ERJEN VAN NIEROP, BART BRONNENBERG, RICHARD PAAP, MICHEL WEDEL, and PHILIP HANS FRANSES*

The authors propose a new model to capture unobserved consideration from discrete choice data. This approach allows for unobserved dependence in consideration among brands, easily copes with many brands, and accommodates different effects of the marketing mix on consideration and choice as well as unobserved consumer heterogeneity in both processes. An important goal of this study is to establish the validity of the existing practice to infer consideration sets from observed choices in panel data. The authors show with experimental data that underlying consideration sets can be reliably retrieved from choice data alone and that consideration is positively affected by display and shelf space. Next, the model is applied to Information Resources Inc. panel data. The findings suggest that promotion effects are larger when they are included in the consideration stage of the two-stage model than in a single-stage model. The authors also find that consideration covaries across brands and that this covariation is mainly driven by unobserved consumer heterogeneity. Finally, the authors show the implications of the model for promotion planning relative to a more standard model of choice.

Keywords: consideration set, brand choice, probit models, marketing-mix allocation

Retrieving Unobserved Consideration Sets from Household Panel Data

The theory of consideration sets, developed in the 1970s from the work of Bettman (1979), Howard and Sheth (1969), and Newell and Simon (1972), has been important ever since in marketing science (for overviews, see Malhotra, Peterson, and Kleiser 1999; Manrai and Andrews 1998; Roberts and Lattin 1997) and has had important implications for marketing practice. Its basic postulate is that consumers follow a two-stage decision process of brand choice. In the first stage, they narrow down the global set of alternatives to a smaller set, the consideration set, from which a choice is made in the second stage. Researchers in marketing have provided ample empirical evidence corroborating this two-stage process of consumer choice (Lussier and Olhavsky 1979; Payne 1976; Wright and Barbour 1977), and from a theoretical perspective, allowing consumers to choose from a limited set of alternatives has much appeal.

Consideration sets vary across households (Alba and Chattopadhyay 1985; Belonax and Mittelstaedt 1978; Chiang, Chib, and Narasimhan 1999; Roberts and Lattin 1991) and are sensitive to marketing instruments, such as promotions (Siddarth, Bucklin, and Morisson 1995) and advertising (Mitra 1995). Ignoring consideration formation in models of choice has been shown to lead to the underestimatation of the impact of marketing control variables (Bronnenberg and Vanhonacker 1996; Chiang, Chib, and Narasimhan). With the rapid proliferation of the number of brands in the marketplace and the increase in cognitive demands placed on consumers choosing among them, understanding consideration set formation and how marketing affects it continues to increase in importance to marketing managers, for whom entering the consideration set has become a key strategic goal (Corstjens and Corstjens 1999).

*Erjen van Nierop is Assistant Professor of Marketing, Department of Marketing, Faculty of Economics and Business, University of Groningen (e-mail: j.e.m.van.nierop@rug.nl). Bart Bronnenberg is Professor of Marketing and CentER Fellow, Faculty of Economics and Business Administration, Department of Marketing, Tilburg University (e-mail: bart.bronnenberg@uvt.nl). Richard Paap is Professor of Econometrics and Marketing (e-mail: paap@few.eur.nl), and Philip Hans Franses is Professor of Applied Econometrics and Professor of Marketing Research (e-mail: franses@few.eur.nl), Econometric Institute, Faculty of Economics, Erasmus University Rotterdam. Michel Wedel is PepsiCo Professor of Consumer Research, Department of Marketing, Robert H. Smith School of Business, University of Maryland (e-mail: mwedel@rhsmith.umd.edu). Sunil Gupta served as associate editor for this article.

ISSN: 0022-2437 (print), 1547-7193 (electronic)

Most approaches used to model consideration sets are rooted in the random utility theory framework (e.g., Guadagni and Little 1983; McFadden 1973). Including the consideration stage into this framework is not trivial, because consideration sets are usually neither observed nor identifiable with certainty (Ben-Akiva and Boccara 1995). There are two streams of consideration set research that have evolved somewhat independently. One stream of research directly assesses consideration set membership of individual brands as stated by consumers through survey instruments and models the marginal distribution of consideration for each brand (e.g., Roberts and Lattin 1991). This approach works well even for large numbers of brands, but it is usually based on an assumption of independence of consideration across brands (Ben-Akiva and Boccara 1995), which has remained untested in empirical research. We call this approach the "stated consideration set approach."

The second stream of research identifies the distribution of consideration sets indirectly from choice data (e.g., Chiang, Chib, and Narasimhan 1999; Manski 1977; Mehta, Rajiv, and Srinivasan 2003). To account for the unobserved nature of consideration and to obtain marginal choice probabilities, this stream integrates over all possible consideration sets of which there are $2^J - 1$, where J is the number of choice options. This method accounts for unobserved dependencies across brands because the realization of the entire consideration set is modeled. This approach, which we call the "revealed consideration set approach," accounts for the dependence of consideration set membership across brands. However, the number of possible consideration sets is exponential in the number of brands contained in the global choice set (see Chiang, Chib, and Narasimhan 1999). With larger numbers of brands, the methods in question become computationally infeasible. These methods do not offer a natural way to study either marginal brand set membership probabilities or their responsiveness to marketing action. The approach that Gilbride and Allenby (2004) developed addresses some of these problems. Their approach models consumers screening on attribute-based rules in a multinomial probit (MNP) choice model, and it is calibrated on stated preference data.

Table 1 provides an overview of consideration set modeling approaches and shows how the approaches deal with explanatory variables, heterogeneity, unobserved correlation of considered alternatives, numbers of brands, and the type of data accommodated. The key takeaway from Table 1 is that the approach we propose in this study enables us to incorporate different variables into the consideration and choice stages and allows for heterogeneity in both stages, correlations of brands in the consideration stage, and a relatively large number of choice alternatives.

An important question that remains unanswered is whether the "consideration probabilities" from the models used in the revealed consideration set approach and

Table 1

CONSIDERATION SET MODELS IN THE LITERATURE

Paper	*Model Type*	*Possible Data*	*Correlation Across Brands*	*Can Variables Differ Across the Two Stages?*	*Heterogeneity*	*Number of Alternatives*
Manski (1977)	Random utility	N.A.	No	No implementation	Yes	N.A.
Lussier and Olshavsky (1979)	Experiment to deduct conjuctive, disjunctive, compensatory rules	CS + BC	N.A.	N.A.	Yes	3–12
Alba and Chattopadhyay (1985)	Experimental design	CS	N.A.	N.A.	N.A.	N.A.
Roberts and Lattin (1991)	Benefit versus costs, MNL, cross-sectional	BC + CS	No	CS: brand evaluations	No	3
Andrews and Srinivasan (1995)	MNL	BC	No	Yes	No	3 + 7
Ben-Akiva and Boccara (1995)	PCS	BC	No	Yes	No	3
Mitra (1995)	Regression of consideration measurements on experimental treatment groups	CS	N.A.	N.A.	No	12
Siddarth, Bucklin, and Morrison (1995)	Restricted MNL	BC	No	No	No	7 + 26
Bronnenberg and Vanhonacker (1996)	MNL	BC	No	Yes	In CS stage	12
Chiang, Chib, and Narasimhan (1999)	Probability for each possible CS	BC	Yes	No variables in CS stage	In both stages	4
Mehta, Rajiv, and Srinivasan (2003)	Structural model	BC	No	Yes	In CS stage	4
Gilbride and Allenby (2004)	MNP, conjunctive, disjunctive, compensatory	BC	No	No	In both stages	7
Zhang (2006)	MNL and probabilistic elimination by aspects	BC	No	Yes	In both stages	6
Our study (2008)	PCS + MNP, experimental and empirical data	CS + BC, or both	Yes	Yes	In both stages	8 and 10

Notes: MNL = multinomial logit model, PCS = probabilistic choice set model, CS = consideration set, and BC = brand choice. N.A. = not applicable.

estimated from choice data reflect consideration sets as stated by consumers and modeled in the stated consideration set approach. This issue bears directly on the validity of the interpretations of parameter estimates and the resultant recommendations for marketing practice. Indeed, Roberts and Lattin (1997, p. 407) conclude that researchers working without explicit measures of consideration "cannot address whether the consideration stage of their model corresponds to a cognitive stage of consideration or if it is just a statistical artifact of the data. . . . Even if what is inferred is consideration, it will be estimated with substantial error." Therefore, it may be surprising that no research to date has addressed the issue of convergent validity of stated versus revealed consideration sets. A possible reason for this gap in the literature is that the aforementioned streams of research have developed more or less independently and that, to demonstrate convergent validity, a joint modeling framework is needed that accommodates stated consideration data, revealed consideration data (choice), or both at the same time. None of the currently available approaches allow for this, and it is the key intended contribution of this article. We develop a model for consideration and brand choice that provides a unifying framework for analysis in the stated and revealed approaches to consideration set identification. It can accommodate either type of data or both simultaneously. We apply the approach to investigate the extent to which consideration probabilities derived from choice data are more than a "statistical artifact" and assess convergent validity of stated and revealed consideration sets.

In the model, we directly specify the joint distribution of the probabilities of brands' consideration set membership by modeling consideration set membership of brands as binary probits that can covary across brands, as in a multivariate probit model (Edwards and Allenby 2003). The model does not suffer from the curse of dimensionality and provides a tractable representation of consideration set formation, the complexity of which is only linear (or, depending on the exact specification, at most quadratical) in the global number of choice options. We develop our model primarily for the purpose of obtaining better substantive insights into consideration and choice processes and the effect of marketing variables on these processes. We do not primarily aim to improve predictive validity but rather to provide richer insight into how marketing-mix variables affect choice and demand (see also Andrews and Srinivasan 1995). Indeed, a core contribution of this study is that we validate the inference of consideration from choice data using actually measured consideration sets.

Next, we lay out the model and its Markov chain Monte Carlo (MCMC) estimation procedure. Then, we investigate whether it is possible to identify consideration sets from choice behavior, using data from an experimental study. We then apply the model to a scanner panel data set and show its implications for marketing decisions—namely, feature ad placement and price discounts. We conclude by discussing the limitations and prospects for further research.

THE MODEL

Preliminaries

In this section, we propose a model to describe the brand choice decision of household i ($i = 1, \ldots, I$) choosing brand j ($j = 1, \ldots, J$) at purchase occasion t ($t = 1, \ldots, T_i$). The brand choice of household i at time t is described by the random variable D_{it}, which can take values from 1 to J. The actual brand choice is given by d_{it}. Without loss of generality, we consider here the more complex situation in which only such choice data are available and the consideration sets themselves are unobserved. Households typically do not consider all brands in their choice decision but rather choose a brand from their consideration or choice set. This choice set may contain one, two, or even all brands that are available to the household. For each household, there are $Q = 2^J - 1$ potential nonempty consideration sets. We model the consideration set of household i at time t by the random variable C_{it}. Because we assume that households choose a brand from their unobserved consideration set, after observing the actual brand choice, the number of potential consideration sets for a household is 2^{J-1}. We denote the collection of potential consideration sets for household i at purchase occasion t as CS_{it}. We include a set of marketing control variables, such as price, feature advertising, and display, denoted as X_{ijt} in the consideration stage, and another, possibly overlapping set, denoted as W_{ijt}, in the brand choice stage.

Stage 1: Consideration Set

The consideration set of household i at time t, C_{it}, is represented by a J-dimensional vector with binary elements, C_{ijt}, which equal 1 if brand j occurs in the consideration set of household i at time t and 0 if otherwise. To describe whether a brand is in the consideration set of household i, we consider the following:

$$C^*_{ijt} = X'_{ijt}(\alpha + \alpha_i) + \varepsilon_{ijt},\ j = 1, \ldots, J, \tag{1}$$

where X_{ijt} is a k_X-dimensional vector containing brand- and purchase-related explanatory variables, including brand-specific intercepts; α describes the average effect and α_i describes the household-specific effect; and ε_{ijt} is an unknown error process, the distribution of which we describe subsequently.

We accommodate the notion that many choice data sets are observed, conditional on a choice being made. That is, the data do not contain records for when no choice is made. This implies that empty consideration sets do not occur. Ignoring this feature of the data leads to biased estimates, as we show with a synthetic data study in Web Appendix A (http://www.marketingpower.com/jmrfeb10). We formally accommodate this data-generating mechanism by having at least one brand (with the highest consideration intensity C^*_{ijt}) in the consideration set. This implies that brand j enters the consideration set if

$$C^*_{ijt} > 0 \text{ or } C^*_{ikt} < C^*_{ijt} < 0 \text{ for all } k \neq j. \tag{2}$$

These constraints reflect that for every purchase occasion, at least one of the brands is in the consideration set. The consideration intensity for this brand is either positive or higher than the intensity of all the other brands, as is the case in standard choice models as well.

The form of the consideration set probabilities depends on the distribution of the ε_{ijt}. We assume that $\varepsilon_{it} = (\varepsilon_{i1t}, \ldots, \varepsilon_{iJt})'$ is normally distributed—that is, $\varepsilon_{it} \sim N(0, \Sigma)$, where the off-diagonal elements in the covariance matrix Σ describe the dependencies among the probabilities that the brands are contained in the consideration set. In this formulation, multiplying all intensities C^*_{ijt} by a positive constant would result in the same consideration set. Therefore, for identification purposes, we set the diagonal elements of Σ all equal to 1.

We assume that the household-specific parameters are drawn from a population distribution—that is, $\alpha_i \sim N(0, \Sigma_\alpha)$, where Σ_α is a diagonal matrix. Note that this approach allows for a nondiagonal covariance structure. The unconditional covariance structure of C^*_{it} equals

$$X_{it}\Sigma_\alpha X'_{it} + \Sigma, \tag{3}$$

where $X_{it} = (X_{i1t}, \ldots, X_{iJt})'$ (for a similar approach and an application, see Allenby and Rossi 1999; Hausman and Wise 1978). Equation 3 shows that covariation in brand consideration at the market level can be decomposed into a component $X_{it}\Sigma_\alpha X'_{it} + \Sigma$ that is due to consumer heterogeneity and a component Σ that captures within-household covariation across brands.

Our modeling approach has the advantage that we model the probability that a brand j is included in the consideration set, which means that we only deal with J instead of $Q = 2^J - 1$ alternatives, as would be the case when probabilities are assigned to all potential consideration sets. Therefore, the number of parameters in this approach increases, at most, quadratical in J.

Stage 2: Brand Choice

Given the consideration sets of households, we describe their brand choice with an MNP model. We assume that household i perceives utility U_{ijt} from buying brand j at purchase occasion t; that is,

$$U_{ijt} = W'_{ijt}(\beta + \beta_i) + \eta_{ijt}, \quad j = 1, \ldots, J, \tag{4}$$

where W_{ijt} is a k_W-dimensional vector containing explanatory variables, including brand-specific intercepts; β describes the average effect and β_i describes the household-specific effect; and η_{ijt} is a random error term. The vector of error terms $\eta_{it} = (\eta_{i1t}, \ldots, \eta_{iJt})'$ is assumed to be normally distributed, $\eta_{it} \sim N(0, \Omega)$. We also assume that the household-specific parameters β_i are drawn from a population distribution—that is, $\beta_i \sim N(0, \Sigma_\beta)$, where Σ_β is a diagonal matrix. Household i purchases brand j at purchase occasion t if the perceived utility of buying brand j is the maximum over all perceived utilities for buying the other brands in the consideration set c_{it}—that is, if

$$U_{ijt} = \max(U_{ikt} \text{ for } k = 1, \ldots, J \mid c_{ikt} = 1). \tag{5}$$

Thus, the probability that household i chooses brand j at purchase occasion t given the consideration set c_{it} and given β_i is as follows:

$$\begin{aligned} \Pr[D_{it} = j \mid c_{it}, \beta_i; \beta, \Omega] &= \Pr[U_{ijt} > U_{ikt}\ \forall\ k \neq j \mid c_{ijt} = c_{ikt} = 1] \\ &= \Pr[\eta_{ikt} - \eta_{ijt} < W'_{ijt} - W'_{ikt}(\beta + \beta_i)\ \forall\ k \neq j \mid c_{ijt} = c_{ikt} = 1]. \end{aligned} \tag{6}$$

Not all elements of the covariance matrix Ω are identified (see Bunch 1991). In addition, the off-diagonal elements are often empirically not identified (Keane 1992), so we opt for a diagonal covariance matrix in which we restrict one of the diagonal elements of Ω to be 1 for identification. The diagonal structure for Ω implies a nondiagonal covariance matrix through the specification of the unobserved household heterogeneity (see Allenby and Rossi 1999; Hausman and Wise 1978):

$$W_{it}\Sigma_\beta W'_{it} + \Omega, \tag{7}$$

where $W_{it} = (W_{i1t}, \ldots, W_{iJt})'$. It is possible to extend this model further through a nondiagonal Ω, but doing so may empirically lead to difficulties in identifying all parameters. For the same reason, it is empirically difficult to model correlations between the two stages, though this would be a potentially useful extension.

PARAMETER ESTIMATION

We consider the case of revealed consideration data, in which only choices of households have been observed. To estimate the model parameters, we consider the likelihood a function of the brand choices of the households $D = \{\{d_{it}\}_{t=1}^{T_i}\}_{i=1}^{I}$; that is,

$$\begin{aligned} &L(D \mid \theta) \\ &= \prod_{i=1}^{I} \int\!\!\int_{\beta_i \alpha_i} \prod_{t=1}^{T_i} \left(\sum_{\forall c_{it} \in CS_{it}} \Pr[C_{it} = c_{it} \mid \alpha_i; \alpha, \Sigma] \Pr[D_{it} = d_{it} \mid c_{it}, \beta_i; \beta, \Omega] \right) \\ &\quad \times \phi(\alpha_i; 0, \Sigma_\alpha)\phi(\beta_i; 0, \Sigma_\beta)\, d\alpha_i d\beta_i, \end{aligned} \tag{8}$$

where $\theta = (\alpha, \Sigma_\alpha, \beta, \Sigma_\beta, \Sigma, \Omega)$ and CS_{it} is the set of potential consideration sets for household i at purchase occasion t. Because we do not observe the consideration set, we sum over all potential consideration sets for household i at time t.

If we apply our model to stated consideration data, the situation simplifies, and next to the choice indicators d_{it}, we also observe the choice set membership indicators, c_{it}. The expression for the likelihood is similar to Equation 8, but the summation across all possible consideration sets vanishes, and the approach reduces to the separate estimation of the consideration and choice components. Because that situation is more straightforward, in the more detailed description of the estimation methodology, we focus on the more complicated case of inferring the joint process of choice and consideration from choice data alone.

We estimate the model with MCMC methods. In Web Appendix B (http://www.marketingpower.com/jmrfeb10), we provide the details of the full conditional posterior distributions and sampling algorithms for the model parameters. For the estimation of the parameters of each model considered in this article, we generate 20,000 iterations of the Gibbs sampler for burn-in and 20,000 iterations for analysis, in which we retain every 20th draw. The (unreported) iteration plots are inspected to determine whether the sampler converges to stationary draws from the posterior distributions of the model parameters. In Web Appendix A (http://www.marketingpower.com/jmrfeb10), we demonstrate that synthetic data analyses show that the parameters are recovered well and that the chains are stationary well before the end of the burn-in. We report the posterior means and standard deviations of the parameters in the subsequent empirical analyses. We compare several models by computing their log-predictive likelihoods (LPLs) (e.g., Bjørnstad 1990; Geweke 2005) and hit rates. We compute these out of sample—that is, for every respondent, we leave out one purchase occasion from the estimation sample. We denote the LPL of our model as $LPL_{proposed}$ and that of the model we compare it with as $LPL_{alternative}$.

IDENTIFYING CONSIDERATION SETS FROM CHOICE DATA

Data from an Online Experiment

We apply our model to a data set consisting of stated choice and consideration protocol data collected in an online experiment. We use this experiment to investigate the validity of stated consideration sets and the sets identified from choice data only. In the online shopping experiment, respondents chose among eight brands of laundry detergent over ten choice occasions. Respondents interfaced with a digital image of a supermarket shelf that contained the universal set of choice options. The choice environment was constant across respondents but varied across choice occasions. We manipulated promotion, brand position on the shelf, and shelf facings.

Figure 1 shows a screenshot from the sixth choice occasion.We simulated a promotion environment by putting "end-of-aisle" displays into the simulation. These were created by showing the brand on promotion before showing the entire shelf. Respondents had the option to choose the promoted brand (and entirely bypass the shelf) or skip the end-of-aisle promotion and visit the regular shelf.

We measured (revealed) choice, information acquisition, and stated consideration set membership. We measured the latter through two questions using 100-point sliders: (1) "Did you consider brand j seriously?" and (2) "Is brand j acceptable to you?" We adopted this operationalization of consideration from Lehmann and Pan (1994 and Nedungadi (1990).

We administered the experiment to graduate students at a U.S. university. Participants received a diskette with the experiment on it and were reminded once a week by e-mail to make a choice. Diskettes were collected after ten weeks. In total, 55 respondents submitted data, yielding 528 observations. We used 48 respondents with 432 purchases for estimation. Table 2 shows the description of the data set.

We computed the stated levels of consideration in Table 2 as the average of the two questions (divided by 100) averaged across purchase occasions and individuals. We construct discrete consideration set memberships by dichotomizing the average of the two questions (divided by 100) around .5 for each choice occasion and each individual. Although we could have chosen other cutoffs, the scale midpoint is the natural choice. The variable shelf space represents the surface of the facings of the six brands. Display frequency is the fraction of purchase occasions that the brand was positioned at the end of the aisle.

Table 2 shows that there is considerable variation in choice shares and consideration across brands. A notable aspect from Table 2 is that the ratio between choice share

Figure 1
SCREENSHOT FROM THE SIXTH CHOICE OCCASION

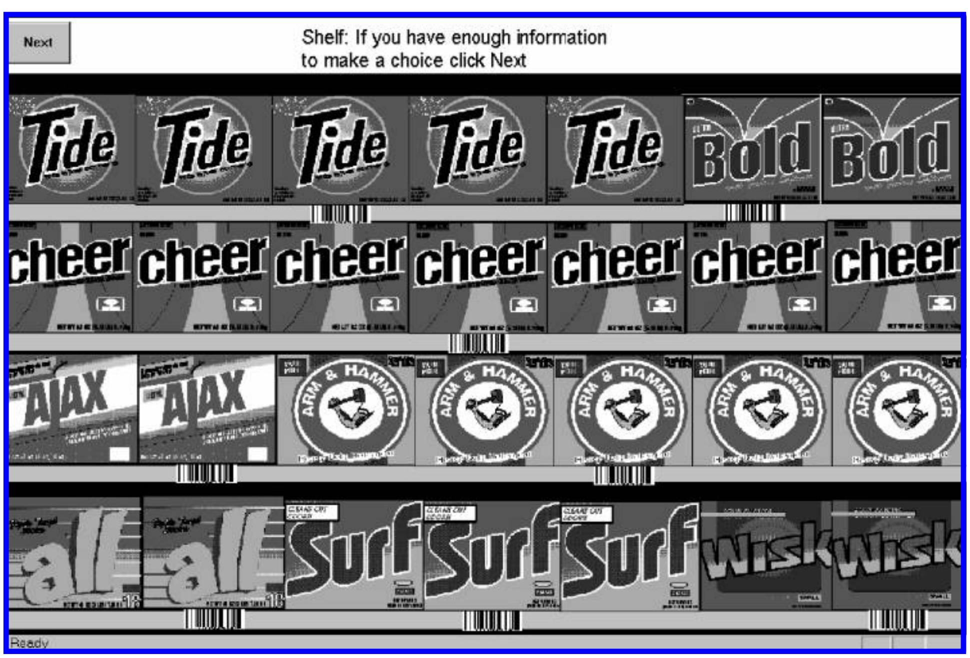

Table 2
DESCRIPTIVE STATISTICS FOR THE EXPERIMENTAL DATA SET (432 OBSERVATIONS)

Brand	*Share* %	*Consideration* %[a]	*Display Frequency* %	*Average Shelf*
Ajax	3	11	0	.34
All	10	28	10	.33
Arm & Hammer	11	21	10	.40
Bold	5	23	10	.38
Cheer	26	58	20	.84
Surf	5	17	0	.40
Tide	40	66	10	.71
Wisk	2	20	0	.36

[a]Consideration share is computed as the mean of the two consideration questions (divided by 100) averaged across purchase occasions and individuals.

and consideration is different across brands (for a similar observation, see Siddarth, Bucklin, and Morisson 1995).

Operationalizations

We assume that consideration is driven by in-store merchandizing activity, specifically display and shelf space, which aims to make a brand more salient at the point of purchase. These variables do not directly drive utility, whereas prices, for example, do. This distinction is not without precedent. It implies that the consumer chooses among the products that are most known (or salient) to him or her and to which he or she has access (Dickson and Sawyer 1990; Mitra and Lynch 1995; Nedungadi 1990; Zhang 2006). In both stages, we allow for consumer-specific brand intercepts that serve to capture the effects of factors that do not depend on the marketing or choice environment, including consumer preferences.

Estimation Results from the Online Experiment

To investigate the extent to which it is possible to infer consideration sets from choice data, we estimate our model on the choice data only and compare the results with a model estimated on choice and consideration data jointly. The estimation results for the consideration set models appear in the top part of Table 3. On the left, the table presents the model estimates of the consideration process estimated from the choice data alone. On the right, the table lists the model estimates based on the stated consideration data. The comparison of the two columns shows that the parameter estimates have comparable values: The correlation between the parameter estimates of the two models is r = .96.[1] However, because the parameters may not be strictly comparable across different models because of the identification constraint, we also compute the posterior distributions of the shelf and display elasticities and compare them across the two models. The results also appear in Table 3. Across the brands, the display and shelf elasticities are slightly higher for our full model, as could be expected from the parameters in the table, but for all pairs of parameters, the corresponding 95% highest posterior density (HPD) intervals overlap, showing high correspondence in the consideration elasticities derived from choice and from stated consideration data. It is also noteworthy that in our model, we find significant shelf space effects, whereas the MNP in the bottom-right part of the table shows an insignificant shelf space parameter.

Model Comparisons

To investigate our choice of the sets of covariates included in the consideration and choice components of the model, we also estimated the full model with all covariates included in both stages. This model leads to worse out-of-sample predictions. The LPLs are as follows: $LPL_{proposed} = -14.1$, and $LPL_{alternative} = -15.6$, suggesting that this specification leads to overfitting.

The intercepts in the choice model estimates in Table 3 are insignificant. This would imply that most of the "action" in our model is in the consideration formation stage (see Hauser and Wernerfelt 1989, 1990). However, because of our formulation of heterogeneity, every respondent is endowed with his or her own preference parameters. On average, respondents differ from 8 other respondents in their brand choice intercepts, and some from as many as 30 others. In addition, of the 48 households, 21 have at least one brand intercept that differs significantly from 0. To investigate this further, we estimated a model with only homogeneous intercepts in the choice part. This model performs worse than our model in terms of out-of-sample predictions: $LPL_{proposed} = -14.1$, and $LPL_{alternative} = -15.3$, indicating that though consideration may be the most important component of our model, conditional on consideration, the choice component still matters.

To empirically verify that the off-diagonals in the error covariance matrix of the consideration component of the full model are indeed indistinguishable from 0, we compute the LPL for a model with a diagonal correlation matrix and for a model with a full correlation matrix. The out-of-sample LPL of the model with diagonal correlation matrix is −14.1, whereas for the model with a full correlation matrix, the LPL is −14.8. These numbers show that the model with the identity correlation matrix is weakly favored. This conclusion is derived from the stated consideration set data and the consideration sets derived from the choice data.

Finally, we compare the LPLs of the models that use consideration set data with that of the model that estimates consideration sets from the choice data. As we expected, the model with observed consideration sets produces better fit in terms of out-of-sample LPLs: $LPL_{proposed} = -14.1$, and $LPL_{alternative} = -13.5$. When examining the consideration set LPLs in Table 4, we observe that for most brands, the combined model is not far behind.

Comparison of Estimated Consideration Sets

We now compare the consideration sets themselves. Using the full model, we can infer the consideration sets

[1]If we omit the unobserved heterogeneity in the consideration stage, we find a lower correlation: r = .75.

Table 3

POSTERIOR MEANS AND STANDARD DEVIATIONS FOR PARAMETERS OF THE FULL MODEL (LEFT) AND SEPARATE CONSIDERATION AND CHOICE MODELS (RIGHT)

	Full Model: Consideration Part[a]				*Consideration-Only Model*[b]			
	Estimates		*Elasticities*[c]		*Estimates*		*Elasticities*	
	M	*SD*	*Display*	*Shelf*	*M*	*SD*	*Display*	*Shelf*
Stage 1: Consideration Set								
α_{Ajax}	**−1.884**	.491	.535	.149	**−2.174**	.288	.546	.180
α_{All}	**−1.583**	.246	.643	.171	**−1.429**	.201	.509	.107
$\alpha_{Arm\&Hammer}$	**−1.285**	.258	.495	.156	**−1.726**	.227	.570	.148
α_{Bold}	**−2.636**	.362	.727	.217	**−2.168**	.306	.547	.170
α_{Cheer}	**−1.258**	.421	.370	.156	−.681	.371	.353	.178
α_{Surf}	−.992	.664	.509	.223	**−1.553**	.274	.508	.171
α_{Tide}	.401	.607	.301	.165	−.244	.331	.256	.124
α_{Wisk}	**−2.131**	.324	.645	.210	**−2.122**	.306	.561	.186
Display	**4.143**	1.402			**1.751**	.181		
Shelf	**1.108**	.456			.584	.412		

	Full Model: Choice Part[a]		*Regular MNP: Choice Model*[d]	
	M	*SD*	*M*	*SD*
Stage 2: Brand Choice				
β_{Ajax}	−1.072	.863	-.031	.314
β_{All}	.765	.435	.157	.383
$\beta_{Arm\&Hammer}$	−.012	.401	-.458	.521
β_{Bold}	−.062	.336	-.422	.328
β_{Cheer}	.674	.429	.194	.419
β_{Surf}	−1.202	1.266	**−1.413**	.641
β_{Tide}	−.189	.426	**1.210**	.429
β_{Wisk}	.000[e]	—	.000[e]	—
Display			**2.621**	.292
Shelf			.818	.521

[a]Our full model estimated on observed choice dummies, ignoring consideration set knowledge. The top part shows the probabilistic choice set part, and the bottom part shows the associated MNP part.

[b]Estimated on observed consideration set dummies.

[c]Elasticities of consideration share with respect to display and shelf changes. For each brand separately, we vary the marketing instrument values in the data set, compute the resulting change in consideration share, and combine these to arrive at the elasticities shown.

[d]Estimated on observed choice dummies.

[e]For identification purposes in the choice models, we need to select a base brand and set its intercept equal to 0. Without loss of generality, we have chosen Wisk.

Notes: Bold numbers indicate that 0 is not contained in the 95% HPD interval.

from which the respondents made their final choices. We call these sets the "inferred consideration sets." The self-reported measures of consideration are the stated consideration sets. Note that both stated and inferred consideration sets comprise numbers between 0 and 1 that vary across brands and purchase occasions. To establish the validity of inferring consideration sets from choice data, for each brand, we compute individual and choice occasion, the revealed set membership, and its correlation with stated set membership. We find that revealed and stated set membership correlate highly for each brand. Specifically, for the eight brands, these correlations are in the range of .44 to .78, with an average of .62. These values are lower when we use alternative consideration set models, such as the model in Bronnenberg and Vanhonacker (1996). With their model, the values range from .37 to .60, with an average of .50.

Table 5 shows a cross-tabulation of consideration set memberships. If our model predicts that the brand is not in the consideration set, this prediction is correct in 81.6% of the cases. Similarly, if the model predicts that a brand is in the consideration set, the model is correct in 76.9% of the cases. These hit rates are above the baseline hit rates that stem from random sampling from the marginal distribution of stated consideration, which would imply hit rates of 2329/3456 (67.4%) and 1127/3456 (32.6%), respectively. On a brand-by-brand basis, the hit rates of (rounded) revealed and stated consideration sets range from 62% to 94%. In total, the proposed model is correct in 80% of the purchase occasions.

With respect to the choices, the full model obtains a holdout hit rate of 73%. The MNP model obtains a holdout hit rate of 72%. In sample, these values are 87% and 78%, respectively. We obtained these values through a "jackknife-type" procedure, in which we held out one of the last seven choices for each individual in turn and averaged the hit rates across them. This indicates that our model predicts choices better, though the differences are small.

Table 4
OUT-OF-SAMPLE LOG-PREDICTIVE LIKELIHOODS OF CONSIDERATION SET MEMBERSHIP

	Ajax	All	Arm & Hammer	Bold	Cheer	Surf	Tide	Wisk
Unobserved consideration sets[a]	−28.5	−27.1	−26.6	−22.5	−37.6	−43.8	−69.1	−19.8
Observed consideration sets[b]	−14.2	−27.0	−20.0	−19.1	−33.5	−24.9	−47.3	−18.0

[a]In this model, we ignore consideration set knowledge and estimate the full model on observed choices only.

[b]In this model, we use the consideration set knowledge and estimate the consideration part of the model on the observed consideration set dummies.

In summary, (1) we obtain similar parameter estimates and elasticities of the consideration process estimates from choice and stated consideration data, and (2) the in-sample consideration set forecasts of our full model estimated on choice data alone have a high hit rate of 80%. We take this as strong supportive evidence for the validity of inferring consideration sets from choice data with our model. Our results strongly support the position that consideration, assessed from choice data only, captures more than "statistical artifacts," and this helps us identify the outcome of the consideration processes. To our knowledge, this analysis is the first to provide such support for the convergence of the stated and revealed consideration approaches.

APPLICATION TO SCANNER PANEL DATA

Data

For the illustration of the model on choice data, we consider an optical scanner panel data set on purchases of ten brands of yogurt. The data set contains information on 2382 purchases of yogurt made by 91 households in a large U.S. city. The brands and their respective number of purchase and descriptive statistics appear in Table 6.

Table 5
CROSS-TABULATION OF CONSIDERATION SET MEMBERSHIP: STATED VERSUS INFERRED

	Stated[a] Out (0)	Stated[a] In (1)	Total
Inferred[a]			
Out (0)	2135	481	2616
	81.6%	18.4%	
In (1)	194	646	840
	23.1%	76.9%	
Total	2329	1127	3456[b]

[a]For this cross-tabulation, we rounded both the inferred consideration set memberships and the stated consideration sets to 0 or 1, whichever is nearest.

[b]The data set contains 432 purchase occasions, with eight brands each. Therefore, we have 3456 observations in this cross-tabulation.

Table 6
DESCRIPTIVES FOR THE YOGURT DATA SET

	Number of Purchases: In Sample	Number of Purchases: Holdout Observations	Marketing Instruments: Display %	Marketing Instruments: Feature %	Marketing Instruments: Price per Ounce ($)
Dannon Plain	478	13	2.0	5.1	.076
Dannon Blended	117	5	2.9	6.3	.084
Dannon Classic Flavor	82	4	2.4	4.8	.078
Dannon Fruit[a]	235	10	1.6	6.7	.080
Private label	65	5	3.1	9.2	.067
Yoplait	249	13	1.3	4.9	.092
Yoplait Custard Style	124	7	1.2	5.3	.104
Dannon Light	343	8	2.9	6.2	.084
Kemps Lite	206	8	1.6	7.0	.069
Yoplait Light	84	7	1.0	5.7	.092
Total	1983	80			

[a]Abbreviation of Fruit on the Bottom.

The variation in choice shares of the brands is somewhat higher than for the experimental data in Table 2. The display frequency is rather low, likely because of limited space for displays in the refrigerated section of a supermarket, and it does not vary across brands to a great extent.

Estimation Results from the Empirical Data

We consider the following three models on the yogurt data, using the same operationalizations as described for the experimental data: First, we consider our own model. Again, this involves estimating the consideration effects α and α_i, the covariance matrix Σ_α of the random consideration set effects α_i, the covariance of consideration Σ, the choice effects β and β_i, the covariance matrix Σ_β of the random consideration set effects β_i, and the (diagonal) covariance matrix of the choice utilities Ω.[2] Second, we consider a single-stage choice model with similar specifications. Third, we estimate Bronnenberg and Vanhonacker's (1996) model.

The posterior results for the yogurt data appear in Table 7. The results of the proposed model show that the posterior means of all marketing parameters are far away from 0 (compared with the posterior standard deviation) and that they are all of the expected sign. Consistent with the controlled choice experiment, the posterior means of the covariance terms in Σ in the consideration model are close to 0, and all HPD regions cover the value zero. This does not appear to be caused by the unobserved heterogeneity, because it also happens when we estimate the model without unobserved heterogeneity in the consideration stage. Just as in the experimental data, it seems that after we

[2]In our empirical application, the estimation of the diagonal elements of Ω appeared to result in instability of the MCMC, which is evidence that these parameters are only weakly identified by the data. Because of this, we chose to set Ω equal to the identity matrix.

Table 7

POSTERIOR MEANS AND STANDARD DEVIATIONS FOR THE FULL MODEL FOR THE YOGURT DATA[a]

	Estimates		*Elasticities*	
Consideration Stage (PCS[b])	*M*	*SD*	*Display*	*Feature*
$\alpha_{DannonPlain}$	–.001	.615	.159	.258
$\alpha_{DannonBlended}$	**–.925**	.361	.523	.604
$\alpha_{DannonClassicFlavor}$	**–1.650**	.412	.871	.837
$\alpha_{DannonFruit}$	–.759	.404	.436	.406
$\alpha_{PrivateLabel}$	**–1.760**	.278	.853	.987
$\alpha_{Yoplait}$	**–.948**	.395	.577	.483
$\alpha_{YoplaitCustardStyle}$	–.211	.673	.155	.292
$\alpha_{DannonLight}$	–.125	.391	.214	.165
$\alpha_{KempsLite}$	–.590	.764	.338	.439
$\alpha_{YoplaitLight}$	–.897	.625	.620	.576
Display	**3.800**	.786		
Feature	**4.110**	.927		
Brand Stage (MNP)	*M*			*SD*
$\beta_{DannonPlain}$	**1.280**			.369
$\beta_{DannonBlended}$	–.084			.394
$\beta_{DannonClassicFlavor}$	–.015			.384
$\beta_{DannonFruit}$	**1.070**			.404
$\beta_{PrivateLabel}$	.047			.362
$\beta_{Yoplait}$	**.931**			.450
$\beta_{YoplaitCustardStyle}$	**–1.130**			.558
$\beta_{DannonLight}$	**.819**			.369
$\beta_{KempsLite}$	–.372			.528
$\beta_{YoplaitLight}$[c]	0			—
Price	**–5.100**			1.250

[a]The posterior means of the covariances in the PCS model are close to 0 and are not displayed here.

[b]PCS = probabilistic choice set model.

[c]For identification purposes, we need to select a base brand and set its intercept equal to 0. Without loss of generality, Yoplait Light is chosen as base brand.

Notes: Bold numbers indicate that 0 is not contained in the 95% HPD interval.

take into account in-store variables and unobserved heterogeneity, little covariation among consideration of brands is left. However, returning to Equation 3, we observe that our specification with unobserved heterogeneity enables us to inspect interbrand correlations. Indeed, we find significant correlations between the following brands: Dannon Plain and Dannon Classic Flavor (positive), Dannon Plain and Yoplait Light (negative), and Dannon Classic Flavor and Yoplait Light (negative). This is intuitive because Dannon Plain and Dannon Classic Flavor are both plain yogurts and share the same brand. In contrast, the two negatively correlated brand pairs do not share any attributes at all.

We computed log-predictive densities for the model with an identity correlation matrix and the model with a full correlation matrix. The LPL of the model with identity correlation matrix is −125. For the model with a full correlation matrix, the out-of-sample LPL is −127. These numbers again show that the model with the identity correlation matrix is weakly favored, and again we find empirical support for the assumption of independence of consideration set membership across brands. This implies that given the individual-level parameters α_i, for an alternative to enter the consideration set, it does not matter greatly which alternatives are already in it. The independence assumption has been extensively used in the stream of research that uses the stated consideration set approach. We calibrated versions of our model on three other data sets (cracker data with four brands, soft drinks with ten brands, and yogurt with ten brands), and after controlling for heterogeneity, we found little or no correlation among the errors in the consideration set stage either.

In Table 8, we present the estimation results of the the single-stage MNP model. Contrasting this with Table 7, the effects for feature and display are much stronger in the consideration stage of the proposed model than in the choice model. We point to the large difference in the price coefficient, which is two times larger than in the choice model. This finding, which has been previously documented in the literature, suggests that when a brand has entered the consideration set, the price instrument is effective in increasing market share. However, the parameters of the full model to the choice model may not be directly compared. To facilitate comparison, we computed brand-specific price elasticities for the full model (unconditional on consideration) and the single-stage MNP model. Across brands, the elasticities for the full model average −.709 and range from −.443 to −.952, while price elasticities for the single-stage model have an average of −.619 and range from −.533 to −.712. For six of the ten price elasticities, the corresponding 95% HPD intervals overlap. Although not true for every brand, these numbers confirm that the same price cut for a particular brand leads to a higher expected market share (unconditional on consideration) in the two-stage consideration and choice model than in the single-stage choice model presented in Table 8. However, there are quantitative differences in the extent to which the two models predict these effects for specific brands.

Table 8

POSTERIOR MEANS AND STANDARD DEVIATION FOR THE MNP MODEL FOR BRAND CHOICE ONLY FOR THE YOGURT DATA[a]

MNP	*M*	*SD*
$\beta_{DannonPlain}$	1.015	.286
$\beta_{DannonBlended}$	.115	.306
$\beta_{DannonClassicFlavor}$	−.613	.386
$\beta_{DannonFruit}$	.681	.269
$\beta_{PrivateLabel}$	−.147	.290
$\beta_{Yoplait}$	.479	.238
$\beta_{YoplaitCustardStyle}$	−.460	.409
$\beta_{DannonLight}$	.910	.252
$\beta_{KempsLite}$	−.565	.403
$\beta_{YoplaitLight}$[b]	0	—
Display	.972	.279
Feature	1.076	.192
Price	−3.343	2.244

[a]To ease the comparison with the PCS (probabilistic choice set)+MNP model results, this single-stage MNP model has an identity covariance matrix and unobserved heterogeneity.

[b]For identification purposes, we need to select a base brand and set its intercept equal to 0. Without loss of generality, we chose Yoplait Light as base brand.

Table 9
SHARES OF PURCHASE, CONSIDERATION, AND PURCHASE CONDITIONAL ON CONSIDERATION

Brand	*Purchased (Observed) %*	*Considered (Retrieved) %*	*Conditional Purchase Share[a] %*
Dannon Plain	31	60	51
Dannon Blended	5	38	13
Dannon Classic Flavor	4	32	13
Dannon Fruit	9	41	22
Private label	4	23	17
Yoplait	12	35	34
Yoplait Custard Style	6	53	12
Dannon Light	15	58	26
Kemps Lite	10	45	22
Yoplait Light	4	36	11

[a]We computed this as the ratio of the purchase and consideration shares.

Next, we consider in greater detail the consideration sets themselves in Table 9, which reports the consideration share, market share, and the share of purchase, conditional on consideration (the ratio of the two aforementioned shares). The table shows, for example, that Yoplait Custard Style has a high consideration share of 53%; however, it has a difficult time converting this consideration into sales, as is evident from the low conditional purchase share. The market leader, Dannon Plain, has no trouble in this respect: The conditional purchase share is more than 50%.

Although the model is not purposely built to make forecasts, out-of-sample predictions show that the LPL for the single-stage MNP model equals −128, whereas our model attains a slightly higher LPL of −125. The holdout hit rate of the full model is 56%. The single-stage choice model produces the same hit rate. When we apply Bronnenberg and Vanhonacker's (1996) model to our data set, we obtain a holdout hit rate of 50%. The posterior distributions of out-of-sample forecast hit rate for the single-stage choice model and our consideration and choice model overlap to a large extent. We would have liked the added complexity of our model to result in substantial improvements in predictive performance, as it did for the experimental data described previously, but as has been found previously, a simpler but theoretically less completely specified model, such as the MNP, also predicts well. We believe that the major advantage of our model accrues from its diagnostic value. We conjecture that the main reason that estimation of consideration set formation is important to a marketing manager may not be prediction but rather the insights it provides into both competitive and positioning issues ("Who are we competing against in the mind of the consumer?" "What is my vulnerability to competitive attacks?") and control issues ("What will be the effect of my marketing-mix variables in various stages?"). With these important issues, the insights derived from single-stage and two-stage models of choice really may differ.

Optimization Under Different Model Assumptions

In this section, we show how optimal decisions on price cuts with feature ad placement can be derived for the full model, and we compare these with those derived from the choice model. In our optimization, we assume that the retailer communicates price cut to its consumers through feature advertising.

To estimate the impact of price discounts, we predict posterior market shares in the holdout sample. We project these shares to a market of 1000 consumers. In this hypothetical market, it costs $10 to implement a feature.[3] We can use the prediction of demand to compute a posterior distribution of gross profits by multiplying the predicted volume by the (assumed) profit margins of the products.[4] The optimization criterion is integrated over the posterior distribution of the parameters (Rossi, Allenby, and McCulloch 2005). In every week of the holdout sample, a decision is made on whether a price cut is featured and, if so, what the magnitude of the price cut is.

The algorithm we use to maximize the profits is a version of steepest descent. In a certain state, we randomly change the price cuts that are present in the data. If this candidate state results in higher profits, the candidate state becomes the current state. This is done for 2500 iterations. We use 1000 random starting points to avoid local optima. These settings appeared to be sufficient: We found no profit improvement well before the end of the run.

In Table 10, we report the results for the full model and the MNP choice model. The optimal promotion strategy is markedly different depending on which assumptions about demand are made. The optimal promotion strategies both call for roughly half the brands to be promoted at least once in the holdout period, but there is only one brand that should be promoted according to both models. As a consequence, if a manager assumes that the MNP choice model is correct, but consumers use a two-stage process and the full model is correct in reality, the promotion strategy will be nonoptimal. Assuming that consumers indeed use a two-stage choice process, the profit associated with the optimal strategy derived from the full model is $93. However, the profit earned from the strategy derived with the MNP choice model is only $37. For comparison, the profit for the actual (current) promotion strategies is $77 when computed under this assumption.

Thus, potentially sizable improvements can be made to the current profits by optimizing the promotional plan, and when demand follows a two-stage process, an optimization using the MNP choice model would result in a large loss of profit. Therefore, we conclude that in making optimal marketing decisions, it is imperative—both theoretically and practically—to use models that accommodate consumer consideration.

CONCLUSION

Entering consumers' consideration set is a top priority in marketing strategy, and the implementation of those strategies is contingent on knowledge of individual consumers' consideration sets. Such knowledge has been obtained either by asking respondents to state their considered set

[3]We experimented with various values of these costs, but it did not change the results substantively.

[4]The results are shown for variable costs equal to 60% of the price. We varied this percentage, but it resulted in the same implications.

Table 10
DESCRIPTIVE STATISTICS FOR OPTIMAL PRICE CUTS FOR THE FULL MODEL AND THE MNP MODEL

	Market Share (%)		*Number of Features*		*Average Price Cut ($)*	
	Full	*MNP*	*Full*	*MNP*	*Full*	*MNP*
Dannon Plain	19.8	20.5	0	1	—	.014
Dannon Blended	11.6	12.3	0	1	—	.034
Dannon Classic Flavor	6.9	5.2	2	0	.012	—
Dannon Fruit	17.4	14.7	1	0	.016	—
Private label	5.3	6.5	0	1	—	.015
Yoplait	11.3	12.4	1	0	.018	—
Yoplait Custard Style	4.4	4.3	1	0	.020	—
Dannon Light	16.4	16.0	0	1	—	.065
Kemps Lite	5.7	5.1	0	0	—	—
Yoplait Light	.2	.3	2	3	.027	.020

of brands or by inferring those sets from their revealed choices. We have proposed, operationalized, and estimated a new model to capture unobserved consideration from discrete choice data. This model accommodates different sets of marketing-mix variables and heterogeneity in both consideration and choice stages, is computationally feasible for large numbers of alternatives, and bridges the stated and revealed approaches, enabling the analysis of either one or both sources of data to infer sets of brands considered for purchase. Thus, the model enables us to shed light on the long-standing question whether consideration sets can be validly inferred from revealed choice data (see Roberts and Lattin 1997) and to study the convergent validity of stated and revealed consideration sets. Although further research in this area is needed, our first findings are promising, and we find support for making inferences about consideration from revealed choice behavior using our model.

We calibrated our model on two data sets. In the first, we observed consideration sets, enabling us to compare the inferred stated consideration sets. Our model appeared to be able to predict the assumed unobserved sets well. Furthermore, the parameters estimated from the revealed and stated consideration sets appear to correlate highly (.96). Elasticities for the marketing instruments are similar as well, thus confirming this resemblance. These findings show that our model is able to pick up unobserved consideration sets well. Finally, the predictive validity of the model compared favorably with that of a simple MNP choice model. However, the differences are small, and it is difficult to make strong claims about statistical superiority of our model in this respect. However, from a theoretical perspective, allowing consumers to choose from a limited set of product alternatives has appeal.

The application of our model to a scanner panel data set revealed that after a product is in the consideration set, price is an extremely effective competitive instrument—more so than predicted by previous single-stage models. A simulation experiment to develop optimal price cuts and featuring showed that our model gives markedly different results than a single-stage model. Perhaps most important, when using the optimum that results from the single-stage model in a world in which two-stage models are the reality, profits are relatively low. This shows that using a two-stage model is crucial for improving the effectiveness of such marketing decisions.

The proposed model enables different explanatory variables to be included in the consideration and choice stages. We included in-store merchandizing (display and feature) in the consideration stage (see Alba and Chattopadhyay 1985), and we included brand intercepts and prices in the choice stage. Theory specifies how these variables should affect consideration and choice (e.g., Mitra and Lynch 1995). Our two-stage model offers a more appealing interpretation for the role of in-store merchandizing on consumer choice than a single-stage model. In the two-stage model, in-store merchandizing has information effects. In contrast, the implication of a single-stage model is that display and feature are components of brand utility. However, feature advertisements and display do not generate the same utility as when paying a low price or receiving high quality of a brand. Rather, the role of these variables is to facilitate—that is, lower the cost of—consideration of brands (see Andrews and Srinivasan 1995). Therefore, in-store merchandizing programs are viewed as suitably fulfilling the goal of lowering the mental cost of information acquisition (for a more detailed representation of the role of feature advertisements and display on consideration and choice, see Zhang 2006). In conclusion, we believe that our model is a useful tool for analyzing both stated and revealed consideration data and for studying the role of consideration set formation in choice behavior.

REFERENCES

Alba, Joseph W. and Amitava Chattopadhyay (1985), "The Effects of Context and Part-Category Cues on Recall of Competing Brands," *Journal of Marketing Research*, 22 (August), 340–49.

Allenby, Greg M. and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89 (1–2), 57–78.

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

Belonax, Joseph A. and Robert A. Mittelstaedt (1978), "Evoked Set Size as a Function of Number of Choice Criterion and Information Variability," in *Advances in Consumer Research*, Vol. 5, Kent Hunt, ed. Ann Arbor, MI: Association for Consumer Research, 48–51.

Ben-Akiva, Moshe and Bruno Boccara (1995), “Discrete Choice Models with Latent Choice Sets,” *International Journal of Research in Marketing*, 12 (1), 9–24.

Bettman, James R. (1979), *An Information Processing Theory of Consumer Behavior*. Reading MA: Addison-Wesley.

Bjørnstad, Jan F. (1990), “Predictive Likelihood: A Review,” *Statistical Science*, 5 (2), 242–65.

Bronnenberg, Bart J. and Wilfried R. Vanhonacker (1996), “Limited Choice Sets, Local Price Response, and Implied Measures of Price Competition,” *Journal of Marketing Research*, 33 (May), 163–74.

Bunch, David S. (1991), “Estimatibility in the Multinomial Probit Model,” *Transportation Research: B*, 25 (1), 1–12.

Chiang, Jeonwen, Siddhartha Chib, and Chakravarthi Narasimhan (1999), “Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity,” *Journal of Econometrics*, 9 (1–2), 223–48.

Corstjens, Judith and Marcel Corstjens (1999), *Store Wars: The Battle for Mindspace and Shelfspace*. Chichester, UK: John Wiley & Sons.

Dickson, Peter R. and Alan G. Sawyer (1990), “The Price Knowledge and Search of Supermarket Shoppers,” *Journal of Marketing*, 54 (July), 42–53.

Edwards, Yancy D. and Greg M. Allenby (2003), “Multivariate Analysis of Multiple Response Data,” *Journal of Marketing Research*, 40 (August), 321–34.

Geweke, John (2005), *Contemporary Bayesian Econometrics and Statistics*. New York: John Wiley & Sons.

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

Guadagni, Peter M. and John D.C. Little (1983), “A Logit Model of Brand Choice Calibrated on Scanner Data,” *Marketing Science*, 2 (3), 203–238.

Hauser, John R. and Birger Wernerfelt (1989), “The Competitive Implications of Relevant-Set/Response Analysis,” *Journal of Marketing Research*, 26 (November), 391–405.

——— and ——— (1990), “An Evaluation Cost Model of Consideration Sets,” *Journal of Consumer Research*, 16 (4), 393–408.

Hausman, Jerry and David Wise (1978), “A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences,” *Econometrica*, 45 (2), 319–39.

Howard, John A. and Jagdish N. Sheth (1969), *The Theory of Buyer Behavior*. New York: John Wiley & Sons.

Keane, Michael P. (1992), “A Note on Identification in the Multinomial Probit Model,” *Journal of Business & Economic Statistics*, 10 (2), 193–200.

Lehmann, Donald. R. and Yigang Pan (1994), “Context Effects, New Brand Entry, and Consideration Sets,” *Journal of Marketing Research*, 31 (August), 364–74.

Lussier, Denis A. and Richard W. Olshavsky (1979), “Task Complexity and Contingent Processing in Brand Choice,” *Journal of Consumer Research*, 6 (2), 154–65.

Malhotra, Naresh K., Mark Peterson, and S. Bardi Kleiser (1999), “Marketing Research: A State-of-the-Art Review and Directions for the Twenty-First Century,” *Journal of the Academy of Marketing Science*, 27 (2), 160–83.

Manrai, Ajay K. and Rick L. Andrews (1998), “Two-Stage Discrete Choice Models for Scanner Panel Data: An Assessment of Process and Assumptions,” *European Journal of Operational Research*, 111 (2), 193–215.

Manski, Charles F. (1977), “The Structure of Random Utility Models,” *Theory and Decision*, 8 (3), 229–54.

McFadden, Daniel (1973), “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers in Econometrics*, P. Zarembka, ed. New York: Academic Press, 105–142.

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

Mitra, Anusree (1995), “Advertising and the Stability of Consideration Sets over Multiple Purchase Occasions,” *International Journal of Research in Marketing*, 12 (1), 81–94.

——— and John G. Lynch Jr. (1995), “Toward a Reconciliation of Market Power and Information Theories of Advertising Effects on Price Elasticity,” *Journal of Consumer Research*, 21 (4), 644–59.

Nedungadi, Prakash (1990), “Recall and Consumer Consideration Sets: Influencing Choice Without Altering Brand Evaluations,” *Journal of Consumer Research*, 17 (3), 263–76.

Newell, Allen and Herbert A. Simon (1972), *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

Payne, John W. (1976), “Task Complexity and Contingent Processing in Decision-Making: An Information Search and Protocol Analysis,” *Organizational Behavior and Human Performance*, 16 (2), 366–87.

Roberts, John H. and James M. Lattin (1991), “Development and Testing of a Model of Consideration Set Composition,” *Journal of Marketing Research*, 28 (November), 429–40.

——— and ——— (1997), “Consideration: Review of Research and Prospects for Future Insights,” *Journal of Marketing Research*, 34 (August), 406–410.

Rossi, Peter E., Greg M. Allenby, and Rob McCulloch (2005), *Bayesian Statistics and Marketing*. New York: John Wiley & Sons.

Siddarth, S., Randolph E. Bucklin, and Donald Morrison (1995), “Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data,” *Journal of Marketing Research*, 32 (August), 255–66.

Wright, P. and F. Barbour (1977), “Phased Decision Strategies: Sequels to Initial Screening,” in *Multiple Criteria Decision Making: North-Holland TIMS: Studies in the Management Science*, M. Starr and M. Zeleny, eds. Amsterdam: North-Holland, 91–109.

Zhang, Jie (2006), “An Integrated Choice Model Incorporating Alternative Mechanisms for Consumers’ Reactions to In-Store Display and Feature Advertising,” *Marketing Science*, 25 (3), 278–90.