ERIC T. BRADLOW, YE HU, and TECK-HUA HO*

Respondents in a conjoint experiment sometimes are presented with successive partial product profiles. First, the authors model how respondents infer missing levels of product attributes in a partial conjoint profile by developing a learning-based imputation model that nests several extant models. The advantage of this approach over previous research is that it infers missing levels of an attribute not only from prior levels of the same attribute but also from prior levels of other attributes, especially ones that match the attribute levels of the current product profile. Second, the authors provide an empirical demonstration of their approach and test whether learning in conjoint studies occurs; to what extent; and in what manner it affects responses, partworths, and the relative importance of attributes. They show that the relative importance of attribute partworths can shift when subjects evaluate partial profiles, which suggests that consumers may construct rather than retrieve partworths and are sensitive to the order in which the profiles are presented. Finally, the results show that consumers' imputation processes can be influenced by manipulating their prior information about a product category. This research is of both theoretical and practical importance. Theoretically, this research sheds light on how customers integrate different sources of information in evaluating products with incomplete attribute information; practically, it highlights the potential pitfalls of imputing missing attribute levels using simple rules and develops a better behavioral model for describing and predicting customers' ratings for partial conjoint profiles.

A Learning-Based Model for Imputing Missing Levels in Partial Conjoint Profiles

Conjoint analysis is perhaps the most celebrated research tool in marketing. It has been applied to solve a wide variety of marketing problems, ranging from understanding consumer preferences, to estimating product demand, to designing a new product line. The method involves presenting customers with a carefully chosen test set of product profiles from the universal set (as defined by the levels of the attributes) and collecting their preferences, which could be ratings, rankings, or profiles, for those profiles in the test set. The power of the method lies in its ability to extrapolate customers' preferences from this test set to the universal set. Conjoint analysis works better when the test set is small and the preference task difficulty is low. Both factors can be influenced significantly by the number of product attributes.

If the number of attributes is large (as for many high-tech durable products), a full-factorial experiment would require respondents to assess their preferences for many profiles, each consisting of many attributes. The large test set problem can be solved by means of a fractional design (Green, Carroll, and Carmone 1978; Plackett and Burman 1946) that divides the test set among several respondents in a common customer segment.

There are two ways to solve the task difficulty problem. The first way is to use a self-explicated conjoint analysis (Green 1984), in which consumers rate the importance of the attributes and then evaluate the attractiveness of each attribute level. By multiplying the normalized importance and attractiveness ratings, the researcher can derive a consumer's overall preference for any profile. This approach typically requires the respondent to answer a smaller set of

*Eric T. Bradlow is Associate Professor of Marketing and Statistics and Academic Director of the Wharton Small Business Development Center, The Wharton School, University of Pennsylvania (e-mail: ebradlow@wharton.upenn.edu). Ye Hu is a visiting assistant professor, Krannert School of Management, Purdue University (e-mail: hu18@purdue.edu). Teck-Hua Ho is William Halford Jr. Family Professor of Marketing, Haas School of Business, University of California, Berkeley (e-mail: hoteck@haas.berkeley.edu). Professor Ho was partially supported by a grant from the Wharton-SMU Research Center, Singapore Management University. The authors thank David Robinson and Young Lee for their help in data collection and seminar participants at University of California at Berkeley, Singapore Management University, University of Michigan at Ann Arbor, the Marketing Science 2002 conference, and the BAMMCONF 2002 for their useful suggestions.

Journal of Marketing Research
Vol. XLI (November 2004), 369–381

questions than a full-profile judgment task does, and it avoids the complexity of judging a profile with too many attributes. However, the self-explicated conjoint analysis method has its own problems, including that respondents' attribute importance ratings are not always consistent with their preference decisions. That is, the experimental condition of separating attribute and level ratings is artificial because real-life purchase decisions are made on whole products.

The second solution is to use orthogonal subsets of all the attributes (Green 1974), or partial-profile conjoint analysis. Because profiles with a smaller number of attributes may be easier to rate, the partial-profile conjoint analysis approach decreases the difficulty of the rating task; however, it may increase the number of profiles needed to determine consumers' utility functions. The partial-profile conjoint analysis approach typically assumes that the attributes that are missing do not influence product evaluation, but several studies cast doubt on this assumption (e.g., Broniarczyk and Alba 1994; Feldman and Lynch 1988; Huber and McCann 1982). Consequently, standard rating conjoint analysis methods that are applied to partial profiles may not produce a highly predictive utility function.

In this article, we investigate how subjects impute missing attribute levels when they evaluate partial conjoint profiles. Our goals are to understand the dependency of ratings of current profiles on all available attribute information (in both the current and the previously shown profiles), including a person's prior knowledge, and to provide insights into how consumers may impute missing levels when evaluating partial conjoint profiles.

We relax the "null effect for missing attribute" assumption and develop a probabilistic model of how respondents impute values for missing attributes based on their priors over the set of attribute levels, a given attribute's previously shown values, the previously shown values of other attributes, and the covariation among attributes (both a priori and learned within the task itself). We conceptualize how consumers infer missing values through a pattern-matching and learning process. Our model assumes that consumers learn and update after each stimulus (partial profile) about the pattern underlying the product attributes, their levels, and the correlations between them. How are strengths of patterns formed and updated? We assume that consumers have prior knowledge about the patterns and use knowledge about the product profiles acquired through the conjoint task to update their strengths. It is this dynamic process of learning about the attribute level occurrences and covariation among attributes that we model and focus on in this article.

We call the fundamental kernel of this updating structure a "pattern-matching" learning model. That is, the respondent uses previously shown profiles that exhibit certain patterns among the attributes to infer the missing attribute levels in the current profile. In essence, this approach can be viewed as a time-varying, multiway contingency table of latent counts for imputation (Little and Rubin 1987). Consequently, the order in which profiles are presented matters in predicting preference ratings.

We model how people *rate* partial conjoint profiles over time. Although rating-based methods may currently be less common than choice-based methods in practice (Wittink and Cattin 1989), our study is relevant for common applications of conjoint analysis in at least three ways:[1]

1. Our research can influence the way that adaptive conjoint analysis (ACA) is used in practice. The ACA engine (e.g., Equation 2 in the ACA 5.0 Technical Report [Sawtooth Software 2002]) requires continuous strength of preference data, treated as a rating score, obtained for pairwise partial profiles. Although ACA can handle up to 30 attributes, each profile should contain no more than 5 attributes, a practice that has been brought into question by others (e.g., Green, Krieger, and Agarwal 1991; Johnson 1991). Our work is directly applicable to the ACA engine, which selects profile pairs on the basis of utility balance, and if those utilities are influenced by the missing attributes that do not cancel across choice pairs because of covariation, learning, and so forth, the resultant partworths may be biased. Our model helps quantify these biases or select pairs that have the highest likelihood of canceling out those missing attributes.
2. Our research has implications for the hybrid approach proposed by Srinivasan and Park (1997) and subsequently extended by Ter Hofstede, Kim, and Wedel (2002). Both share the same data structure, in which the authors use a subset of the most important attributes, based on self-explicated data, in a subsequent full-profile conjoint study. Respondents are aware of all the attributes before they rate the profiles that contain only a subset of attributes. In this manner, the profiles shown in the approaches are partial, and our model sheds light on the role of those attributes that are excluded in the rating task.
3. Our study is also relevant for choice-based conjoint methods, despite the assumption of ignorability across pairs of not shown attributes (Elrod, Louviere, and Davey 1992). Conceptually, all profiles, even if they are called full, have missing attributes that could be inferred. Therefore, despite the common practice of stating that "respondents were instructed that profiles were similar in every respect except possibly for those attributes shown in the profile description," it is an open empirical question whether respondents do or, possibly more important, can follow this instruction. Our model can verify whether subjects are able to ignore the levels of those attributes that are not included in the study.

The rest of this article is composed of four sections. In the next section, we develop our imputation model. We then explain the design of our experiment. Next, we estimate our model on two sets of experimental data and report the results. In the last section, we conclude with caveats and provide directions for further research.

THE IMPUTATION MODEL

Notation and Model Setup

We investigate how I respondents (indexed by $i = 1, \ldots, I$) rate a series of product profiles (partial or full) in a conjoint experiment. Each product profile is characterized by J attributes (indexed by $j = 1, \ldots, J$), and each attribute j has two levels. Each respondent i rates T profiles (denoted by $M_i(t)$, $t = 1, \ldots, T$) one by one. Respondent i's rating for product profile $M_i(t)$ is given by $y_i(t)$. Profile $M_i(t)$ takes a level of $x_{ij}(t) = 1$ or 0 for attribute j. We denote whether attribute j is missing in the tth profile shown to respondent i by $r_{ij}(t)$. If attribute j is missing in profile $M_i(t)$, $r_{ij}(t)$ is 0; otherwise, $r_{ij}(t)$ is 1. The basic premise of our model is that

[1]We thank an anonymous reviewer for pointing out the link between our study and these approaches.

respondent i does not ignore a missing attribute level but rather constructs an imputed value for it. Let $x'_{ij}(t)$ be that imputed value, determined as follows:

$$x'_{ij}(t) = \begin{cases} x_{ij}(t) & \text{if } r_{ij}(t) = 1, \text{ and} \\ Ix_{ij}(t) & \text{if } r_{ij}(t) = 0. \end{cases}$$

Note that $Ix_{ij}(t)$ is a random variable that takes the value of 1 or 0 (or, in general, the possible values of $x_{ij}(t)$), and the imputation modeling effort is to determine its probability distribution over the possible levels. If an attribute is not missing (i.e., $r_{ij}(t) = 1$), we assume that the shown attribute level $x_{ij}(t)$ occurs with probability 1.

To determine the partworths of the attributes, we postulate a regression with heterogeneous coefficients given by

$$(1) \qquad y_i(t) = \alpha_i + \sum_{j=1}^{J} \beta_{ij} x'_{ij}(t) + \varepsilon_i(t),$$

where β_{ij} is respondent i's partworth for attribute j.

Johnson, Levin, and their colleagues (Johnson 1987; Johnson and Levin 1985; Levin et al. 1986), however, suggest that subjects have different partworths for the same attribute when it is missing compared with when it is not. To control for this, we modify regression Equation 1 to yield the following:

$$(2) \qquad y_i(t) = \alpha_i + \sum_{j=1}^{J} \left[\beta_{ij} x'_{ij}(t) + \beta'_{ij} r_{ij}(t)\right] + \varepsilon_i(t).$$

In Table 1, we show the model's partworths in different conditions. If subjects have different partworths when an attribute is missing, then β'_{ij} will be significantly different from 0. Thus, our model nests the work of Johnson, Levin, and others and generalizes theirs by including the imputed attribute levels $x'_{ij}(t)$ when $r_{ij}(t) = 0$.

Basic Ideas

In Table 2, we show a hypothetical example that introduces the basic ideas of our imputation model and demonstrates current extant models. There are four attributes (i.e., J = 4) and three profiles, that is, $M_i(1)$, $M_i(2)$, $M_i(3)$. Each profile has one missing attribute (denoted as MA), where subject i rates $M_i(1)$, $M_i(2)$, and $M_i(3)$ sequentially. At time t = 3, $x_{i1}(3) = 1$, $x_{i3}(3) = 1$, and $x_{i4}(3) = 0$. Attribute 2 is missing at time t = 3, and we denote its imputed level, 1 or 0, by $x'_{i2}(3)$. In a real experiment, the respondent views a product profile at time 3 with only Attributes 1, 3, and 4 and does not see "MA" for Attribute 2; we include it in Table 2 only to describe the design.

Assume that the subject has finished rating profiles $M_i(1)$ and $M_i(2)$ and that profile $M_i(3)$ is the current product profile. To investigate how information from different attributes might influence the imputed value for Attribute 2 at t = 3, we divide the attributes into three types: (1) the omnipresent (OM) set, (2) the presence-manipulated (PM) set with present attributes (nonmissing PM), and (3) the PM set with missing attributes (missing PM). The OM attributes are always presented, whereas the PM attributes may or may not be. A PM attribute is called a "nonmissing PM attribute" if it is not missing in the current conjoint profile and a "missing PM attribute" if it is missing in the current conjoint profile, though it may not be missing in others. In profile $M_i(3)$, Attribute 1 is an OM attribute, Attributes 3 and 4 are nonmissing PM attributes, and Attribute 2 is a missing PM attribute. Existing models use only prior information (values) from the currently missing PM attribute (Attribute 2) for imputation of the missing level $x_{i2}(3)$. Our model uses all three sources: missing and nonmissing PM and OM attributes.

There are several different ways to treat missing attribute levels. The first way is to assume that respondents ignore them (Green 1974). Such an assumption implies that all the MAs in Table 1 are filled in as 0 (the default level). Formally, this assumption leads to the following prediction of the missing attribute level: $\Pr[x'_{i2}(3) = 0] = 1$ and $\Pr[x'_{i2}(3) = 1] = 0$. Note that in this case, the imputation process of the missing levels depends on which level is coded as 0, a potential theoretical problem.

An alternative approach is to assume that respondents impute values using all available information. That is, people infer the levels of the missing attributes from previously viewed product profiles and weight each profile pattern accordingly. This latter view is consistent with the work of Meyer (1981), who shows that when a subject has no information about certain attributes, he or she does not ignore that attribute but rather assigns it a score equal to his or her adaptation level.

There are two common ways to model how consumers make inferences about missing attribute levels. The first way is based on the so-called recency effect (Lynch and Srull 1982); people assume that the missing attribute level is the last level of the same attribute they saw. According to such a model, in Table 2, Attribute 2 in profile $M_i(3)$ takes level 1 (following the level of Attribute 2 in profile $M_i(2)$). Formally, we have $\Pr[x'_{i2}(3) = 0] = 0$ and $\Pr[x'_{i2}(3) = 1] = 1$.

The second commonly used imputation approach is the averaging model (Yamagishi and Hill 1981); people impute the missing attribute level by averaging all the previously shown levels of the missing attribute. For example, in Table 2, this yields $\Pr[x'_{i2}(3) = 0] = ½$ and $\Pr[x'_{i2}(3) = 1] = ½$.

Table 1
PARTWORTHS OF THE SAME ATTRIBUTE

		Attribute Level	
		High: $x'_{ij}(t) = 1$	*Low:* $x'_{ij}(t) = 0$
Attribute Shown?	Yes	$\beta_{ij} + \beta'_{ij}$	β'_{ij}
	No	β_{ij}	0

Table 2
AN ILLUSTRATIVE EXAMPLE

	OM	*PM*		
Time (t)	*Attribute 1*	*Attribute 2*	*Attribute 3*	*Attribute 4*
1	1	0	MA*	1
2	0	1	1	MA
3	1	MA	1	0

*MA = missing attribute.

In imputing the missing values, the recency and averaging models make strong assumptions about the similarity between the current and previous profiles. The recency-based model assumes that the current profile is similar only to the most recently shown profile and is dissimilar from the rest.[2] The averaging model assumes that the current profile is equally similar to all the previously shown profiles. However, we expect that some previously shown profiles are more similar ("count more" in imputing) to the current profile than others.

These models use only prior data from Attribute 2 to impute $x'_{i2}(3)$. By doing so, they ignore two important pieces of information. First, there is the complete set of patterns shown to the subjects $[M_i(1), M_i(2)]$, not just the values for Attribute 2 $[x_{i2}(1), x_{i2}(2)]$. Some of these patterns might occur more frequently, so their values for Attribute 2 might be more salient and memorable. Second, levels of other attributes from the current profile (i.e., $M_i(3)$) might be diagnostic about the missing level. For example, if Attribute 1 is negatively correlated with Attribute 2, as in Table 2, a respondent might infer a 0 for Attribute 2 from a 1 in Attribute 1. Such a correlation structure could be based on people's long-term memory or learning in the conjoint task. Huber and McCann (1982) show that people use their belief about the correlation structure between price and quality to infer the missing price or quality when either one is missing. Broniarczyk and Alba (1994) also show that consumers' intuitions (priors) influence their inference making. Our model captures these covariances, as well as the priors they already have, in a parsimonious way.

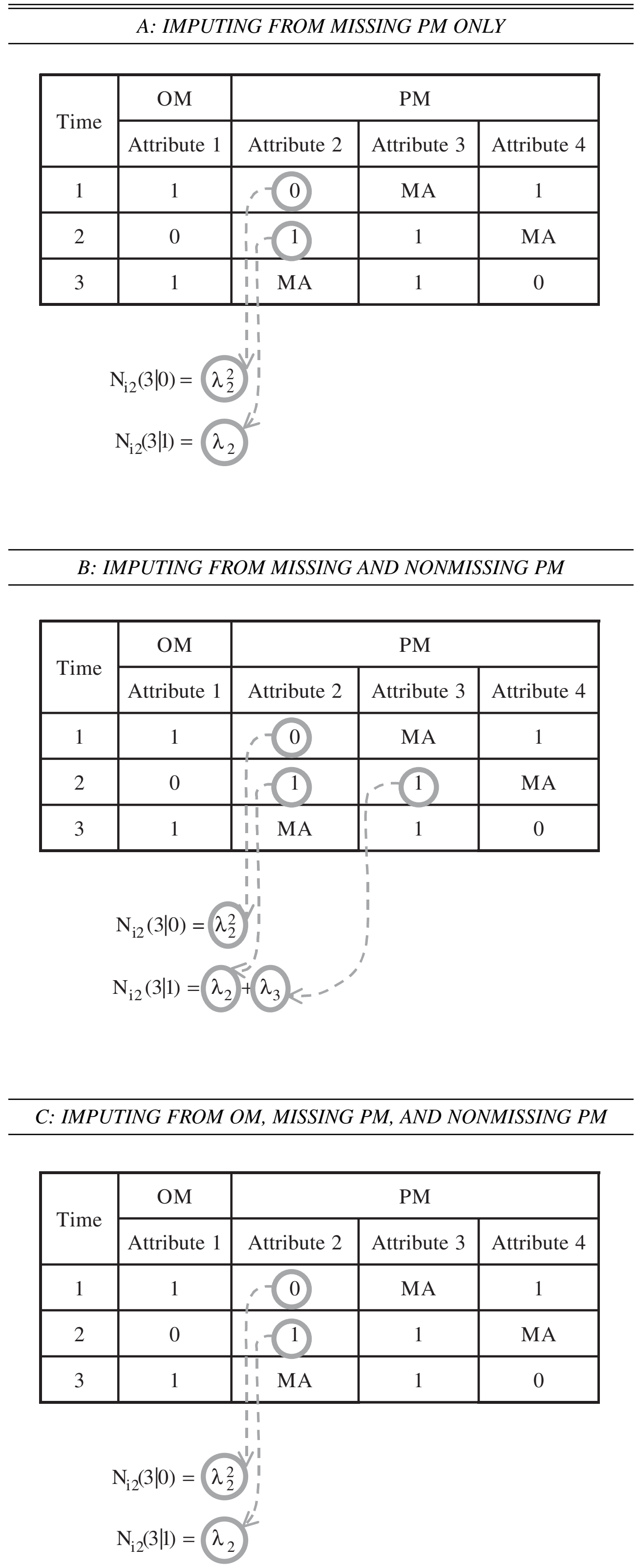
As do existing models, our imputation model derives probabilities that the missing PM Attribute 2 in profile $M_i(3)$ takes a value of 1 or 0. The parameterization of the probabilities is based on the work of Hoch, Bradlow, and Wansink (1999), who describe a similarity measure between a pair of categorical objects (conjoint profiles in this research), and Camerer and Ho (1998, 1999) and Ho and Chong (2003), who describe how learning and memory decay occur over time. The two basic concepts we use here are what we call "pattern matching" and "experience count" (EC). We combine them to yield our imputation model, which defines the probabilities over the missing attribute levels.

We develop three potential classes of models (Figure 1) to demonstrate how the generality of our model is built sequentially, using varying information sources. The model in Figure 1, Panel A, uses only previously shown information about the missing PM attribute (i.e., j = 2) to impute missing levels. It is a natural extension of the recency and averaging model that allows for decay. The imputation is based on historical levels of Attribute 2 (0 in profile $M_i(1)$ and 1 in profile $M_i(2)$), but the more recent level (1 in profile $M_i(2)$) may be more influential. To capture the recency effect in a decay-weighed averaging model, we introduce a decay parameter, λ_2 ($0 < \lambda_2 \leq 1$; the subscript denotes Attribute 2). We also introduce the EC concept (Camerer and Ho 1999), such that $N_{ij}(t|l)$ denotes respondent i's latent EC of attribute j at time t, taking level l.

[2]Note that if $M_i(2)$ were the first profile and $M_i(1)$ were the second profile, the prediction of the recency model would be reversed; that is, $\Pr[x'_{i2}(3) = 1 = 0]$. The averaging model, however, would give the same prediction.

Figure 1

THREE CLASSES OF MODELS



As illustrated in Figure 1, Panel A, $N_{i2}(3|0) = \lambda_2^2$ because $M_i(1)$ has an observed level of 0 for Attribute 2 at time 1 (i.e., $x_{i2}(1) = 0$). If $x'_{i2}(3)$ were to be imputed as 0, λ_2 would have a power of 2 because there are two periods of time difference between profile $M_i(1)$ and profile $M_i(3)$, during which they would match on PM Attribute 2. Similarly, $N_{i2}(3|1) = \lambda_2$ because $M_i(2)$ has $x_{i2}(2) = 1$ and would match with $x'_{i2}(3)$ if it were to be imputed as a 1. Therefore, the probability that $x'_{i2}(3)$ takes level 1 or 0 is as follows:

$$\Pr\left[x'_{i2}(3) = 1\right] = \frac{N_{i2}(3|1)}{N_{i2}(3|1) + N_{i2}(3|0)} = \frac{\lambda_2}{\lambda_2 + \lambda_2^2}.$$

$$\Pr\left[x'_{i2}(3) = 0\right] = 1 - \Pr\left[x'_{i2}(3) = 1\right].$$

The averaging model is a special case of this class of models in which $\lambda_2 = 1$. The recency model has $\lambda_2 \to 0$, with $\Pr[x'_{i2}(3) = 1] \to 1$ and $\Pr[x'_{i2}(3) = 0] \to 0$.

The more general model in Figure 1, Panel B, uses information from both the missing and nonmissing PM attributes (Attributes 3 and 4). That is, in addition to using information from Attribute 2 itself, we use possible conditional match patterns between profiles on the nonmissing PM attributes. Because we assume that the missing attribute levels are not used for imputation,[3] we only need to check whether there is a match between $[x_{i3}(3)]$ and $[x_{i3}(2)]$ and between $[x_{i4}(3)]$ and $[x_{i4}(1)]$. Because $[x_{i3}(3)]$ and $[x_{i3}(2)]$ match, we expect $M_i(2)$ to influence imputation more than $M_i(1)$ on $M_i(3)$. We add another decay parameter, λ_3 ($0 < \lambda_3 \leq 1$), to capture this reinforcement. Consequently, $N_{i2}(3|1)$ becomes $\lambda_2 + \lambda_3$, whereas $N_{i2}(3|0)$ stays the same. The probability that $x'_{i2}(3)$ is imputed as 1 or 0 now becomes:

$$\Pr\left[x'_{i2}(3) = 1\right] = \frac{N_{i2}(3|1)}{N_{i2}(3|1) + N_{i2}(3|0)} = \frac{\lambda_2 + \lambda_3}{(\lambda_2 + \lambda_3) + \lambda_2^2}.$$

$$\Pr\left[x'_{i2}(3) = 0\right] = 1 - \Pr\left[x'_{i2}(3) = 1\right].$$

Note that $\Pr[x'_{i2}(3) = 1]$ becomes larger, compared with Figure 1, Panel A, because $M_i(2)$, compared with $M_i(1)$, is more similar to $M_i(3)$ than when we consider only the missing PM attribute.

The model in Figure 1, Panel C, uses all of the available information from both the PM and OM attributes to impute the missing level. Following the same procedure, the ECs are $N_{i2}(3|0) = \lambda_1^2 + \lambda_2^2$ and $N_{i2}(3|1) = \lambda_2 + \lambda_3$. The corresponding probabilities become

$$\Pr\left[x'_{i2}(3) = 1\right] = \frac{N_{i2}(3|1)}{N_{i2}(3|1) + N_{i2}(3|0)}$$

$$= \frac{\lambda_2 + \lambda_3}{(\lambda_2 + \lambda_3) + (\lambda_1^2 + \lambda_2^2)}, \text{ and}$$

$$\Pr\left[x'_{i2}(3) = 0\right] = 1 - \Pr\left[x'_{i2}(3) = 1\right].$$

The most general model, Figure 1, Panel C, has two desirable properties: First, it uses all available information in the previously shown and current profiles in a sensible way. Furthermore, the model highlights the potential pitfalls of the averaging and recency models. For example, it implies that these simpler models would yield the same prediction if $M_i(3)$ were to take any of these patterns {[1, MA, 1, 1], [1, MA, 1, 0], [1, MA, 0, 1], [1, MA, 0, 0], [0, MA, 1, 1], [0, MA, 1, 0], [0, MA, 0, 1], [0, MA, 0, 0]}, which seems very unlikely. Second, it enables respondents to apply different weights to different attributes, depending on their preference.

General Formulation

In general, the pattern matching between two profiles can be formally defined as follows: Assume $r_{ij}(t) = 0$ and the level of attribute j at time t for respondent i is to be imputed. For each possible level of $x'_{ij}(t)$, we consider all previously shown profiles ($t' < t$) that have attribute j of the same value (i.e., $x_{ij}(t') = x'_{ij}(t)$). That is, we find all t′ in which the indicator function $I[x_{ij}(t'), x_{ij}(t)] = 1$. In addition, we set $I[x_{ij'}(t), x_{ij'}(t')]$ equal to 1 for those profiles $M_i(t')$ that have a match along a different attribute j′ with the current profile $M_i(t)$ but also on the missing PM attribute j. We call this a conditional match-up model, because the pairs of profiles must match on the missing PM attribute for it to be added to the EC.

It is also important to note the following properties of our pattern-matching approach: (1) We do not match profiles based on imputed values of previous attributes or, in the case in which more than one attribute is missing, imputed values of the missing PM attribute(s) $j' \neq j$. (2) The way we match a given pattern is binary (yes it matched/no it did not). Although a metric-based degree of matching model is possible, we chose a binary Hamming metric approach because it is parsimonious, easy to describe, and cognitively simple.

Let $N_{ij}(t|l_j)$ denote the latent EC of respondent i for attribute j to take level l_j at time t. With attribute j as the missing PM attribute, our model for $N_{ij}(t|l_j)$ in complete generality is given by

$$\text{(3)}\quad N_{ij}(t|l_j) = N_{ij}(0|l_j) + \sum_{t'=1}^{t-1}\Big(\lambda_{i(j,j)}^{t-t'} \times I\left[x'_{ij}(t), x_{ij}(t')\right] + \sum_{j'=1, j'\neq j}^{J}\Big\{\lambda_{i(j,j')}^{t-t'} \times I\left[x_{ij'}(t), x_{ij'}(t')\right] \times I\left[x'_{ij}(t), x_{ij}(t')\right]\Big\}\Big),$$

$$= \text{prior count} + \{\text{missing PM count} + [\text{nonmissing PM attribute} + \text{OM attribute count}]\},$$

where $N_{ij}(0|l_j)$ denotes the prior count of person i on attribute j, level l_j at time 0, and $0 < \lambda_{i(j,j')} \leq 1$ is the decay parameter for person i relating attribute j′ (j′ = 1, …, J) to j. In addition, $N_{ij}(0|l_j)$ allows for the possibility of prior knowledge of the marginal frequency of attribute levels and prior correlation between attribute levels.

In our experimental results, we fit a fairly general (reduced-form) version of the model (Equation 3) with the following set of specifications for λ_{ij}. This version corresponds to Table 3, an example with digital cameras in which j = {1, 2, 3, 4} are PM attributes (delay between shots, storage media, maximum resolution, and camera size) and j = {5, 6} are OM attributes (price and mini-movie).

[3]This is an assumption/limitation of our approach that we discuss subsequently as an area for further research.

Table 3

DIGITAL CAMERA ATTRIBUTES

Attribute	*Delay Between Shots*	*Storage Media*	*Maximum Resolution*	*Camera Size*	*Price*	*Mini-Movie*
Level 0	Four seconds	Floppy disk	800 × 600	SLR*	\$239	No
Level 1	Two seconds	Removable memory	1024 × 768	Medium	\$159	Yes

*SLR = single-lens reflex camera, larger than "medium."

$$\lambda_{i(j,j')} = \lambda_{ij} \quad \text{if } j' = j, j' \in \text{PM}; \forall j = 1,2,3,4;$$

$$\lambda_{i(j,j')} = \lambda_{i5} \quad \text{if } j' \neq j, j' \in \text{PM; and}$$

$$\lambda_{i(j,j')} = \lambda_{i6} \quad \text{if } j = \text{maximum resolution and } j' = \text{price}.$$

Note that λ_{i6} is included in the model, as we describe subsequently, because of a prior manipulation of the covariance between price and maximum resolution.

This structure defines the entire imputation process for partial-profile conjoint designs as a time-varying latent contingency table with counts, $N_{ij}(t|l_j)$, given in Equation 3. Thus, $\Pr[x'_{ij}(t) = l_j]$, $l_j = 1$ or 0, is given by

$$(4) \quad \Pr\left[x'_{ij}(t) = l_j\right] = \begin{cases} 1 & \text{if } r_{ij}(t) = 1 \text{ and } x'_{ij}(t) = l_j, \\ 0 & \text{if } r_{ij}(t) = 1 \text{ and } x'_{ij}(t) \neq l_j, \text{ and} \\ \dfrac{N_{ij}(t|l_j)}{N_{ij}(t|0) + N_{ij}(t|1)} & \text{if } r_{ij}(t) = 0. \end{cases}$$

That is, the probability that a given attribute level is imputed when the attribute is missing is its proportion of the total EC for that attribute. Because $N_{ij}(t|l_j)$ incorporates information across patterns to reinforce each pattern and allows for differing importance across time, this model satisfies our basic pattern-matching and reinforcement requirements. When multiple attributes are missing, we assume independence of counts to derive the joint probability of the missing pattern; however, the counts correlated as attribute levels that occur together have counts that will be updated together and prior counts that are related.

We denote the vector of imputed values at time t by a row vector $\mathbf{x}'_i(\mathbf{t}) = [x'_{i1}(t), x'_{i2}(t), \ldots, x'_{iJ}(t)]$. For a conjoint design with J attributes, in which each attribute has two levels, the imputed value vector $\mathbf{x}'_i(\mathbf{t})$ may assume one of the $K = 2^J$ possible potential profiles. We denote these potential profiles by Z_k (k = 1, …, K), and we determine the probability that $\mathbf{x}'_i(\mathbf{t})$ equals potential profile Z_k as follows:

$$(5) \quad \Pr\left[\mathbf{x}'_i(\mathbf{t}) = Z_k\right] = \prod_{j=1}^{J} \Pr\left[x'_{ij}(t) = l_j\right].$$

Heterogeneity

We allow the rate of information decay for a specific attribute pattern to be individual and attribute specific and recognize that considerable heterogeneity is likely to exist across people in their decay attribute imputation parameters, λ_{im} (m = 1, …, J). In addition, the basic parameters of the conjoint model, the individual conjoint intercepts α_i, the attribute partworths β_{ij} and β'_{ij}, and the residual variances may contain considerable heterogeneity yet share commonalities across the population of inference. To account for this heterogeneity in a coherent fashion, we nest our model in a Bayesian framework (Gelfand and Smith 1990). From Equation 2, we have the following:

$$\varepsilon_i(t) = y_i(t) - \left\{\alpha_i + \sum_{j=1}^{J}\left[\beta_{ij}x'_{ij}(t) + \beta'_{ij}r_{ij}(t)\right]\right\}.$$

We use an AR(1) (first-order autoregressive) model to capture the potential correlation of error terms over time (that is, people may anchor somewhat on the previously provided rating):

$$(6) \quad \varepsilon_i(t) = \gamma_i \varepsilon_i(t-1) + u_i(t),$$

where $u_i(t) \sim N(0, \sigma_i^2)$. In addition, we assume $\varepsilon_i(0) = 0$, $\forall$ i.

Prior and hyperprior specifications for the conjoint parameters ($\forall$ i, j) are given by[4]

$$\gamma_i \sim U(-1, 1),$$

$$\alpha_i \sim N(\bar{\alpha}, \sigma_\alpha^2),$$

$$\beta_{ij} \sim N(\bar{\beta}_j, \sigma_{j\beta}^2),$$

$$\beta'_{ij} \sim N(\bar{\beta}'_j, \sigma_{j\beta'}^2),$$

$$\sigma_\alpha^2, \sigma_{j\beta}^2, \sigma_{j\beta'}^2 \sim \text{Inv} - \Gamma(\cdot,\cdot),$$

and prior specification for the attribute decay parameters, $0 < \lambda_{im} \leq 1$, is given by

$$(7) \quad \lambda_{im} \sim \text{Beta}(a_m, b_m).$$

We assume that the prior ECs of each respondent follow a Poisson distribution with parameters varying by respondents and attribute levels:

$$N_{ij}(0|l_j) \sim \text{Poisson}\left[\exp(\zeta_i + \omega_j)\right],$$

with slightly informative priors on ζ_i and ω_j. We note that $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2; U(g, h) a uniform distribution with a lower bound g and an upper bound h; Inv – $\Gamma(\cdot,\cdot)$ an inverse gamma distribution with corresponding parameters; and

[4]We are aware that a more general setup would be to allow the βs to follow a multivariate normal distribution with nonzero, off-diagonal covariances. The current setup avoids overparameterization of the model. It is commonly used in economics literature (e.g., Berry, Levinsohn, and Pakes 1995).

Beta(a, b) a beta distribution with parameters a, b. To complete the model specification, we placed slightly informative hyperpriors on σ_α^2, $\sigma_{j\beta}^2$, and $\sigma_{j\beta'}^2$, $\forall$ j (inverse gamma distribution: Inv – Γ(.001, .001)); $\bar{\beta}_j$ and $\bar{\beta}'_j$, $\forall$ j (normal distribution: N(0, 1000)); and (a_m, b_m) (uniform distribution: U(0, 1000)). Sensitivity analyses indicate that the results were not affected by the exact choice of uninformative hyperprior values.

To summarize, let the imputation model parameters be denoted by row vectors $\boldsymbol{\lambda}_i = [\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{i6}]$ (the length of $\boldsymbol{\lambda}_i$ varies with different models as described in the experimental section) and the conjoint parameters by $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{i6}]$ and $\boldsymbol{\beta}'_i = [\beta'_{i1}, \beta'_{i2}, \ldots, \beta'_{i6}]$. Given $\Pr[x'_i(\mathbf{t}) = Z_k]$ from Equation 5, the likelihood function is as follows:

$$(8)\quad L(\alpha_i, \boldsymbol{\beta}_i, \boldsymbol{\beta}'_i, \boldsymbol{\lambda}_i, \gamma_i, \sigma_i^2, \mathbf{y}_i) = \prod_{t=1}^{20}\left\{\sum_{k=1}^{26}\left[\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{\{[\varepsilon_i(t)|Z_k] - \gamma_i\varepsilon_i(t-1)\}^2}{2\sigma_i^2}\right) \times \Pr[\mathbf{x}'_i(\mathbf{t}) = Z_k]\right]\right\}.$$

That is, we integrate the conjoint regression model with respect to the imputation model by sticking in the considered value for attribute j to person i for all possible profiles and weight them by their probability of being the imputed corresponding level. We use the notation $[\varepsilon_i(t)|Z_k]$ to emphasize that the value of $\varepsilon_i(t)$ is conditional on profile Z_k. Although there are 2^6 potential profiles in our study, at each time t, only two profiles have nonzero $\Pr[x'_i(\mathbf{t}) = Z_k]$ in the one-missing attribute case, and only four do in the two-missing attributes case.

We derived inferences from all models by obtaining posterior samples using a Markov chain Monte Carlo sampler. We performed all computation using the software package WinBUGS (Bayesian inference using Gibbs sampling; Spiegelhalter et al. 1996).[5] The results reported in the experimental section are the posterior means we obtained from aggregating the draws of three runs of the sampler from different starting points with a burn-in period of 6000 draws and a total run length of 10,000 draws. We assessed convergence using the F-test approach of Gelman and Rubin (1992).[6]

EXPERIMENT

We designed an experiment to provide a basic demonstration of our model for rating conjoint data with missing attributes. Our interest is in providing not only a demonstration of our approach but also a preliminary understanding of the following questions:

- Do people use missing attributes and their levels to evaluate products?
- If yes, do they infer missing attribute levels from all the information they learn about the product profiles, and do they reinforce patterns?

We assume that a consumer has minimal prior information about the product, though we estimate this as given in Equation 7. Therefore, we are able to impose a prior structure that varies across respondents in a systematic way. First, through a learning process, we create a prior for each respondent by controlling the products that he or she sees in a learning phase. Second, we ask participants to rate products with (or without, in the control group) missing attributes. Another control group, which worked on a self-explicated conjoint task, acts as a second baseline.

Stimulus

We selected digital cameras for this experiment because we wanted a relatively new product category for which the frequency of attribute levels and the correlation structure of the attributes are mostly unknown. Therefore, we could manipulate the frequencies and impose a prior as we desired. According to our demographic questions, less than 10% of our subjects owned digital cameras or claimed to have extensive prior expertise. We chose digital cameras with six attributes as the full-profile task because our research indicated that digital cameras could be described well using six features. A summary of the digital camera attributes we used are listed in Table 3. In our experimental condition, all attributes are simplified to have two realistic levels. This product setup provides a stylized empirical test of our model.

Experimental Groups

The experiment was designed to run on a university network. A total of 130 undergraduate students from a large East Coast U.S. university participated in the experiment. Subjects were obtained from six sections of a large class; these sections were assigned randomly, as follows: one section each to receive the self-explicated and full-profile (zero missing) cases and two sections each to receive the one-missing and two-missing-attributes cases. This assignment resulted in four groups of 17, 23, 47, and 43 respondents, respectively.[7] Group size differences were due to different section sizes and the participation rate of students in those sections. Across the conditions, less than 10% owned digital cameras, 40% were women, and 60% were men.

The experiment was composed of two phases: learning (prior) and rating. In the learning phase, we provided the subjects with text information about digital cameras and their attributes and then showed them 20 digital camera profiles listed in a single table. We controlled the subjects' priors by manipulating the digital camera profiles they saw in the learning phase. In the rating session, we asked the subjects to rate, on a 0–9 Likert scale, the attractiveness of different digital cameras (some with partial product profiles, depending on the treatment condition).

[5]To assess the ability of our most general model (Model 6) to recover the true underlying model structure, we ran a simulation study using synthetic data. The simulation results indicate that the model is able to recover the underlying conjoint regression coefficients (α_i, β_{ij}, β'_{ij}, and γ_i) very accurately and the imputation parameters (λs) with reasonable accuracy. Details are available from the authors on request.

[6]Because of the complexity of the model, the lack of familiarity that readers may have with the WinBUGS software program, and a desire for other researchers to apply our model easily, we have included an annotated version of the WinBUGS code for our most general model in an online appendix (see http://mktgweb.wharton.upenn.edu/ebradlow/research_files.htm).

[7]We note that a better design would have been to assign people randomly, not sections. In Study 2, we use random assignment.

Learning Phase

In the learning phase, we showed all subjects 20 digital camera profiles. The priors of the subjects before the rating phase were manipulated by the learning phase profiles. The purpose of this learning phase manipulation is twofold. First, if the relationship between, for example, price and maximum resolution (as we describe next) can be influenced by showing subjects profiles of a given structure, then managerial practice suggests that prior manipulations of this type could be valuable. Second, we wanted to test our model for a given attribute correlation. For example, as we manipulated the priors between digital camera price and maximum resolution, we wanted to test whether λ_{i6} affects the EC for resolution when it is missing. (Note that price is never missing because it is an OM attribute.)

We assigned each subject randomly to 1 of 11 prior coincidence structures that represented a different level of coincidence between price and maximum resolution (the coincidences between other attributes remained orthogonal). Specifically, subjects were assigned to read a table with a specific coincidence value (between 0 and 10) between price and maximum resolution. For example, a coincidence value of 10 indicates that among the 10 profiles that have low price ($159), all have low resolution (800 × 600); a value of zero would indicate that among the 10 profiles that have low price ($159), all of them have high resolution (1024 × 768). Such a coincidence structure could affect $\Pr[x'_{ij}(t) = l_j]$ if the learning phase carries over to the rating phase. That is, we empirically test our ability to manipulate the rating phase data by covarying price and maximum resolution at the 0, 1, …, 10 levels in the learning phase and then by estimating λ_{i6} in our model and finding its correlation with the subject's prior manipulation.

To ensure that subjects followed and attended to all information in the table of 20 learning profiles provided, we asked them to count five of the pairwise coincidences after they had read the tables. Among the five questions, one asked the subjects to count the coincidence between the price $159 and maximum resolution 800 × 600 (the manipulated coincidence), whereas the other four questions were randomly chosen to ask the subjects to count other coincidences. The sequences of these questions were randomized so as not to bias the results. Subjects' responses to these questions suggested that they had paid attention to the coincidence counts.

Rating Phase

In the rating phase, the design is orthogonal. In the one- or two-missing-attributes group, one or two of four PM attributes are removed from the designed conjoint cards, respectively. We fixed two attributes to be OM because we wanted to determine the impact of imputation of missing levels on observed attribute partworths. We used a Plackett–Burman design (Green, Carroll, and Carmone 1978; Plackett and Burman 1946) to create the profile cards. The sequences in which the profile cards were shown were generated randomly and varied across respondent. Each subject saw 24 profiles in the rating phase. Debriefing questions after the experiments provided evidence that the subjects noticed that attributes are missing and used them in their ratings. Response time was recorded, which could be used as a proxy for the difficulty of the task. Each response time was the time (in seconds) between subjects' keying in of successive rating score responses. An analysis of the response time data across missing attribute conditions (0, 1, and 2) indicates that respondents in the two-missing-attributes case spent considerably less time than did respondents in either of the other two cases ($p < .01$), corresponding to an average of 31 seconds less across the 24 profile rating tasks. No significant differences were found in response time between the zero- and one-missing-attribute conditions.

RESULTS

We used the first 20 profiles for each subject to calibrate the model and the last 4 as holdouts for validation. We estimated a total of six models with differing degrees of generality, grouped into four categories: (1) prior models; (2) imputation based on missing PM attributes only; (3) imputation based on missing and nonmissing PM attributes; and (4) imputation based on the OM attribute (price), missing, and nonmissing PM attributes. In Table 4, we show these models and their relationships for both the one- and two-missing-attributes cases. We estimated the models using the Bayesian hierarchical structure described in the "Heterogeneity" section.

In Table 5, we show the relative performance for the six models. We also show results for the three extant models (ignore, recency, and averaging), as well as one model in each of the three classes (Models 4, 5, and 6). For each model, we report the log of the marginal likelihood as computed by the log of the harmonic mean of the likelihood values (Congdon 2001, p. 475) and the mean absolute errors (MAE), both in-sample and out-of-sample. Using all three measures, our models (Models 4–6) perform better than prior models (Models 1–3). Specifically, Models 4–6 consistently perform better as more information is used for imputation.

Imputation Based on Missing PM Attributes

As we discussed previously, the extant models assume that subjects ignore the missing attribute, use the most recent occurrence of a missing attribute, or compute the average of all past occurrences. Of Models 1–3, the averaging model (Model 3) performs better in terms of log-marginal likelihood and in-sample and out-of-sample MAE. Model 4 relaxes the assumptions of Model 3; it allows a separate λ for each missing PM attribute, which decays geometrically.

Compared with the averaging model, Model 4 performs better in terms of the log-marginal likelihood, in-sample MAE, and out-of-sample MAE. Such results suggest that the relaxation of allowing for heterogeneous geometric decay helps in terms of model performance and better captures the actual rating process when missing attributes exist.

Imputation Based on Missing and Nonmissing PM Attributes

The preceding models assume that subjects impute a missing level of an attribute using only information within that PM attribute. A natural extension is to account for covariation from the nonmissing PM attributes. In Model 5, we assume that each attribute takes a different λ when it is present and when it is missing; however, the value of λ is assumed to be common across all PM attributes when they

Table 4
DESCRIPTION OF MODELS

			λ Values of PM Attributes					*λ Values of OM Attributes*	
			Missing PM						
Model Category	*Model*	*Number of Individual-Level Parameters*	*Delay*	*Storage*	*Resolution*	*Size*	*Nonmissing PM*	*Price*	*Mini-Movie*
Prior Models	1[§]	12	—				—	—	—
	2[†]	12	$\lambda_{i1} = \lambda_{I2} = \lambda_{i3} = \lambda_{i4} \rightarrow 0$				—	—	—
	3[‡]	12	1				—	—	—
Imputed based on missing PM	4	14	λ_{i1}	λ_{i2}	λ_{i3}	λ_{i4}	—	—	—
Imputed based on missing and nonmissing PM	5	17	λ_{i1}	λ_{i2}	λ_{i3}	λ_{i4}	λ_{i5}	—	—
Imputed based on OM and missing and nonmissing PM	6	18	λ_{i1}	λ_{i2}	λ_{i3}	λ_{i4}	λ_{i5}	λ_{i6}	—

[§]Ignore/missing model.
[†]Recency model.
[‡]Averaging model.

Table 5
PERFORMANCE OF DIFFERENT MODELS

	One Missing			*Two Missing*		
		MAE			*MAE*	
Model	*Log-Harmonic Mean of Likelihood*	*In-Sample*	*Out-of-Sample*	*Log-Harmonic Mean of Likelihood*	*In-Sample*	*Out-of-Sample*
1	–1285	.807	1.366	–1232	.861	1.382
2	–1245	.778	1.386	–1187	.823	1.390
3	–977	.733	1.314	–685	.500	1.381
4	–904	.694	1.310	–580	.461	1.360
5	–884	.602	1.292	–513	.459	1.341
6	–782	.580	1.281	–356	.434	1.300

are nonmissing. As we indicate in Table 5, Model 5 fits better than Models 1–4 in terms of log-marginal likelihoods, in-sample MAE, and out-of-sample MAE.

Imputation Based on OM Attribute, Missing, and Nonmissing PM Attributes

To test fully whether subjects use all information when inferring missing attribute levels, in addition to the last set of models, we add price (the OM attribute) to impute the missing level of maximum resolution. Recall that we manipulated the correlation structure between price and resolution in the learning phase. Model 6 extends Model 5 by allowing price to be used in the imputation process for missing maximum resolution levels. This relaxation improves the log harmonic mean likelihood, as well as the in-sample and out-of-sample MAE. The results suggest that subjects use OM attributes to infer missing attribute levels; whether this inference goes beyond price is an open question. Model 6 outperforms Models 1–3 significantly by all the measures we consider in Table 5. Specifically, Model 6 decreases the out-of-sample MAE by 4.1%, 5.3%, and 6.5% over Models 1–3, respectively, in the one-missing-attribute case and by 1.6%, 3.0%, and 5.9%, respectively, in the two-missing-attributes case.

We also performed a more detailed analysis at the individual level between the estimated effect λ_{i6} (price and maximum resolution) and the prior manipulated covariation between price and maximum resolution (0, 1, …, 10). First, we note that λ_{i6} is significantly different from zero (the [2.5%, 97.5%] percentile of its posterior is [.151, .315] in the one-missing-attribute case and [.628, .863] in the two-missing-attributes case), suggesting significant effects overall. Second, analysis at the individual level (without shrinkage) indicates a significant effect in the one-missing-attribute case (correlation = .345, p = .017) and insignificance in the two-missing-attributes case between the prior manipulation and λ_{i6}. Overall, these findings suggest that the subjects' priors can be manipulated to influence the way they infer missing attributes, but to what extent remains an open empirical question.

We provide the average and the standard deviations of the best-fitting model (Model 6) λ values in Table 6. The estimated λ values are different in the one- and two-missing-attributes cases, which is not surprising; when different numbers of attributes are missing, the weights that reflect how information from nonmissing attributes is used change accordingly.

Note that all the λ values are significantly larger than 0 and smaller than 1, thus indicating that the actual imputation procedure is different from the pure effect of Models 1–3.

Estimated Partworths and Priors

In Table 7, we report the mean and standard deviation of the partworths of Model 6, the best-fitting model. Elements β'_{ij} (j = 1, …, 4, ∀ i) are the partworths of the attributes

Table 6

λS OF THE BEST-FITTING MODEL (MODEL 6)

Model	*Values*	*Missing PM*				*Nonmissing PM*	*OM (Price)*
		Delay	*Storage*	*Resolution*	*Size*		
One missing	Average	.306	.095	.829	.240	.977	.247
	S.D.	.056	.066	.044	.067	.018	.046
Two missing	Average	.981	.429	.032	.070	.697	.743
	S.D.	.012	.091	.018	.061	.074	.070

Notes: S.D. = standard deviation.

when they take level 0 and are present (compared with taking level 0 and not being present, or being imputed as a zero); $\beta_{ij} + \beta'_{ij}$ (j = 1, …, 4, ∀ i) are the partworths of the attributes when they take level 1 and are present (compared with taking level 0 and not being present). To compare our results with traditional conjoint partworths, therefore, we note that with all attributes present and the "low"-level attributes coded as zero (as is standard), the partworths represent the difference in utility between the high- and low-attribute levels. To align with our case, the traditional partworths from our model are β_{ij} (j = 1, …, 4, ∀ i), that is, the effect of being high when shown less the effect of being low when shown.

As we mentioned previously, we get a bonus, in that we can assess the effect of imputed versus not imputed attribute levels in our conjoint design, in addition to level 1 (high) versus level 0 (low) effects. In our model, these are the partworths β'_{ij}. We find that all elements β'_{ij} (j = 1, …, 4, ∀ i) have a 95% posterior interval that does not contain 0, which means that when an attribute is present, it is given significantly greater weight. This finding is consistent with extant research (Johnson 1987; Levin et al. 1986; Louviere and Johnson 1990; Meyer 1981). Alternatively, we note that the combined tests (present or not, levels 1 or 0) for each of the attributes suggest that it is not the attribute level inferred but the presence of the attribute that influences the weight put on the attribute. We believe this is an interesting area for additional study.

In Table 8, we report the relative importance of (traditional) partworths, such as when a high level is shown versus when a low level is shown, from both Model 6 and the case when there are no missing attributes. Although we observe relatively high stability in the rankings (e.g., storage and size are always last, resolution is always most important, and the other three are relatively close in importance), there are changes in the magnitude of the relative importance of the partworths. The finding that partworths themselves are biased (compared with the full-profile condition) is consistent with extant research (Johnson 1987; Levin et al. 1986; Louviere and Johnson 1990). However, becasue we also find that the relative rankings stay fairly stable, there is prima facie evidence that similar rating processes are occurring. In the two-missing-attributes case, when fewer attribute levels are available for imputation, the more important attributes in the zero-missing-attributes case become less important, and the less important ones become more important. Thus, there is a regression effect in partworths when subjects evaluate partial profiles when less information is provided.

We note that a way to interpret the observed changes in partworths is that consumers construct rather than retrieve utilities. Because the set of all available information changes with successive profiles, the utilities can change, even for identical profiles, if they appear at different points in time. This view is not new and has been established by consumer researchers (Bettman and Zins 1977; Payne, Bettman, and Johnson 1992).

Finally, we report on the model results with regard to the carryover effect from ones rating's error, $\varepsilon_i(t-1)$, to another and from the priors $N_{ij}(0|l_j)$. The AR(1) carryover effect is

Table 7

AVERAGE AND STANDARD DEVIATION (S.D.) OF THE PARTWORTHS OF THE BEST-FITTING MODEL (MODEL 6)

Model	*Coefficients*		*Delay*	*Storage*	*Resolution*	*Size*	*Price*	*Mini-Movie*	*Intercept*
One missing	β_{ij}	Mean	1.328	.046	1.384	.092	.829	1.296	.592
		S.D.	.124	.182	.133	.144	.124	.158	(.426)
	β'_{ij}	Mean	.844	.745	.876	.737	—	—	
		S.D.	.117	.145	.124	.136	—	—	
Two missing	β_{ij}	Mean	1.001	.565	1.267	.469	1.000	1.085	.526
		S.D.	.004	.177	.088	.150	.004	.130	(.140)
	β'_{ij}	Mean	1.196	1.238	1.168	1.119	—	—	
		S.D.	.065	.106	.063	.063	—	—	

Table 8

COMPARISON OF RELATIVE IMPORTANCE OF PARTWORTHS

Model	*Delay*	*Storage*	*Resolution*	*Size*	*Price*	*Mini-Movie*
None missing	.126	.029	.317	.056	.198	.275
One missing, Model 6 (β_{ij})	.267	.009	.278	.018	.167	.261
Two missing, Model 6 (β_{ij})	.186	.105	.235	.087	.186	.201

statistically significant with a mean of $\gamma \approx .1$ (in both the one- and two-missing-attributes cases), which suggests that people anchor somewhat on previous values. This result also suggests that the order in which previous profiles are presented influences subjects' ratings of a current profile.

None of the estimated prior parameters ζ_i (i = 1, …, 47) or ω_j (j = 1, …, 6) is significantly different from zero according to the [2.5%, 97.5%] percentile of their posterior draws in the one-missing-attribute case, an indication that the subjects have weak priors for the product category. Consequently, as we show in Table 9, the average values of the prior ECs are typically small for Model 6. These initial ECs thus exert minor influence on the imputation of the early profiles but decay quickly when more profiles are shown. However, some of the prior parameters become significant in the two-missing-attributes case. Specifically, the ωs for resolution and size, the PM attributes people are probably most familiar with, are fairly significant. Thus, when less information becomes available, people may depend more on their priors to make judgments. This finding certainly requires further study beyond the empirical example provided here.

Robustness of Results

In our experiment, we impose a prior on subjects' beliefs about the relationship between price and maximum resolution in the learning phase and subsequently measure whether it exists in the calibration phase. Our process of having people count relationships between pairs of attributes (which would normally not be done in practice) may bias people toward imputing attribute levels when they are missing, due to priming.[8] To check whether our results are robust to this manipulation, we ran a second study, with zero- and one-missing-attribute cases only, that does not include a learning phase; in all other ways, it was identical to the first study. Our goal was to demonstrate the existence of imputation (as in the first experiment) and replicate the patterns of superiority of Models 4–6 over Models 1–3.

Specifically, 91 subjects from a large West Coast U.S. university, to partially fulfill requirements for a course, were obtained for our conjoint computer-based study of digital cameras with the same six attributes as in our first study. Subjects were randomly assigned to either the zero-missing-attributes case as a baseline (41 subjects) or the one-missing-attribute case (50 subjects). As in Study 1, the first 20 rating tasks were used to calibrate the model, the remaining 4 for out-of-sample validation. Profiles were presented in a random order within each design.

We present a detailed set of findings for this study in Tables 10, 11, and 12, but at a summary level, our findings are as follows: We find an identical pattern of overall fit, both in-sample and out-of-sample, to that for Study 1, in that the recency model has the worst fit, followed by the model that ignores the missing attributes and the averaging model, and then the three learning-based models. Other findings, such as the mean value of $\gamma = .167$ and the pattern of relative partworths for Model 6 (the best-fitting model), indicate that our findings are robust overall to the learning phase manipulation and are replicated.

CONCLUSION AND FURTHER RESEARCH

We develop a learning model to describe how consumers impute missing levels in partial conjoint profiles. In our model, consumers match patterns and develop inferences on the basis of their prior exposures. Our model extends aver-

[8]We thank an anonymous reviewer for suggesting this and the second study.

Table 9
AVERAGE OF ESTIMATED PRIORS $N_{ij}(0|\cdot)$

Models	*One Missing*		*Two Missing*	
Levels	*0*	*1*	*0*	*1*
Delay	1.025	.931	4.198	2.461
Storage	.194	.722	2.898	4.812
Resolution	.422	.539	10.068	16.816
Size	.453	.441	4.439	4.858
Price	.967	1.004	2.062	2.006
Mini-Movie	1.005	1.017	2.002	2.022

Table 10
PERFORMANCE OF DIFFERENT MODELS (NO LEARNING)

	One Missing		
	Log-Harmonic Mean of Likelihood	*MAE*	
Model		*In-Sample*	*Out-of-Sample*
1	–582	.905	1.182
2	–599	.928	1.214
3	–526	.866	1.182
4	–459	.820	1.177
5	–424	.808	1.167
6	–414	.802	1.159

Table 11
COMPARISON OF RELATIVE IMPORTANCE OF PARTWORTHS (NO LEARNING)

Model	*Delay*	*Storage*	*Resolution*	*Size*	*Price*	*Mini-Movie*
None missing	.144	.114	.308	.026	.190	.217
One missing, Model 6 (β_{ij})	.162	.114	.254	.028	.204	.237

Table 12
AVERAGE OF λS OF THE BEST-FITTING MODEL (MODEL 6, NO LEARNING)

		Missing PM					
Model	*Values*	*Delay*	*Storage*	*Resolution*	*Size*	*Nonmissing PM*	*OM (Price)*
One missing	Average	.624	.126	.555	.037	.065	.598
	Standard Deviation	.092	.037	.068	.013	.027	.128

aging and recency models and shows that consumers may infer missing attribute levels using both missing and nonmissing attribute information. We show that our best-fitting models outperform the prior models both in-sample and out-of-sample.

The ignore model is inadequate because consumers appear to consider missing attribute levels. Neither the averaging nor the recency model performs significantly better because consumers impute missing attribute levels using prior levels of nonmissing attributes (PM or OM). At the same time, significant correlation between the manipulated coincidence in the learning phase and the estimated decay parameter provides evidence that consumers' priors could be influenced by communication and experience. Consequently, managers may be able to influence the overall attractiveness of a product to a consumer by making consumers learn prior knowledge that favors the product.

This research has two caveats. First, the product used in our experiment has only six attributes, each with two levels. Thus, our study is best considered a demonstration of the potential of our imputation model for predicting preferences in more complicated product categories. Second, our rating-based conjoint experiment does not provide direct evidence of the applicability of our model to choice-based conjoint, though theoretically, such application is possible, as we described previously.

We foresee at least three research opportunities:

1. An interesting area to pursue is modeling the trade-off between the number of profiles and the number of attributes shown in each profile. From an econometrics perspective, it would be interesting to keep the total number of attribute levels shown fixed and determine how different combinations of the number of attributes and profiles lead to different levels of information content.
2. We assume here a pattern-matching model in which attributes either match or do not (0/1). A more general distance model can explicitly account for the relative differences between attribute levels. Such machinery is already in marketers' toolboxes; multiple dimensional scaling studies are used for such purposes. Thus, two promising areas for future studies would be to (a) combine conjoint analysis and multiple dimensional scaling studies to impute missing attribute levels and (b) create a latent perceptual mapping model for missing attribute levels in conjoint.
3. As we mentioned previously and as has been shown in prior research, missing attributes may change the relative importance of attributes. Although our work confirms this hypothesis, and for the most important attributes, whether this is true generally is unclear, and what may moderate this effect may be of interest. Thus, it would be interesting to conduct studies to determine the degree of this change and its moderating variables.

In conclusion, we believe that the general theoretical framework presented here, as well as its empirical validations, is a good first step that we hope will lead to a stream of managerially important research.

REFERENCES

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–90.

Bettman, J.R. and M. Zins (1977), "Constructive Processes in Consumer Choice," *Journal of Consumer Research*, 4 (2), 75–85.

Broniarczyk, S.M. and J.W. Alba (1994), "The Role of Consumers' Intuitions in Inference Making," *Journal of Consumer Research*, 21 (3), 393–407.

Camerer, Colin F. and T.-H. Ho (1998), "EWA Learning in Coordination Games: Probability Rules, Heterogeneity, and Time Variation," *Journal of Mathematical Psychology*, 42 (2-3), 305–326.

——— and ——— (1999), "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, 67 (4), 827–74.

Congdon, P. (2001), *Bayesian Statistical Modeling*. West Sussex, England: John Wiley & Sons.

Elrod, T., J.J. Louviere, and K.S. Davey (1992), "An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models," *Journal of Marketing Research*, 29 (August), 368–77.

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

Gelfand, A.E. and A.F.M. Smith (1990), "Sampling-Based Approaches to Computing Marginal Densities," *Journal of the American Statistical Association*, 85 (410), 398–409.

Gelman, A. and D.B. Rubin (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7 (4), 457–511.

Green, P.E. (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," *Journal of Consumer Research*, 1 (2), 61–68.

——— (1984), "Hybrid Model for Conjoint Analysis: An Expository Review," *Journal of Marketing Research*, 21 (May), 155–69.

———, J.D. Carroll, and F.J. Carmone (1978), "Some New Types of Fractional Factorial Designs for Marketing Experiments," in *Research in Marketing*, Vol. 1, J.N. Sheth, ed. Greenwich, CT: JAI Press, 99–122.

———, A.M. Krieger, and M.K. Agarwal (1991), "Adaptive Conjoint Analysis: Some Caveats and Suggestions," *Journal of Marketing Research*, 28 (May), 215–22.

Ho, T.-H. and J.-K. Chong (2003), "A Parsimonious Model of SKU Choice," *Journal of Marketing Research*, 40 (August), 351–65.

Hoch, S.J., E.T. Bradlow, and B. Wansink (1999), "The Variety of an Assortment," *Marketing Science*, 18 (4), 527–46.

Huber, J. and J.W. McCann (1982), "The Impact of Inferential Beliefs on Product Evaluations," *Journal of Marketing Research*, 19 (August), 324–33.

Johnson, R.D. (1987), "Making Judgments When Information is Missing: Inferences, Biases, and Framing Effects," *Acta Psychologica*, 66 (1), 69–82.

——— and I.P. Levin (1985), "More than Meets the Eye: The Effect of Missing Information on Purchase Evaluations," *Journal of Consumer Research*, 12 (2), 169–77.

Johnson, R.M. (1991), "Comment on 'Adaptive Conjoint Analysis: Some Caveats and Suggestions,'" *Journal of Marketing Research*, 28 (May), 223–25.

Levin, I.P., R.D. Johnson, P.J. Deldin, L.M. Carstens, L.J. Cressey, and C.R. Davis (1986), "Framing Effects in Decisions with Completely and Incompletely Described Alternatives," *Organizational Behavior and Human Decision Processes*, 38 (1), 48–65.

Little, R.J.A. and D.B. Rubin (1987), *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Louviere, J.J. and R.D. Johnson (1990), "Reliability and Validity of the Brand-Anchored Conjoint Approach to Measuring Retailer Images," *Journal of Retailing*, 66 (4), 359–81.

Lynch, J.G., Jr. and T.K. Srull (1982), "Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods," *Journal of Consumer Research*, 9 (1), 18–37.

Meyer, R. (1981), "A Model of Multiattribute Judgments Under Attribute Uncertainty and Informational Constraint," *Journal of Consumer Research*, 18 (4), 428–41.

Payne, J.W., J.R. Bettman, and E.J. Johnson (1992), "Behavioral Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology*, 43, 87–131.

Plackett, R.L. and J.P. Burman (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, 33 (4), 305–325.

Sawtooth Software (2002), ACA 5.0 Technical Paper, Sawtooth Software Technical Paper Series.

Spiegelhalter, D.J., A. Thomas, N.G. Best, and W.R. Gilks (1996), WinBUGS (Version 1.3). Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.

Srinivasan, V. and C.S. Park (1997), "Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement," *Journal of Marketing Research*, 34 (May), 286–91.

Ter Hofstede, F., Y. Kim, and M. Wedel (2002), "Bayesian Prediction in Hybrid Conjoint Analysis," *Journal of Marketing Research*, 39 (May), 253–61.

Wittink, D.R. and P. Cattin (1989), "Commercial Use of Conjoint Analysis: An Update," *Journal of Marketing*, 53 (July), 91–96.

Yamagishi, T. and C.T. Hill (1981), "Adding Versus Averaging Models Revisited: A Test of a Path-Analytic Integration Model," *Journal of Personality and Social Psychology*, 41 (1), 13–25.

This article has been cited by:

1. Rebecca Walker Naylor,, Cait Poynor Lamberton,, David A. Norton. 2011. Seeing Ourselves in Others: Reviewer Ambiguity, Egocentric Anchoring, and Persuasion. *Journal of Marketing Research* **48**:3, 617-631. [Abstract] [PDF] [PDF Plus]
2. Sören W. Scholz,, Martin Meissner,, Reinhold Decker. 2010. Measuring Consumer Preferences for Complex Products: A Compositional Approach Based on Paired Comparisons. *Journal of Marketing Research* **47**:4, 685-698. [Abstract] [PDF] [PDF Plus]
3. Young-Hoon Park,, Min Ding,, Vithala R. Rao. 2008. Eliciting Preference for Complex Products: A Web-Based Upgrading Method. *Journal of Marketing Research* **45**:5, 562-574. [Abstract] [PDF] [PDF Plus]
4. Min Ding. 2007. An Incentive-Aligned Mechanism for Conjoint Analysis. *Journal of Marketing Research* **44**:2, 214-223. [Abstract] [PDF] [PDF Plus]
5. Teck H. Ho, Noah Lim, Colin F. Camerer. 2006. Modeling the Psychology of Consumer and Firm Behavior with Behavioral Economics. *Journal of Marketing Research* **43**:3, 307-331. [Abstract] [PDF] [PDF Plus]
6. Vithala R. Rao. 2004. Comments on Conjoint Analysis with Partial Profiles. *Journal of Marketing Research* **41**:4, 388-389. [Citation] [PDF] [PDF Plus]
7. Joseph W. Alba,, Alan D. J. Cooke. 2004. When Absence Begets Inference in Conjoint Analysis. *Journal of Marketing Research* **41**:4, 382-387. [Citation] [PDF] [PDF Plus]
8. Donald B. Rubin. 2004. Design and Modeling in Conjoint Analysis with Partial Profiles. *Journal of Marketing Research* **41**:4, 390-391. [Citation] [PDF] [PDF Plus]
9. Eric T. Bradlow,, Ye Hu,, Teck-Hua Ho. 2004. Modeling Behavioral Regularities of Consumer Learning in Conjoint Analysis. *Journal of Marketing Research* **41**:4, 392-396. [Abstract] [PDF] [PDF Plus]