empfang: Empfang auch unter Bäumen
Display-farben: farbig
Genauigkeit: auf 3 Meter präzise
Display-helligkeit: auch in der Sonne gut lesbar
Wegauf-zeichnung: nein
Display-größe: groß
Mini-USB-Schnittstelle: Mini-USB-Anschluss
Display-auflösung: hoch
Keybord-beleuchtung: nein
Aktualisie-rungszeiten: 2 Sekunden
Auftrieb im Wasser: schwimmt
Preis: €399.00
kommt in Frage
kommt nicht in Frage
Set 1 von 2
engere Auswahl
Hilfe
Im Anschluss lassen wir Sie diese Produkte in eine Rangfolge bringen.
Marke: Garmin
Größe: groß
Akku-Laufzeit: 30 Stunden
Gewicht: 125 Gramm
Empfang: Empfang auch unter Bäumen
Display-farben: monochrom
Genauigkeit: auf 15 Meter präzise
Display-helligkeit: auch in der Sonne gut lesbar
Wegauf-zeichnung: ja
Display-größe: klein
Mini-USB-Schnittstelle: Mini-USB-Anschluss
Display-auflösung: gering
Keybord-beleuchtung: nein
Aktualisie-rungszeiten: 10 Sekunden
Auftrieb im Wasser: geht unter
Preis: € 249
Marke: Magellan
Größe: klein
Akku-Laufzeit: 12 Stunden
Gewicht: 125 Gramm
Empfang: normaler Empfang
Display-farben: monochrom
Genauigkeit: auf 3 Meter präzise
Display-helligkeit: auch in der Sonne gut lesbar
Wegauf-zeichnung: nein
Display-größe: klein
Mini-USB-Schnittstelle: kein USB-Anschluss
Display-auflösung: gering
Keybord-beleuchtung: nein
Aktualisie-rungszeiten: 2 Sekunden
Auftrieb im Wasser: schwimmt
Preis: € 299
nicht gewählte Produkte ansehen
Sie werden 5 aussortierte Produkte sehen

Display Colors

One GPS unit displays graphics in 256 colors, the other offers a monochrome display. Colors might be useful for reading maps . They help to distinguish objects on the screen.

Display Brightness

The standard brightness level allows you to read out information from the screen under many circumstances; however, it might be difficult to read in very bright sunlight.

The extra bright display allows you reading under any circumstances. Special transflective Thin-Film-Transmitters enable reading even under bright sunlight.

next

Brand

Two major brands exist in the market for handheld GPS-units. One is called *Garmin*, the other one is *Magellan*. Both companies are very experienced in developing and manufacturing GPS devices.

Price

GPS units vary in price between $249 and $399. Prices depend on features, materials, and level of technology. If you win the lottery, you will receive a GPS unit that closely matches your preferences (based on your answers). In addition, you will receive the difference to $500 in cash.

next

We will show you 32 GPS devices (in a random order).

- We will first ask you which device you would consider purchasing.
  - For example, you might include devices you would evaluate further
  - Or, you might include devices that you would purchase if your most-preferred device were unavailable.
- We then show these devices on a new screen and ask you to choose the device you most prefer, which device you would next prefer, etc.
- After a few fun questions to give you a break, we then show you a second set of 32 GPS devices.

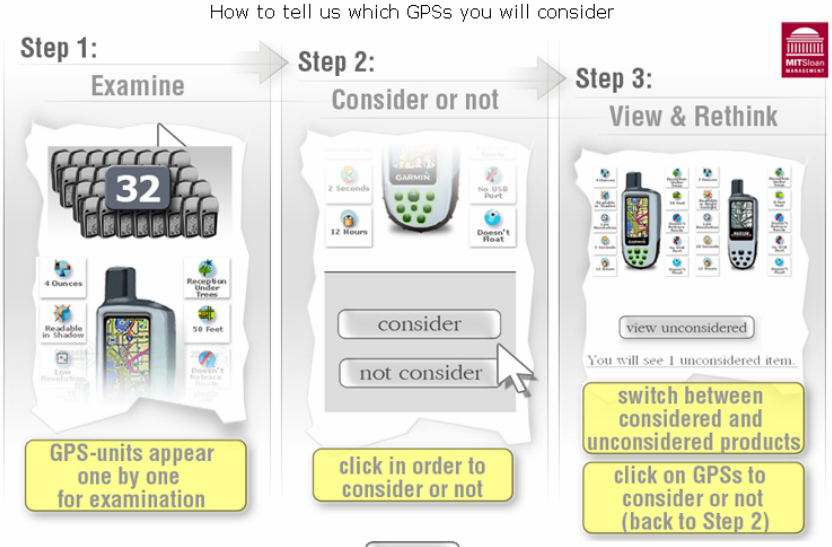
next

The next page explains how you will tell us which device you would consider.

Remember, if you win a GPS, these tasks will determine which GPS (plus cash) you receive!

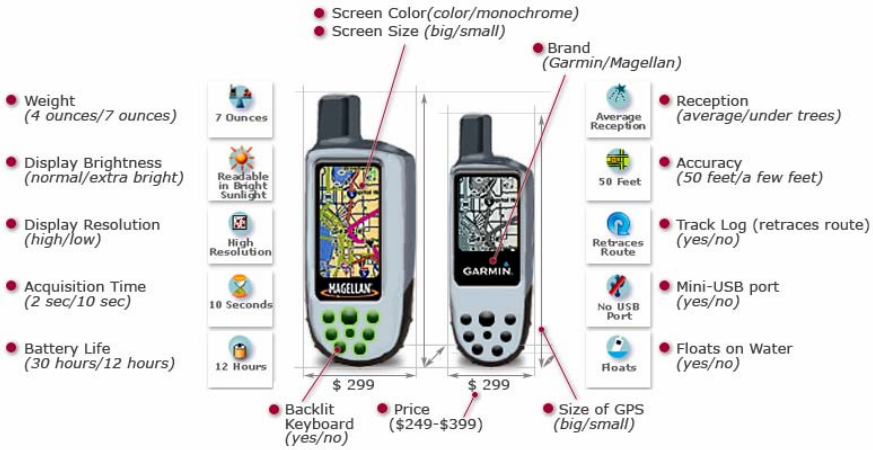
next

How to tell us which GPSs you will consider



next

Reminder: We provide you with 16 GPS features.



next