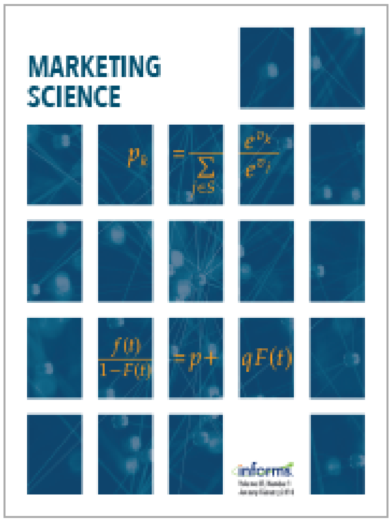


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Benefit-Based Conjoint Analysis

Dong Soo Kim, Roger A. Bailey, Nino Hardt, Greg M. Allenby

informs
http://pubsonline.informs.org/journal/mksc/

MARKETING SCIENCE
Vol. 36, No. 1, January–February 2017, pp. 54–69
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Benefit-Based Conjoint Analysis

Dong Soo Kim,[a] Roger A. Bailey,[a] Nino Hardt,[a] Greg M. Allenby[a]

[a] Fisher College of Business, Ohio State University, Columbus, Ohio 43210

Contact: kim.5042@osu.edu (DSK); bailey.1117@osu.edu (RAB); hardt.8@osu.edu (NH); allenby.1@osu.edu (GMA)

Received: November 19, 2014
Accepted: May 4, 2016
Published Online in Articles in Advance: October 4, 2016

https://doi.org/10.1287/mksc.2016.1003

Abstract. Firms develop products by manipulating the attributes of offerings, and consumers derive utility from the benefits that the attributes afford. While the field of marketing has long been aware of the distinction between attributes and benefits, it has not developed methods for understanding how attributes and benefits are related. This paper develops a benefit-based model for conjoint analysis that assumes consumers satiate on attributes that are perceived to provide the same benefit. A latent-variable model is proposed that estimates the map between attributes and benefits, and is applied to data from two conjoint studies involving a durable product and a household consumable. The model is shown to fit the data better, provide improved predictions, and lead to different product design implications than the standard conjoint model.

History: K. Sudhir served as the editor-in-chief and Eric Bradlow served as associate editor for this article.

Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2016.1003.

Keywords: economic separability • satiation • Bayesian estimation

1. Introduction

At the heart of any conjoint analysis is a model of how the values of products are created. In the standard model, product utility is modeled as an additively separable function of the attributes that describe each alternative. Although this linear model can be viewed as a robust approximation to any underlying model of value formation, directly relating the attributes that a firm can manipulate to the perceived value assigned by consumers is overly simplistic. The value of a product does not directly come from the product's attributes themselves, but through the perceived benefits that the product provides to consumers. It follows that models of value formation in conjoint analysis should recognize the relationship between attributes and benefits, and not assume that benefits are well described by simple functions of the attributes under study.

While at first glance this may seem to be a minor issue, note that in most cases the number of perceived benefits will be much smaller than the number of attributes and will depend on factors such as respondent experience, expertise, and intended use of the product. Some attributes will be perceived to not be important, while others will be perceived as providing the same benefit. When attributes are related in the sense of providing the same benefit, it is likely that their value decreases in the presence of the others. That is, consumers satiate on attributes that are perceived to be providing the same thing.

Consider, for example, a consumer who is trying to decide between toothpastes described by the attributes "whitening," "stain protection," "cavity protection," "tartar control," and "flavor." Suppose that this consumer perceives only two benefits from toothpaste: "cleaning power" and "flavor," wherein the first four attributes are all believed to provide the cleaning power benefit. While the consumer may value all four of the attributes that increase the cleaning power of the toothpaste, the marginal contribution of tartar control to the value of a toothpaste is likely smaller when all three of the other cleaning power attributes are present than if the tartar control attribute was the only cleaning power attribute present. In other words, the consumer satiates on the attributes that map to the cleaning power benefit. Although preferences may still be modeled as a function of attributes, the standard model using a function that is additively separable in attributes is insufficient. Compounding the problem is the fact that the firms providing toothpaste may not know a priori what attributes map to the same benefit or even the dimension of the benefit space in which preferences are being formed. A model constructed via a priori assumptions about the mapping from attributes to benefits is unlikely to yield improved results over the standard model.

This paper proposes a model for identifying the benefits provided by a product by partitioning its attributes into latent groups that provide additively separable utility. Preferences for products are assumed to be a function of its benefits, with consumers satiating on benefits rather than attributes. We propose a model structure that nests the standard attribute-based conjoint model, and illustrate its performance with two conjoint studies for common durable and household products where we find evidence that the attributes are being mapped to a much lower-dimensional benefit space. Our model provides better in-sample and predictive fit to the data, and describes a simplified benefit space that has implications for product design and brand positioning.

The organization of the paper is as follows. Section 2 develops a benefit-based model that nests the standard model, exhibits satiation through its properties of monotonicity and subadditivity, and possesses good mixing properties to facilitate statistical estimation. Section 3 presents empirical applications of the model. The results are discussed in Section 4, and concluding comments are provided in Section 5.

2. Model Development

Standard conjoint analysis assumes that consumer utility for an alternative is an additively separable function of the alternative's attributes (Green and Srinivasan 1978, 1990). Given the levels of N attributes of option j in a choice task, denoted $x_{j1}, x_{j2}, \ldots, x_{jN}$, its utility can be expressed as

$$u_j(x_{j1}, x_{j2}, \ldots, x_{jN}) = \sum_{n=1}^{N} w_n(x_{jn}) + \epsilon_j, \tag{1}$$

where w_n is the part-worth function for attribute n and ϵ_j is an alternative-specific error term. The part-worth functions, $w_1, w_2, \ldots, w_N$, are typically estimated using dummy variable coding. The size of the benefit, or part-worth, of attribute n for option j is given by $w_n(x_{jn}) = \mathbf{x}'_{jn}\boldsymbol{\beta}_n$, where $\mathbf{x}_{jn}$ is a vector denoting the level of attribute n for option j and $\boldsymbol{\beta}_n$ is a vector of the part-worths for these levels. The marginal utility for option j with this notation is given by

$$u_j = \sum_{n=1}^{N} w_{jn} + \epsilon_j = \sum_{n=1}^{N} \mathbf{x}'_{jn}\boldsymbol{\beta}_n + \epsilon_j. \tag{2}$$

The standard model in Equation (2) is quite flexible in modeling the value of each attribute individually, as it avoids the requirement of an a priori specification of the part-worth functions over the levels (for details, see Green and Srinivasan 1978, pp. 105–106).

Additive separability of the utility function implies that each attribute is assumed to provide its own benefit independently. This becomes problematic when consumers cannot distinguish among some of the attributes under study, either because of a lack of experience with the product or because they lack an understanding of how the attributes are different. A consumer purchasing a digital camera, for example, may not understand how the camera's resolution, sensor size, aperture, focal length, and image stabilization attributes work to produce clear pictures, and may perceive that all are elements of a common dimension, or benefit, called "picture clarity." If different attributes are perceived to be providing the benefits on the same dimension, the attributes possibly *interact* with each other for benefit formation and it would be required to relax additive separability in conjoint analysis. While the linear approximation is often viewed as adequate for prediction purposes (Johnson and Meyer 1984), it could be not enough to exactly understand consumers' choice behaviors and precisely develop firms' managerial strategies.

A challenge in modeling benefits is to identify the underlying structure of the benefit formation, which is usually unobservable in practice. While benefits have been studied in the segmentation (Haley 1968, Calantone and Sawyer 1978), market structure (Day et al. 1979, Wedel and Steenkamp 1991, Lamberton and Diehl 2013), and conjoint literatures (Wind 1973, Myers 1976), their relationship to product attributes has always been assumed to be a priori known, or known up to including interaction terms that improve model fit. The model that we propose in this section uses a standard conjoint experimental design with no additional, or prior, knowledge required to identify the benefit formation. This could be a distinctive advantage in practice because other general grouping methods, such as hierarchical information integration (HII) (Louviere 1984, Oppewal et al. 1994) and confirmatory factor analysis (CFA) (Ansari et al. 2002, Batra et al. 2010), would also require either a complex experimental design or additional attribute-level data.

Some supply-side optimization methods employ the standard additive conjoint model as a demand-side input in the optimization, assuming that producers and consumers engage in decision making at the same level of resolution. For example, analytical target cascading (Michalek et al. 2005) integrates constrained engineering design models with the standard attribute-based model to optimize new product development decisions. We note that these methods can be extended with the use of our benefit-based model. Using our model in place of the standard conjoint model would allow the engineering model to be connected with the underlying benefits consumers use to make choices. As we demonstrate below, our model embeds the standard conjoint model and therefore could improve the performance of analytical target cascading and other supply-side optimization methods.

2.1. A Benefit-Based Conjoint Model

We introduce a benefit-based conjoint model, a new model describing the choices based on the perceived benefits that the product provides to consumers. We start by relieving the additive separability assumption in the standard conjoint model. If different attributes are related to each other—i.e., perceived to provide the same benefit—their part-worths may not be additively separable. This implies that if the part-worths of different attributes are additively separable, those attributes may be independent of each other. The benefit formation then can be interpreted as a partitioning behavior: Consumers partition the N attributes into K mutually exclusive benefits, denoted $B_1, B_2, \ldots, B_K$, with $K \leq N$. The utility of option j is modeled as the sum of the part-worths of the benefits

$$u_j = \sum_{k=1}^{K} b_k(a_{jk1}, a_{jk2}, \ldots, a_{jkM_k}) + \epsilon_j, \tag{3}$$

where b_k is the part-worth function for benefit k, a_{jkm} is the value of the mth attribute of alternative j in benefit k, and M_k is the number of attributes that belong to benefit k. Utility in Equation (3) is an additively separable function of the benefit part-worths, implying that the benefits are exchangeable in generating value to the consumer, not the attributes, and that attributes in one benefit do not interact with attributes in other benefits.

Adding interaction terms could be a way to define the part-worth function, b_k, but it leads to a curse of dimensionality for implementation: The number of interaction terms increases as the number of attributes (N) increases. Even if there are only 10 binary attributes, a total of 1,013 ($= 2^{10} - 1 - 10$) interaction terms are possible and this number is not empirically feasible, compared to the number of observations per individual in a usual marketing data set. We therefore model how consumers form benefits from various attributes rather than empirically capture it through interaction terms.

Our model specification of the part-worth function (b_k) has three properties. First, it retains much of the simplicity of the standard dummy-variable conjoint model. This property allows our framework to nest the standard conjoint analysis with no extra data gathering effort. Second, it allows for satiation in attributes that map to the same benefit. Multiple attributes providing the same benefit are expected to have diminishing marginal returns, and it is this property that allows us to identify the presence of benefits. Diminishing marginal returns imply that the attributes are not universally exchangeable in generating utility, but depend on the presence of other attributes within the benefit. Third, it provides good statistical properties for estimating the assignment of attributes to benefits.

A form of b_k satisfying these properties is

$$b_k(a_{jk1}, a_{jk2}, \ldots, a_{jkM_k}) = g\left(\sum_{m=1}^{M_k} g^{-1}(a_{jkm})\right), \tag{4}$$

where

$$g(a) = \operatorname{sgn}(a)\log(|a|+1) = \begin{cases} \log(a+1), & \text{for } a \geq 0, \\ -\log(-a+1), & \text{for } a < 0, \end{cases} \tag{5}$$

and $a_{jkm} = \mathbf{x}'_{jkm}\boldsymbol{\beta}_{km}$ as in the standard conjoint model. This is the benefit function that we actually used in empirical applications. Justification for this form and its resulting implications are discussed below.

2.2. Properties of the Benefit Function b_k

2.2.1. Nested Structure. The first property of the benefit function is the ability to nest the standard conjoint analysis specified with dummy variable coding. We assume that $a_{jkm} = a_{km}(\mathbf{x}_{jkm}) = 0$ if $\mathbf{x}_{jkm} = \mathbf{0}$; i.e., when attribute m in benefit k is specified at its null level. If the values of all attributes in the benefit are equal to their null level—i.e., $a_{jk1}, \ldots, a_{jkM_k} = 0$—the part-worth for the benefit is assumed to be equal to zero. This assumption is similar to the normalizing assumption used with dummy-variable coding in conjoint models and is consistent with the notion that the statistical estimation of a utility function requires its location to be fixed. If an attribute is horizontally defined—i.e., its levels are not ordinal but categorical, such as colors—expanding it to multiple binary variables and handling the different levels as different attributes can still be used as in the standard conjoint model.

We note that this property does not imply that the benefit function is unable to handle continuous variables. It is needed only to assume that $a_{jkm} = a_{km}(x_{jkm}) = 0$ if x_{jkm} is equal to its null level, which is usually zero. For example, if price is continuous in analysis, the value of the price attribute, $a_{k,\text{price}}(p_j)$, is assumed to be zero at the zero-price as in the standard conjoint model. We employ a linear function, $\beta_{\text{price}}\, p_j$, for the continuous price variable in our empirical applications.

2.2.2. Satiation. The second property of the benefit function is satiation, requiring the presence of monotonicity and subadditivity of the benefit function. Monotonicity implies that additional units of an attribute lead to additional utility. Let $\mathbf{a}_j$ and $\mathbf{a}_{j'}$ denote the values of the attributes for option j and j', respectively. The benefit function is monotonic if

$$b(\mathbf{a}_j) \geq b(\mathbf{a}_{j'}), \quad \text{if } \mathbf{a}_j \geq \mathbf{a}_{j'}. \tag{6}$$

We note that monotonicity holds regardless of the sign of $\mathbf{a}$. People can become worse off with increasing levels of an attribute, resulting in a negative sign for the function $b(\cdot)$. Monotonicity allows us to define an ordinal structure of the part-worth for a benefit. Let $\mathbf{a}_S$ denote the vector whose elements are set equal to those of $\mathbf{a}$ for attribute $m \in S$ and 0 for $m \notin S$ for any $S \subseteq B$. Let B^+ be the set of attributes that have positive value and B^-

be the set of attributes that have negative value; i.e., $B^+ = \{m \mid a_m \geq 0\}$ and $B^- = \{m \mid a_m < 0\}$. If b is monotonic, then

$$b(\mathbf{a}_{B-S^+}) \leq b(\mathbf{a}) \leq b(\mathbf{a}_{B-S^-}), \quad \text{for all } S^+ \subseteq B^+ \text{ and } S^- \subseteq B^-. \tag{7}$$

This inequality consists of two parts. The first part, $b(\mathbf{a}_{B-S^+}) \leq b(\mathbf{a})$, implies that the part-worth function is increasing if any set of the attributes that have a positive value (S^+) is additionally included. By contrast, the second part, $b(\mathbf{a}) \leq b(\mathbf{a}_{B-S^-})$, implies that the part-worth function is decreasing if any set of attributes that have a negative value (S^-) is additionally included. The ordinal structure of the part-worth bounds the benefit function. In Equation (7), if $S^+ = B^+$ and $S^- = B^-$, then $B - S^+ = B^-$, $B - S^- = B^+$, and

$$b(\mathbf{a}_{B^-}) \leq b(\mathbf{a}) \leq b(\mathbf{a}_{B^+}). \tag{8}$$

Furthermore, since $\mathbf{a}_{B^-} \leq \mathbf{a}_{S^-} \leq \mathbf{0} \leq \mathbf{a}_{S^+} \leq \mathbf{a}_{B^+}$ and $b(\mathbf{0}) = 0$, we have

$$b(\mathbf{a}_{B^-}) \leq b(\mathbf{a}_{S^-}) \leq 0 \leq b(\mathbf{a}_{S^+}) \leq b(\mathbf{a}_{B^+}), \quad \text{for all } S^+ \subseteq B^+ \text{ and } S^- \subseteq B^-. \tag{9}$$

Satiation of the benefit function deals with the size of the increment or decrement in the part-worths. Since people satiate if they are repeatedly exposed to similar stimuli, the absolute change in the part-worth of a benefit is decreasing as more positive attributes are included. This implies that the part-worth function is subadditive: Consider partitions of B^+ and B^-, $\{S_1^+, \ldots, S_{L^+}^+\}$ and $\{S_1^-, \ldots, S_{L^-}^-\}$ such that $S_l^+, S_l^- \neq \varnothing$, $S_l^+ \cap S_{l'}^+ = \varnothing$ and $S_l^- \cap S_{l'}^- = \varnothing$ for $l \neq l'$, $\bigcup_{l=1}^{L^+} S_l^+ = B^+$, and $\bigcup_{l=1}^{L^-} S_l^- = B^-$. Then, b is subadditive[1] if

$$|b(\mathbf{a}_{B^+})| \leq \sum_{l=1}^{L^+} |b(\mathbf{a}_{S_l^+})| \quad \text{and} \quad |b(\mathbf{a}_{B^-})| \leq \sum_{l=1}^{L^-} |b(\mathbf{a}_{S_l^-})|, \tag{10}$$

for all possible partitions of B^+ and B^-. We note that subadditivity describes only the cases that the values of all attributes are positive (B^+) or negative (B^-). For the entire set, B, the following property is derived from Equations (8)–(10):

$$\sum_{l=1}^{L^-} b(\mathbf{a}_{S_l^-}) \leq b(\mathbf{a}_{B^-}) \leq b(\mathbf{a}) \leq b(\mathbf{a}_{B^+}) \leq \sum_{l=1}^{L^+} b(\mathbf{a}_{S_l^+}). \tag{11}$$

Returning to the toothpaste example, suppose that "stain protection," "cavity protection," and "tartar control" are grouped into the benefit "dental health." Because they positively affect "dental health," they may be positively evaluated. Monotonicity within the "dental health" benefit means that

$$0 \leq b(\text{stain}) \leq b(\text{stain, cavity}) \leq b(\text{stain, cavity, tartar}),$$

and subadditivity implies that

$$\begin{aligned} b(\text{stain, cavity, tartar}) &\leq b(\text{stain, cavity}) + b(\text{tartar}) \\ &\leq b(\text{stain}) + b(\text{cavity}) + b(\text{tartar}). \end{aligned}$$

Now, suppose that "whitening" is added into the benefit. Some people might feel that "whitening" is harmful to "dental health." Monotonicity for those who think of "whitening" negatively implies that

$$\begin{aligned} b(\text{whitening}) &\leq 0 \quad \text{and} \\ b(\text{stain, whitening}) &\leq b(\text{stain, cavity, whitening}) \\ &\leq b(\text{stain, cavity}). \end{aligned}$$

To verify that a potential part-worth function demonstrates monotonicity and subadditivity, we provide one more useful property. If b can be written as a single-valued function, denoted as v, of the sum of $a_1, a_2, \ldots, a_M$; i.e.,

$$b(a_1, a_2, \ldots, a_M) = v(y(\mathbf{a})), \quad y(\mathbf{a}) = \sum_{m=1}^{M} a_m, \tag{12}$$

and v satisfies the following conditions, b is monotonic and subadditive:

(i) v is a continuous and second-order differentiable real-valued function.[2]

(ii) v is monotonically increasing; i.e., $\partial v/\partial y \geq 0$.

(iii) v is fixed to zero at zero—i.e., $v(0) = 0$—implying that if the net value of the attributes in a benefit is zero, consumers do not have any gains and losses from the benefit.

(iv) v is concave for $y \geq 0$ and convex for $y < 0$; i.e., $\partial^2 v/\partial y^2 \leq 0$ for $y \geq 0$ and $\partial^2 v/\partial y^2 \geq 0$ for $y < 0$.

The proof is provided in the online appendix.

2.2.3. Statistical Mixing. One simple way to find an appropriate function for v in Equation (12) is to modify a monotonically increasing and concave function on the positive axis. Let $f(y)$ be a function that is continuous, second-order differentiable real-valued, monotonically increasing, fixed to zero at $y = 0$, and concave for $y \geq 0$. Then, we can build the following function that combines f and its reflection about the origin:

$$g(y) = \operatorname{sgn}(y) f(|y|) = \begin{cases} f(y), & \text{for } y \geq 0, \\ -f(-y), & \text{for } y < 0, \end{cases} \tag{13}$$

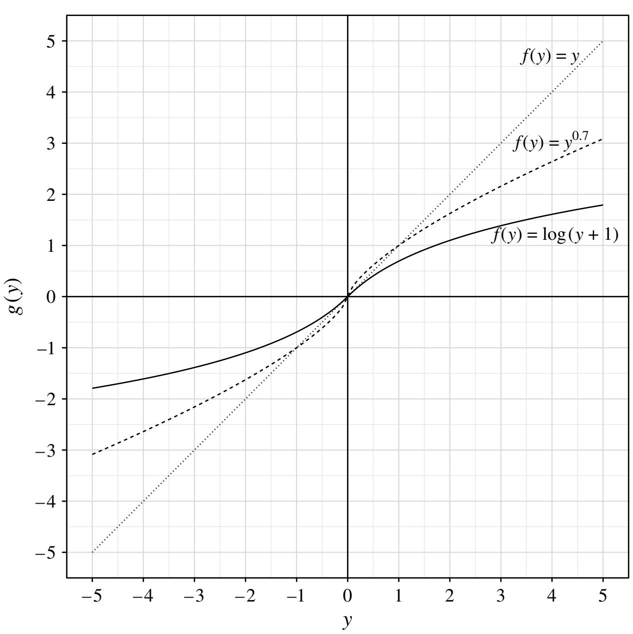
where $\operatorname{sgn}(y)$ is the sign function that equals 1 if $y > 0$, 0 if $y = 0$, and -1 if $y < 0$. Since f is a continuous and second-order differentiable real-valued function and $f(0) = -f(-0) = 0$, g is also a continuous and second-order differentiable real-valued function and $g(0) = 0$. Since $(\partial f/\partial y)(y) \geq 0$ for $y \geq 0$, we obtain that $(\partial g/\partial y)(y) = (\partial f/\partial y)(-y) \geq 0$ for $y < 0$. Therefore, g is monotonically increasing. Since $(\partial^2 f/\partial y^2)(y) \le 0$ for $y \ge 0$, we obtain that $(\partial^2 g/\partial y^2)(y) = -(\partial^2 f/\partial y^2)(-y) \ge 0$ for $y < 0$. Therefore, g is concave for $y \ge 0$ and convex for $y < 0$; i.e., g satisfies the conditions for v.

Two functions have been frequently employed as the concave function, f, in the literature. The first one is a power function (e.g., Dixit and Stiglitz 1977, Kim et al. 2002, Bhat 2008): y^δ, where δ is bounded in $(0, 1]$ to make the function concave. The second one is a logarithmic function (e.g., Satomura et al. 2011, Lee et al. 2013): $\log(\gamma y + 1)$, where "+1" is necessary to fix the function value to zero at $y = 0$ and γ is constrained to be positive. Additionally, a simple linear function αy is also able to be f if $\alpha \ge 0$. Figure 1 shows an example for each functional form. All functions are concave on the positive axis but convex on the negative axis.

We introduce a latent stochastic variable, $\mathbf{B}$, with a discrete distribution that indicates the assignment of attributes to latent benefits. Let $\mathbf{B} = \{B_1, \ldots, B_K\}$ be a partition of A, the set of all attributes. The overall utility, u, then depends on $\mathbf{B}$ as well as the values of the attributes, $a_1, a_2, \ldots, a_N$. The statistical estimation of the attribute–benefit partition requires a functional form for the benefits that allows for smooth transitions as different partitions $\mathbf{B}$ are investigated. As an attribute's assignment is changed from one benefit to another during estimation, it is desirable that its value in the utility function remains relatively consistent.

Consider, for example, the benefit functions with $f(y) = y$, $\log(y+1)$, and $y^{0.7}$ displayed in Figure 1, where it is seen that different values of y may be needed to produce the same benefit, $g(y)$. All statistical estimation algorithms require some regularity

Figure 1. The Shape of $g(y)$ for $f(y) = y$, $\log(y+1)$, and $y^{0.7}$



in the likelihood surface so that alternative values of model parameters can be explored. In our model, this includes the attribute assignments to benefits as indicated by $\mathbf{B}$. A challenge in estimating our model is to allow for smooth transitions in attribute assignment from, say, a linear model with no satiation to a logarithmic model with satiation, so that the level of the attribute has a similar effect on utility.

Our solution to the problem is to introduce an adjustment function for the values of the attributes. Let $q_k(a)$ be a function that adjusts the scale of a for benefit k and rewrite Equation (12) as

$$\begin{aligned} b_k(a_{jk1}, a_{jk2}, \ldots, a_{jkM_k}) &= g_k(y_k(\mathbf{a}_{jk})), \\ y_k(\mathbf{a}_{jk}) &= \sum_{m=1}^{M_k} q_k(a_{jkm}), \end{aligned} \tag{14}$$

for option j. Since the scale of the part-worth is the same across the benefits, the output value of b_k must be the same with the same single input value a; i.e.,

$$g_k(q_k(a)) = g_{k'}(q_{k'}(a)), \quad \text{for all } k, k' = 1, 2, \ldots, K. \tag{15}$$

This equation implies that if only one attribute is assigned into a benefit, there are no interactions within the benefit and, therefore, the part-worth of the benefit must be the same regardless of the functional form for the benefit. To determine the form of $q_k(a)$, we set a hypothetical benefit, denoted as $k = 0$, that has a linear part-worth function, $g_0(y) = y$, and assume that the adjustment function for this benefit is also a linear function, $q_0(a) = a$. The part-worth structure of this hypothetical benefit is equivalent to the linear part-worth structure in the standard conjoint model. Since Equation (15) has to also hold for this hypothetical benefit, $a = g_0(q_0(a)) = g_k(q_k(a))$, for all $k = 1, 2, \ldots, K$. Therefore, the appropriate adjustment function for benefit k becomes the inverse function of g_k as follows:

$$q_k(a) = g_k^{-1}(a). \tag{16}$$

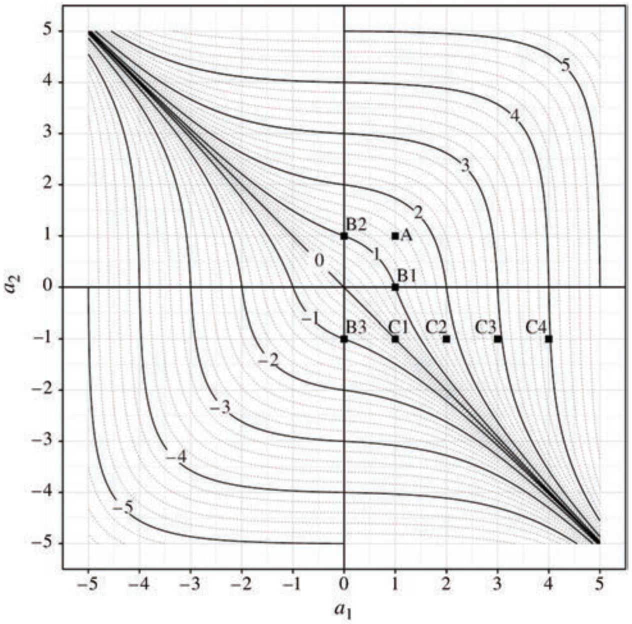
Substituting Equation (16) into Equation (14), we have

$$b_k(a_{jk1}, a_{jk2}, \ldots, a_{jkM_k}) = g_k\left(\sum_{m=1}^{M_k} g_k^{-1}(a_{jkm})\right). \tag{17}$$

Equation (17) clearly distinguishes between the roles of the value of an attribute, a_{jkm}, and the part-worth function for a benefit, b_k. If only one attribute exists in a benefit, the part-worth function becomes the value of the attribute itself, implying that the scale of the value of an attribute is equivalent to that of the part-worth. Therefore, a_{jkm} still captures the part-worth of an attribute as in the standard conjoint model. However, b_k describes how the part-worths of the attributes are combined within a benefit.

Figure 2 shows the contour plot of g in Equation (13) for $f(y) = \log(y+1)$ with two attributes. The function

Figure 2. Contour Plot of $g(g^{-1}(a_1)+g^{-1}(a_2))$ for $f(y)=\log(y+1)$



is convex to the origin (i.e., subadditive) in the first and third quadrants where the value of the attributes are either both positive or negative. If there is only one attribute for a benefit (points B1, B2, and B3), the part-worth of the benefit is equivalent to the value of the attribute; i.e., the scale adjustment allows us to still interpret a_{jkm} as the part-worth of the attribute. Consider "stain protection" (denoted a_1) and "whitening" (denoted a_2) for "dental health" in the toothpaste example. If a consumer thinks that both attributes are good for "dental health" as in point A, the part-worth of the composition of both attributes is less than the sum of the values of both attributes since the consumer satiates within "dental health." However, if only one of the attributes exists in a toothpaste (points B1 and B2), the consumer has the part-worth of the single attribute itself even though this consumer thinks of "whitening" negatively (point B3).

We do not need to make assumptions for the composition of positive and negative attributes: subadditivity is defined for the attributes that have the same sign in Equation (10), and the part-worth function in Equation (17) combines their values for a net score. Figure 2 graphically illustrates this property where the part-worth function is comprised of differently signed attributes in the second and fourth quadrants as compared to similarly signed attributes in the first and third quadrants. The 45° downward sloping line in the graph are points for which the attribute values are the same magnitude and the benefit has zero net value. Positive values are obtained above the line and negative values are below the line. It can be seen that the algebraic sign of the benefit is the same as that of the attribute with greater value. That is, if a consumer thinks that "stain protection" is good for "dental health" but "whitening" is harmful, and that both have equal value (point C1), the sum of the values is zero and the part-worth of "dental health" is zero. However, if the consumer thinks that "whitening" is harmful but less so than "stain protection" (point C2), the sum of the values of both attributes and the part-worth are greater than zero.

Subadditivity of the value function implies that as the importance of one attribute strictly increases, the impact of other attributes decreases. For example, as the importance of "stain protection" increases from point C1 to C4, the part-worth of "dental health" becomes closer to the value of "stain protection": the function value is zero at C1, 1.74 at C2, 2.91 at C3, and 3.97 at C4. If there are more than one positive and one negative attributes, subadditivity implies that one dominant positive (negative) attribute more strongly affects the part-worth than several minor negative (positive) attributes: We need only one attribute whose value equals −1 to make the part-worth be 0, if the value of an attribute in a benefit is 1. However, if the value of the positive attribute is 2 we need more than three attributes whose values are equal to −1, and if the value of the positive attribute is 3 we need more than 11 attributes whose values are equal to −1, to make the part-worth be 0.[3] This property could be useful to take into account "must have" or "must avoid" attributes. If there is a salient attribute, our model implies that people will tend to focus on the salient attribute and brush aside the other attributes. For example, if a consumer really hates "whitening" in terms of "dental health," the consumer may never buy a "whitening" toothpaste even though it is the best for "stain protection," "cavity protection," and "tartar control."

Substituting Equation (17) into Equation (3), we have the final form of the utility function. For convenience of implementation, we introduce auxiliary variables that describe the attribute–benefit mapping. Since N attributes are partitioned into K benefits, the attribute–benefit mapping could be described by N multinomial variables: $\tau_n^* \in \{1,2,\ldots,K\}$ for $n=1,2,\ldots,N$, meaning that attribute n is assigned into benefit k if $\tau_n^*=k$. For exposition purposes, we also define $N\times K$ binary variables, $\{\tau_{nk}\}$, which equal 1 if $\tau_n^*=k$ and 0 otherwise. The model can then be written as

$$u_j=\sum_{k=1}^{K} g_k\left(\sum_{n=1}^{N}\tau_{nk}\cdot g_k^{-1}(\mathbf{x}'_{jn}\boldsymbol{\beta}_n)\right)+\epsilon_j. \tag{18}$$

We note that K defines the maximum number of benefits in empirical analysis. When we fit the model to the data, both the number of benefits taken into account for product evaluation (K) and the attribute–benefit mapping ($\{\tau_{nk}\}$) are the unknown that we estimate.

Equation (18) does not have any restrictions on $\{\tau_{nk}\}$, so it allows for navigating all of the possible partitions including a case where there exist benefits into which no attributes are assigned. For example, it is possible that consumers utilize only one or two benefits even though $K = 3$. This leads to an additional property for the nested structure of the proposed model: the model with large K always nests the model with small K. We will discuss this with an empirical result in Section 3.

2.3. Null Group of Unique Attributes

Benefits do not necessarily have two or more attributes. If an attribute is perceived as unique, it may stand alone and make up a benefit by itself. In this case, since $g_k(g_k^{-1}(a)) = a$, the value of the unique attribute becomes the part-worth and, therefore, the contribution of the unique attribute can be described by a linear combination as in the standard conjoint model. We introduce a "null group" to separate the unique attributes from the other attributes and handle them together. We denote this group as "0" and redefine τ_n^* and $\{\tau_{nk}\}$: $\tau_n^* \in \{0, 1, 2, \ldots, K\}$ for $n = 1, 2, \ldots, N$ and $\{\tau_{nk}\}$ are $N \times (K+1)$ variables, which equal 1 if $\tau_n^* = k$ and 0 otherwise. Then, the final model is given by

$$u_j = \sum_{n=1}^{N} \tau_{n0} \cdot \mathbf{x}'_{jn} \boldsymbol{\beta}_n + \sum_{k=1}^{K} g_k \left(\sum_{n=1}^{N} \tau_{nk} \cdot g_k^{-1}(\mathbf{x}'_{jn} \boldsymbol{\beta}_n) \right) + \epsilon_j, \quad (19)$$

where the first term, $\sum_{n=1}^{N} \tau_{n0} \cdot \mathbf{x}'_{jn} \boldsymbol{\beta}_n$, is the total contribution of the attributes in the null group.

We note that introducing the null group provides another property for the nested structure of the proposed model. The proposed model with the null group nests the standard conjoint model as a special case: If $K = 0$, the proposed model is equivalent to the standard conjoint model as in Equation (2).

An example of the unique attributes is price. Price sometimes plays a role of a key criterion in choice context. In the second empirical application, we assume that price is always assigned into the null group. The uniqueness of price can be a strong assumption, and not only price but also other variables could be unique in some cases. We note that our model specification can be flexibly adapted for any assumptions of unique variables: it is needed only to put variables into the null group (i.e., set $\tau_{n0} = 1$ and $\tau_{nk} = 0$ for $k = 1, 2, \ldots, K$) or remove them from the null group (i.e., set τ_{n0} free). Moreover, conventional model selection schemes, such as likelihood ratio, Bayesian information criterion (BIC) (Schwartz 1978), Bayes factor, and log-marginal density (LMD), could be used for hypothetically testing the uniqueness of any variables. For example, we additionally estimated the model not assuming that price is unique in the second empirical application and found no significant differences caused by the assumption.

2.4. Heterogeneity

Different consumers may group the attribute differently as we discussed above; i.e., τ_n^* is heterogeneous across respondents in a conjoint analysis. We probabilistically model the individual assignment, τ_{hn}^*, as follows:

$$\tau_{hn}^* \sim \text{Multinomial}_{K+1}(\theta_{n0}, \theta_{n1}, \ldots, \theta_{nK}), \quad (20)$$

where θ_{nk} indicates the probability of assigning attribute n into benefit k. This probabilistic model implies that we a priori assume that attributes are assigned into only one benefit at the individual level, but there is uncertainty in this assignment; i.e., an attribute can be linked to multiple benefits with probabilities, $\{\theta_{nk}\}$, a posteriori.

The price coefficient and part-worth parameters can be also heterogeneous across respondents. As in previous conjoint studies (e.g., Allenby and Ginter 1995), we assume that the individual parameters are normally distributed as follows:

$$\boldsymbol{\beta}_h = [\boldsymbol{\beta}'_{h1} \; \boldsymbol{\beta}'_{h2} \; \cdots \; \boldsymbol{\beta}'_{hN}]' \sim N(\bar{\boldsymbol{\beta}}, \mathbf{V}_\beta). \quad (21)$$

With individual-specific subscript h and observation-specific subscript t, the full model specification is as follows:

$$u_{hjt} = \sum_{n=1}^{N} \tau_{hn0} \cdot \mathbf{x}'_{hjnt} \boldsymbol{\beta}_{hn} + \sum_{k=1}^{K} g_k \left(\sum_{n=1}^{N} \tau_{hnk} \cdot g_k^{-1}(\mathbf{x}'_{hjnt} \boldsymbol{\beta}_{hn}) \right) + \epsilon_{hjt}. \quad (22)$$

Finally, heterogeneity imposed for two different types of parameters, part-worth and assignment probabilities, is identified by data from a standard conjoint survey in our empirical application. The presence of multiple observations per respondent allows for estimation of the heterogeneity for both types of parameters. We have verified our ability to recover the model parameters with simulated data. Details of our simulation experiment are provided in the online appendix.

2.5. Comparison to Other Methods

Our model integrates the identification of the underlying attribute–benefit mapping into a conjoint analysis, whereas existing methods for investigating benefits require the a priori specification of this mapping. For example, HII (Louviere 1984, Oppewal et al. 1994) utilizes prior or external information about the benefits and requires a complex experimental design to estimate part-worths of attributes. Researchers alternatively might consider using an exploratory analysis to identify the underlying mapping, prior to using the existing methods. Factor analysis can be an alternative to construct groupings of attributes (e.g., Ansari et al. 2002, Batra et al. 2010), but this is a statistical dimension-reduction procedure that relies on variations in the attribute space. By contrast, the proposed model identifies benefits through a theoretical model structure based on the intuitive notion of satiation in attributes, leading to a model of choice. To see this, imagine a situation in which the research collects data on the importance of each attribute from each respondent. The resulting groups from a standard exploratory factor analysis will depend only on the variance/covariance of these importances across respondents, whereas the attribute–benefit mapping consumers actually use to make choices can be completely independent of this variation. Furthermore, the covariation in the importances across respondents can be captured through the upper-level random effect of heterogeneity, $\mathbf{V}_\beta$, in Equation (21) in the proposed model.

While our model is equipped to handle heterogeneity, the identification of the attribute–benefit mapping is not dependent on heterogeneity in attribute importances, and the mapping can be identified even in the case when there is no heterogeneity in the importances. Gathering reliable data describing how each attribute fits within the choice process would be challenging, and the resulting groups may still not yield the true underlying benefits used in making real choices. By comparison, the modeling and data requirements of our model are quite parsimonious, and the resulting benefits are identified through choices rather than variations in self-explicated ratings and other such data.

3. Empirical Applications

We apply our proposed benefit-based conjoint (BBC) model to two different data sets. The first application, digital cameras, illustrates the general features of the proposed model. The second application is a frequently purchased consumer packaged good and focuses on a detailed comparison between the proposed and standard multinomial logit (MNL) model.

3.1. Digital Cameras

The first data set was previously used by Allenby et al. (2014) to estimate the value of product features in the "point and shoot" digital camera market. We apply our model to a sample of 492 respondents who were presented 16 choice tasks, each consisting of four hypothetical products and an outside no-choice option. Table 1 displays the attributes and associated levels in the study. Each option in the choice task has seven attributes. The first attribute is a horizontal attribute of brand name that is coded using four binary variables, one for each of the brands under study. The last attribute is price and ranges from \$129 to \$279. We treat price as a linear attribute and estimate a single price coefficient in each of the estimated models; i.e., $\beta_{\text{price},h}\, p_{hjt}$. All of the other attributes are ordinal and binary. As a result, we have a total of $N = 10$ part-worth coefficients for nine binary variables ($x_{hj,1,t}, x_{hj,2,t}, \ldots, x_{hj,9,t}$) and one continuous variable ($x_{hj,10,t} = p_{hjt}$) for each respondent (h), option (j), and task (t).

Table 1. Attributes and Levels: Digital Camera Data

Attributes	Level 1	Level 2	Level 3	Level 4
Brand	Canon	Sony	Nikon	Panasonic
Pixels	10 M-pixels	16 M-pixels		
Zoom	4× optical	10× optical		
Video	HD (720p)	Full HD (1080p)		
Swivel screen	No	Yes		
WiFi	No	Yes		
Price	Continuous variable ranging from \$129 to \$279			

3.1.1. Alternative Models.

We use the first 14 tasks for model calibration, with the two remaining tasks used for evaluating predictive performance. We estimate the proposed BBC model assuming that $f(y) = \log(y+1)$ for all benefits except the null group; i.e.,

$$u_{hjt} = \sum_{n=1}^{N} \tau_{hn0} \cdot x_{hjnt}\beta_{hn} + \sum_{k=1}^{K} g\left(\sum_{n=1}^{N} \tau_{hnk} \cdot g^{-1}(x_{hjnt}\beta_{hn})\right) + \epsilon_{hjt}, \quad (23)$$

where $g(y) = \operatorname{sgn}(y)\log(|y|+1)$. We also estimate a standard MNL model, the special case of the BBC model where $K = 0$, as a benchmark model

$$u_{hjt} = \sum_{n=1}^{N} x_{hjnt}\beta_{hn} + \epsilon_{hjt}. \quad (24)$$

We assume that $\epsilon_{hjt} \sim \mathrm{EV}(0,1)$ in all models.

We employ a hierarchical Bayesian Markov Chain Monte Carlo (MCMC) method for estimation. As discussed in the online appendix, we control for label switching (Stephens 2000) by running the Markov chain for a burn-in period and then imposing ordinal constraints on the assignment probabilities $\theta_{n1}, \theta_{n2}, \ldots, \theta_{nK}$ based on initial parameter estimates. The burn-in period was set at 100,000 MCMC iterations, although only 10,000 iterations were needed for convergence. We then ran additional 100,000 iterations for parameter estimation.

Table 2 displays in-sample and predictive fit performance of the proposed model for various values of K. In-sample fit is measured with the log-marginal density using both the Newton–Raftery (NR)'s (1994) harmonic mean estimator and the Gelfand–Dey (GD)'s (1994) estimator. Both measures indicate that the BBC models provide a significant improvement in fit relative to the MNL model, and that there is little difference in fit among the BBC models. The in-sample hit ratio and hit probability measures echo these findings.

Table 2. In-Sample and Predictive Fit Statistics: Digital Camera Data

	In-sample				Holdout sample	
Models	LMD-NR	LMD-GD	Hit ratio	Hit prob.	Hit ratio	Hit prob.
MNL, $K=0$	−4,591.552	−7,024.096	0.750	0.653	0.596	0.539
BBC, $K=1$	−4,020.185	−5,319.209	0.774	0.693	0.626	0.576
BBC, $K=2$	−3,965.552	−5,238.508	0.781	0.701	0.625	0.578
BBC, $K=3$	−3,940.828	−5,190.446	0.782	0.702	0.624	0.578
BBC, $K=4$	−3,926.250	−5,163.957	0.782	0.702	0.621	0.576
BBC, $K=5$	−4,066.147	−5,241.950	0.781	0.700	0.617	0.572

Details of calculating the fit measures are provided in the online appendix. Holdout sample fit statistics confirm the in-sample results.

3.1.2. Parmeter Estimates. The posterior mean and standard deviation of $\bar{\beta}$ are reported in Table 3. All models show similar results in terms of ranking the relative importance of the different attributes. Canon is perceived as the best brand and Panasonic the worst. As for the camera features, zoom is the most important and swivel screen is the least important feature in all models. The magnitude of the BBC estimates are larger than those in the MNL model because of the presence of satiation effects in the BBC model.

Table 4 displays the estimated assignment probabilities (θ). All attributes except price are consistently assigned into the first benefit. This implies that people may not consider each feature of the "point and shoot" digital cameras independently. The basic quality across usual cameras may be good enough to take a picture in daily life, and thus people may satiate on additional "better" camera features. However, about one-third of people assign pixels and zoom attributes into the null group, implying that there are still a small number of people who consider some features separately. As for price, half of the respondents assign it into the null group (when $K=1$) or the second benefit (when $K=2$), but the other half assign it into the first benefit. Apart from price, there appears to be just one underlying benefit sought by consumers in this category.

Table 3. Posterior Estimates of $\bar{\beta}$: Digital Camera Data

	MNL	BBC	BBC	BBC	BBC	BBC
Variables	$K=0$	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$
Canon	5.40	5.36	5.49	5.35	5.47	5.36
	(0.37)	(0.50)	(0.51)	(0.51)	(0.52)	(0.50)
Sony	5.22	5.21	5.34	5.21	5.36	5.25
	(0.37)	(0.50)	(0.51)	(0.51)	(0.52)	(0.50)
Nikon	5.17	5.07	5.24	5.14	5.24	5.13
	(0.38)	(0.50)	(0.51)	(0.50)	(0.51)	(0.51)
Panasonic	4.89	4.81	5.02	4.90	5.01	4.90
	(0.37)	(0.50)	(0.51)	(0.51)	(0.52)	(0.50)
Pixels	1.12	3.06	2.75	2.56	2.45	2.11
	(0.09)	(0.29)	(0.28)	(0.25)	(0.23)	(0.21)
Zoom	1.47	3.67	3.29	3.04	2.81	2.55
	(0.10)	(0.31)	(0.28)	(0.26)	(0.30)	(0.23)
Video	0.92	3.48	2.87	2.46	2.24	1.81
	(0.08)	(0.43)	(0.36)	(0.29)	(0.32)	(0.26)
Swivel	0.39	1.94	1.78	1.48	1.26	0.95
	(0.07)	(0.40)	(0.34)	(0.28)	(0.27)	(0.20)
WiFi	0.73	3.23	2.66	2.17	1.92	1.56
	(0.08)	(0.48)	(0.31)	(0.28)	(0.29)	(0.25)
Price	−2.72	−3.01	−2.93	−2.89	−2.91	−2.86
	(0.15)	(0.16)	(0.17)	(0.18)	(0.17)	(0.17)

Note. Standard deviation is in parentheses.

3.2. Consumer Packaged Good

We apply the BBC model to data from a sample of 281 U.S. consumers who participated in a choice-based conjoint study investigating a new chemical formulation of a frequently used household product. The conjoint experiment was part of a larger study conducted by a major packaged goods manufacturer.[4] This manufacturer was interested in understanding how best to deploy its chemical expertise in the arena of the product category, and sought to understand how consumers valued different aspects of what it meant to solve something for the best product design. Thus, the attributes focus exclusively on the chemical formulation of the product. Since the formulation is subject to trade-offs, it is important to understand the relative importance of the different performance dimensions. Exploring the impact of peripheral attributes, such as the packaging, was not part of this study. Also, the study was administered at a central location where subjects were presented physical samples illustrating different attribute levels. Using physical representations of the attribute levels helps standardize stimuli and reduces idiosyncratic interpretation of attribute level descriptions.

Respondents were presented 20 choice tasks, each consisting of three alternatives and an outside option. The alternatives are characterized by 10 different attributes, of which five are shown in each choice task. An example choice task is displayed in Figure 3. The first seven attributes are described on three levels and the remaining three attributes on two levels. Levels correspond to the performance of a "typical" product, a

Table 4. Posterior Estimates of Assignment Probabilities (θ): Digital Camera Data

	$K=1$		$K=2$			$K=3$				$K=4$					$K=5$					
Attributes	Null	B1	Null	B1	B2	Null	B1	B2	B3	Null	B1	B2	B3	B4	Null	B1	B2	B3	B4	B5
Canon	0.03	**0.97**	0.03	**0.88**	0.09	0.03	**0.84**	0.08	0.05	0.03	**0.80**	0.07	0.05	0.05	0.04	**0.74**	0.07	0.05	0.05	0.05
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)
Sony	0.05	**0.95**	0.03	**0.89**	0.07	0.03	**0.84**	0.07	0.05	0.03	**0.82**	0.05	0.05	0.04	0.04	**0.76**	0.06	0.05	0.05	0.04
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.02)
Nikon	0.08	**0.92**	0.04	**0.87**	0.09	0.04	**0.84**	0.07	0.05	0.04	**0.80**	0.07	0.05	0.05	0.04	**0.74**	0.07	0.05	0.05	0.05
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)
Panasonic	0.04	**0.96**	0.03	**0.93**	0.04	0.04	**0.88**	0.04	0.04	0.04	**0.84**	0.04	0.04	0.04	0.05	**0.78**	0.04	0.04	0.05	0.04
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)	(0.05)	(0.02)	(0.03)	(0.03)	(0.02)
Pixels	0.33	**0.67**	0.30	**0.54**	0.16	0.21	**0.51**	0.15	0.14	0.14	**0.49**	0.13	0.16	0.08	0.12	**0.42**	0.14	0.08	0.17	0.08
	(0.04)	(0.04)	(0.05)	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.06)	(0.07)	(0.07)	(0.06)	(0.04)	(0.05)	(0.08)	(0.07)	(0.04)	(0.06)	(0.04)
Zoom	0.39	**0.61**	0.38	**0.51**	0.11	0.28	**0.45**	0.13	0.14	0.20	**0.40**	0.13	0.12	0.15	0.14	**0.35**	0.13	0.13	0.11	0.13
	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.08)	(0.08)	(0.07)	(0.06)	(0.07)
Video	0.21	**0.79**	0.20	**0.64**	0.16	0.13	**0.56**	0.16	0.15	0.10	**0.51**	0.14	0.12	0.13	0.10	**0.39**	0.14	0.16	0.12	0.08
	(0.04)	(0.04)	(0.05)	(0.08)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.05)	(0.10)	(0.07)	(0.06)	(0.06)	(0.05)	(0.10)	(0.08)	(0.06)	(0.06)	(0.04)
Swivel	0.11	**0.89**	0.07	**0.77**	0.16	0.06	**0.74**	0.12	0.08	0.06	**0.68**	0.10	0.08	0.08	0.07	**0.56**	0.12	0.09	0.09	0.08
	(0.03)	(0.03)	(0.03)	(0.07)	(0.07)	(0.03)	(0.07)	(0.06)	(0.05)	(0.03)	(0.08)	(0.06)	(0.05)	(0.04)	(0.04)	(0.11)	(0.07)	(0.05)	(0.05)	(0.05)
WiFi	0.14	**0.86**	0.09	**0.72**	0.19	0.07	**0.66**	0.17	0.11	0.07	**0.63**	0.12	0.10	0.09	0.06	**0.52**	0.13	0.09	0.10	0.09
	(0.03)	(0.03)	(0.04)	(0.07)	(0.07)	(0.04)	(0.07)	(0.07)	(0.06)	(0.04)	(0.08)	(0.06)	(0.05)	(0.05)	(0.04)	(0.10)	(0.07)	(0.05)	(0.06)	(0.05)
Price	**0.58**	0.42	0.09	0.34	**0.56**	0.07	0.31	**0.44**	0.18	0.06	0.31	**0.33**	0.16	0.14	0.06	0.28	**0.31**	0.12	0.13	0.11
	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)	(0.04)	(0.05)	(0.09)	(0.08)	(0.03)	(0.04)	(0.07)	(0.07)	(0.06)	(0.03)	(0.05)	(0.08)	(0.06)	(0.06)	(0.05)

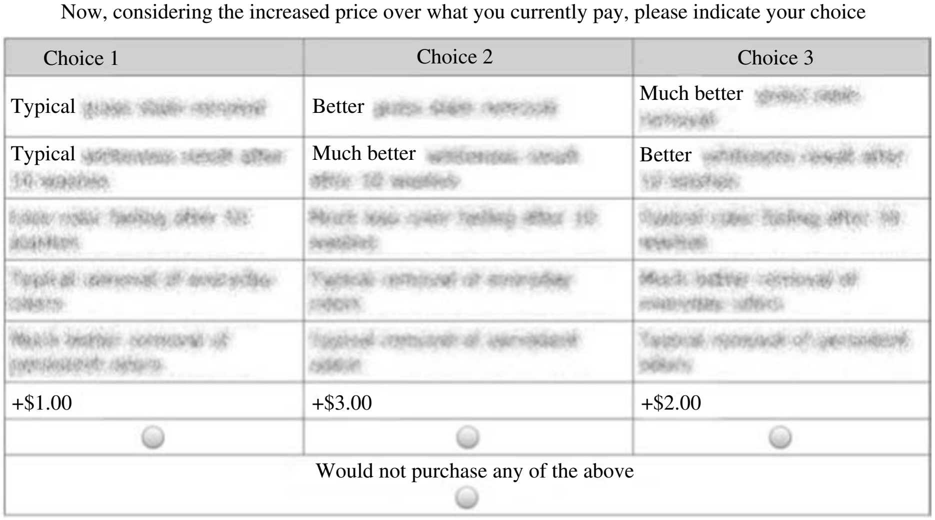
Notes. Bold numbers indicate the maximum assignment probability of an attribute. Standard deviation is in parentheses.

"better" higher-tier product, and a "much better" state-of-the-art product. The first seven attributes are illustrated using physical samples, while the latter three attributes are only described verbally. Respondents were told to assume that attributes not shown in a task are at their "typical" level. Price entered as a price premium over a currently used product. This is consistent with the attribute levels reflecting "typical" or "better" attribute levels. There are 3 levels of price premiums ($1, $2, $3), and we use one linear price coefficient in each of the models; i.e., $\beta_{\text{price},h}\, p_{hjt}$. Based on the nature of the stimuli, the outside option may be perceived as a "typical" product in terms of all of the attributes. We therefore fix the value of each "typical" level to zero for identification. As a result, we have a total of 18 variables including price for each option in a choice task.

3.2.1. Alternative Models. We randomly selected 18 tasks for model calibration for each respondent. The remaining two tasks are used for evaluating the holdout sample performance. As in the first application, we estimate the BBC model assuming that $f(y) = \log(y+1)$ for all benefits except the null group for parsimony. There is only one difference from the model used in

Figure 3. An Example Choice Task: CPG Data

Now, considering the increased price over what you currently pay, please indicate your choice



(1 of 20)

the first application in that the price attribute is always assigned into the null group; i.e.,

$$u_{hjt} = \sum_{n=1}^{N} \tau_{hn0} \cdot \mathbf{x}'_{hjnt}\boldsymbol{\beta}_{hn} + \sum_{k=1}^{K} g\left(\sum_{n=1}^{N} \tau_{hnk} \cdot g^{-1}(\mathbf{x}'_{hjnt}\boldsymbol{\beta}_{hn})\right) + \beta_{\text{price},h}\, p_{hjt} + \epsilon_{hjt}, \tag{25}$$

where $g(y) = \operatorname{sgn}(y)\log(|y|+1)$.[5] We also estimate a standard MNL model, the special case of the BBC model where $K = 0$, as a benchmark model[6]

$$u_{hjt} = \sum_{n=1}^{N} \mathbf{x}'_{hjnt}\boldsymbol{\beta}_{hn} + \beta_{\text{price},h}\, p_{hjt} + \epsilon_{hjt}. \tag{26}$$

We assume that $\epsilon_{hjt} \sim \text{EV}(0, 1)$ in all models.

The model parameters are estimated in the same manner as in the first application. We ran 20,000 MCMC iterations[7] first and set the order of the assignment probabilities based on the initial iterations. We then ran additional 50,000 MCMC iterations with the order restriction on the assignment probabilities. We used the last 20,000 draws to obtain the posterior distribution of the model parameters. For the benchmark model, we ran 70,000 MCMC iterations and used the last 20,000 draws to obtain the posterior distribution.

We investigated the BBC model with various values of K and calculated the same fit measures as in the first application. Table 5 displays the fit measures. Although the NR and GD approximations of the LMD estimate pick different models (BBC models with $K = 2$ and $K = 3$) as the best, the difference between models in terms of both measures is small. As discussed in the first application, this marginal difference between the models with $K = 2$ and $K = 3$ implies that only two benefits may be enough and the additional groups may not be necessary to describe the choice data. This interpretation is supported by the posterior estimates of the assignment probabilities: the third benefit is the benefit where any attribute is not dominantly assigned, and therefore it might be redundant. We therefore focus on $K = 2$ (henceforth, $\text{BBC}_{K=2}$ model), which favorably compares to the MNL model in terms of in-sample and predictive fit.

3.2.2. Parameter Estimates. The posterior distribution of $\bar{\boldsymbol{\beta}}$ for the $\text{BBC}_{K=2}$ and MNL models is reported in Figure 4. Both models show similar results in terms of ranking the relative importance of the different attributes. The respective "high" levels of attributes 3 and 4 have the highest part-worths in both models. However, there are some clear differences, which can be seen from Figure 4: The standard MNL model implies a price coefficient of zero, whereas the $\text{BBC}_{K=2}$ model yields a negative price coefficient. Within the $\text{BBC}_{K=2}$ model, the estimates of several elements of $\bar{\boldsymbol{\beta}}$ are larger than within the MNL model, especially for the "high" attribute level (attributes 1, 3, 4, and 5, and to some extent attributes 8, 9, and 10). The remaining coefficients are at similar values. Thus, if considered

Figure 4. Comparing Posteriors of $\bar{\boldsymbol{\beta}}$ for MNL and $\text{BBC}_{K=2}$: CPG Data

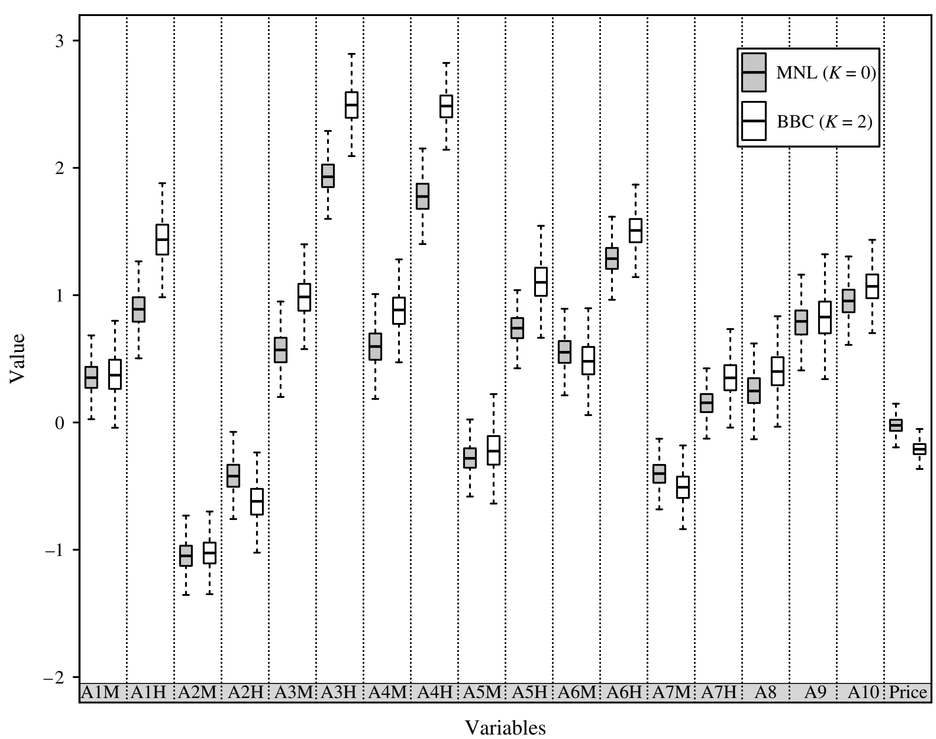


Table 5. In-Sample and Predictive Fit Statistics: CPG Data

	In-sample				Holdout sample	
Models	LMD-NR	LMD-GD	Hit ratio	Hit prob.	Hit ratio	Hit prob.
MNL, $K = 0$	−4,126.333	−5,478.135	0.678	0.581	0.487	0.444
BBC, $K = 1$	−4,027.136	−5,166.228	0.688	0.592	0.488	0.446
BBC, $K = 2$	−3,959.641	−5,029.382	0.693	0.598	0.493	0.451
BBC, $K = 3$	−3,961.891	−5,009.718	0.694	0.599	0.489	0.448

independently, the attributes associated with one benefit yield relatively larger increases in the overall preference, consistent with the estimation result reported for the first application.

Table 6 displays the estimated assignment probabilities (θ). In the $\text{BBC}_{K=2}$ model, only attribute 6 has a high probability of being assigned into the null group (0.52). Consistently, the estimates of $\bar{\boldsymbol{\beta}}$ for attribute 6 do not differ much between both models. We therefore conclude that there is a three-benefit solution: one single-attribute benefit (attribute 6) and two multiattribute benefits, where attributes 1, 2, 4, and 8 contribute to benefit 1 and attributes 3, 5, 7, 9, and 10 contribute to benefit 2.

Finally, we compared the estimated assignment probabilities between the BBC models with $K = 2$ and $K = 3$, and find consistent results. Relatively important attributes (attributes 1–5 and 10) are assigned into one of the benefits with satiation, and the pattern of partitioning of these attributes is identical in both models. Moreover, none of the attributes are assigned into the third benefit in the model with $K = 3$ with high probability. This implies that the third benefit may be redundant, supporting our choice of the $\text{BBC}_{K=2}$ model.

4. Discussion

The empirical results of the second application indicate that consumer utility is formed on the basis of a three-dimensional benefit space rather than a 10-dimensional attribute space. This section provides a discussion of our benefit solution. We first examine the nature of the benefit space and then explore product design implications within and across benefits. We focus analysis on the benefits with satiation in the $\text{BBC}_{K=2}$ model.

4.1. Benefit Characterization

We asked the company sponsoring the study to characterize the different attributes, hoping that the attribute–benefit map would be meaningful to them. Table 7 displays their interpretation of the results.

Table 6. Posterior Estimates of Assignment Probabilities (θ): CPG Data

	$K = 1$		$K = 2$			$K = 3$			
Attributes	Null	Benefit 1	Null	Benefit 1	Benefit 2	Null	Benefit 1	Benefit 2	Benefit 3
A1	0.09	0.91	0.07	**0.85**	0.07	0.07	0.76	0.08	0.09
	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.07)	(0.04)	(0.05)
A2	0.16	0.84	0.12	**0.74**	0.14	0.13	0.52	0.13	0.23
	(0.09)	(0.09)	(0.07)	(0.10)	(0.07)	(0.07)	(0.16)	(0.07)	(0.15)
A3	0.83	0.17	0.15	0.12	**0.73**	0.14	0.11	0.59	0.16
	(0.08)	(0.08)	(0.07)	(0.06)	(0.08)	(0.08)	(0.06)	(0.12)	(0.09)
A4	0.13	0.87	0.07	**0.87**	0.05	0.08	0.78	0.06	0.08
	(0.07)	(0.07)	(0.04)	(0.05)	(0.03)	(0.05)	(0.07)	(0.03)	(0.05)
A5	0.89	0.11	0.14	0.07	**0.79**	0.13	0.07	0.63	0.18
	(0.06)	(0.06)	(0.07)	(0.04)	(0.07)	(0.07)	(0.04)	(0.13)	(0.11)
A6	0.63	0.37	**0.52**	0.26	0.22	0.29	0.19	0.19	0.33
	(0.13)	(0.13)	(0.12)	(0.10)	(0.10)	(0.12)	(0.09)	(0.09)	(0.14)
A7	0.52	0.48	0.30	0.32	**0.38**	0.20	0.26	0.30	0.24
	(0.12)	(0.12)	(0.12)	(0.11)	(0.12)	(0.10)	(0.11)	(0.13)	(0.12)
A8	0.61	0.39	0.35	**0.38**	0.28	0.24	0.26	0.24	0.25
	(0.16)	(0.16)	(0.14)	(0.15)	(0.12)	(0.12)	(0.12)	(0.11)	(0.13)
A9	0.33	0.67	0.19	0.34	**0.47**	0.16	0.26	0.28	0.29
	(0.13)	(0.13)	(0.10)	(0.13)	(0.15)	(0.09)	(0.13)	(0.13)	(0.15)
A10	0.76	0.24	0.23	0.12	**0.65**	0.19	0.13	0.45	0.23
	(0.09)	(0.09)	(0.10)	(0.06)	(0.11)	(0.10)	(0.06)	(0.14)	(0.12)

Notes. Bold numbers indicate the maximum assignment probability of an attribute. Standard deviation is in parentheses.

Table 7. Attribute Focus Characteristic as Described by the Company: CPG Data

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
New	New	Old	New	Old	New	New	Both	Old	Both

Some attributes are relevant to solving "new" problems. These problems surface eventually as the consumer takes care of the household. Other attributes are seen as being related to existing or "old" problems. These problems have already developed over time and have not been adequately dealt with yet. The assignment of some attributes into the two groups is not clear and these are labeled as "both." Consumers may be concerned with either or both benefits when purchasing and using the product.

The first benefit is characterized mostly by attributes 1, 2, 4, and 8 (85%, 74%, 87%, and 38%). These attributes relate to solving immediate problems that consumers face and try to solve by applying the product. Attributes 1 and 2 had previously been considered very important attributes in the product category, and research has focused on improving them. Attribute 4 is an attribute that has traditionally been associated with the product category. Attribute 8, with an assignment probability of 38%, is a somewhat surprising assignment that the company feels could be construed as having either a new- or old-problem orientation and is therefore labeled as "both." There is a 28% probability that attribute 8 is assigned into the second benefit. Overall, the grouping is seen by the firm to be plausible, with the biggest surprise being the existence of satiation among attributes within the benefit. It was hoped that each of the benefits would be viewed independently by consumers.

The second benefit is associated with attributes 3, 5, 7, 9, and 10. Attributes 3 and 5 refer to problems that have been developed over time, and using the product may finally solve these problems. Their corresponding assignment probabilities are large (73% and 79%) and are clearly seen by the firm as being related to the "old" problems. The remaining attributes have somewhat lower assignment probabilities. It is important to remember that the assignment probabilities reported in Table 6 reflect respondent heterogeneity as well as modeling error, and that the assignment probabilities clearly point to one of the benefits as being most plausible for each attribute.

The null group absorbs all attributes that constitute their own single-attribute benefits. In the empirical application, only attribute 6 relates to the null group. While it also has a 26% chance of being assigned into benefit 1, along with the other "new" focus attributes, it is most likely to relate to its own benefit.

4.2. Within-Benefit Analysis

The presence of attribute satiation within a benefit has implications for product design. If consumers quickly satiate on the attributes within a benefit, then it may not be advantageous to the firm to attempt to solve all of the problems it is capable of solving, and instead the firm could design their products to solve enough of them. Figure 5 investigates this issue by computing the aggregate cumulative utility for the addition of an attribute at the "high" level. Figure 5(a) shows the cumulative utility for the addition of attributes within the first benefit. Figure 5(b) shows the cumulative

Figure 5. Satiation in Benefits: CPG Data

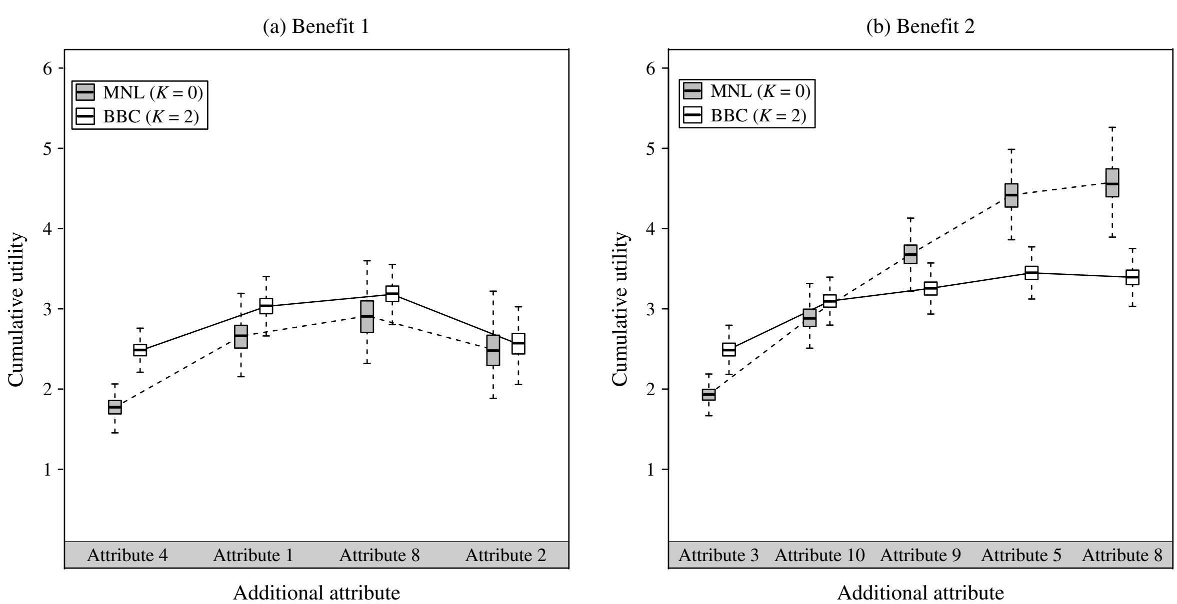
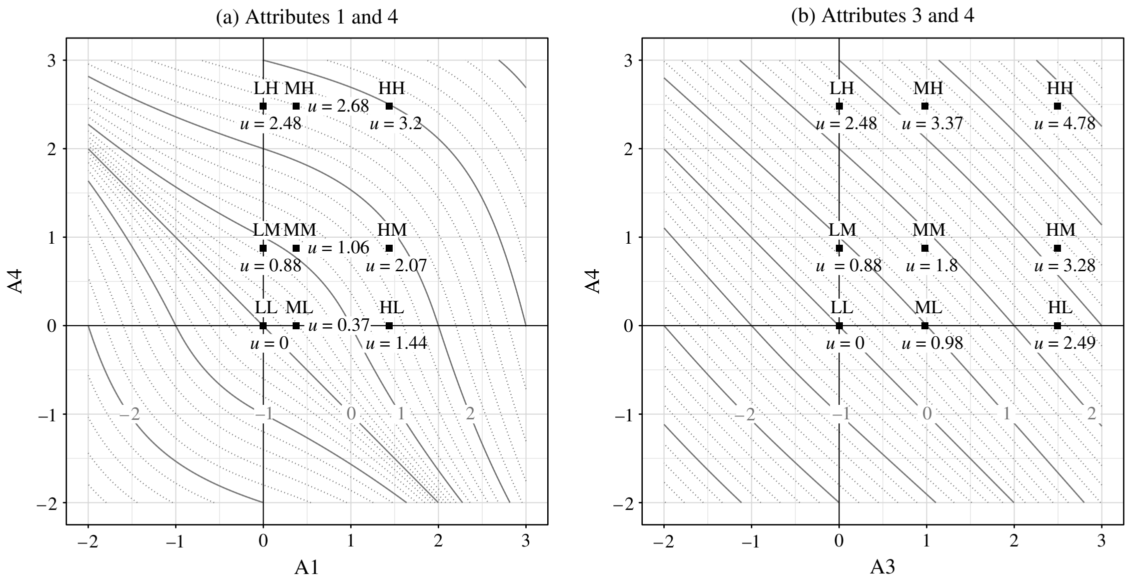


Figure 6. Improving Same- or Different-Benefit Attributes: CPG Data



Notes. The value of u is not perfectly matched with the contour curves since the contour curves are drawn under the same values of beta across individuals.

utility for the second benefit. The order of entry of the attributes is in terms of the magnitude of the part-worth coefficient displayed in Figure 4 for the MNL model. The boxplots represent the posterior distribution of the average cumulative utility from the benefits from the base case of all attributes set at their "typical" level.

The effect of satiation is present for both benefits in the $\text{BBC}_{K=2}$ model and is particularly large for the second benefit. Satiation is not present in the standard MNL model. Our results indicate that there is little incremental gain to having "high" levels of multiple attributes for the second benefit, in contrast to the MNL model. The design implications of the $\text{BBC}_{K=2}$ and MNL models are different.

4.3. Across-Benefit Analysis

The presence of satiation within benefits suggests that firms can increase consumer utility more easily with attributes belonging to different benefits. Figure 6 illustrates the change in overall utility for improvements in two attributes. The attributes in Figure 6(a) belong to the same benefit, while the attributes in Figure 6(b) belong to different benefits. The indifference curves plotted in the figure reflect the integrated preferences of the entire sample of respondents, and the presence of the assignment probabilities (θ) in the model (see Table 6) leads to the curvature in the across-benefit plot in Figure 6(b) despite the additive separable utility specification for the benefits. In both graphs, the origin represents a product with all attributes set to the "typical" or "low" level.

We can see that improvements in attribute 3 hardly affect the preference derived from improvements in attribute 4. This is expected because attributes 3 and 4 are associated with different benefits. However, investments in attribute 1 decrease the preference contribution of improving attribute 4. That is, when attribute 1 is at its low level, an improvement in attribute 4 to its high level is 2.48 utils, whereas the same increase when attribute 1 is set to its high level is 1.76 $(3.20 - 1.44)$.

The overall preference of a product with both attributes set to the "high" or "much better" level is 3.20 in Figure 6(a), which is lower than the utility gain for each attribute individually $(2.48 + 1.44 = 3.92)$. In Figure 6(b), the improvement afforded by each attribute individually is nearly identical to the joint gain $(2.48 + 2.49 = 4.97$ is close to $4.78)$. Overall, our benefit-based conjoint results point to larger increases in utility through investment in attributes belonging to different benefits.

5. Conclusion

This paper proposes a new model for conjoint and demand analysis based on the idea that consumer utility is derived from benefits, which are in turn related to attributes whose effect satiates within benefits. We propose a new benefit-based model for conjoint analysis (BBC) that nests the standard multinomial model, exhibits monotonicity and subadditivity of the value function, and has good mixing properties for estimation. We demonstrate the BBC model using two data sets, one involving demand for digital cameras and another involving a household product. We find that the BBC model results in improved model fit and prediction in both data sets, with the benefit space much smaller than the space of attributes.

The identification of an attribute that leads to a unique benefit, where diminishing marginal returns are not yet realized, offers manufacturers the ability to set their offerings apart from the competition and achieve higher sales. Examples of this are found in the auto industry, where the presence of seemingly minor attributes is believed to have a large effect on demand. A dated example is the presence of beverage holders in cars, and a more recent example is the presence of external stereo speakers in pickup trucks. In our digital camera study, the presence of just one underlying benefit indicates that brands are not well differentiated by the current set of product features because consumers view the attributes as exchangeable in providing the benefit "picture quality." The digital point-and-shoot camera market is in need of attributes providing a new benefit different from the generic benefit currently perceived to be present. Our BBC model can be used to identify attributes with this property.

We expect the within-benefit satiation detected in our data to increase as the number of attributes increases. As the number of attributes under investigation grows in any study, there is an increased likelihood that consumers process the product descriptions in a lower dimensional space. This is particularly true in studies involving new product characteristics consumers may not yet understand. An advantage of our model is that it allows companies to study a large number of attributes without implicitly imposing a model structure that is unrealistic.

Our model can be viewed as a generalization of a Lancasterian model that allows for nonlinear effects of the attribute levels on consumer utility. Lancasterian models (Lancaster 1966, Kim et al. 2007, Dubois et al. 2014) focus on segregating utility-generating sources, such as food ingredients, mixed in a product, whereas our model seeks to aggregate the impacts of attributes on utility. Benefits could be the intersection of both modeling approaches, where consumers actually derive utility. Lancasterian models traditionally have required some linear forms of quantity measurements of the attributes that affect utility. Our model estimates the quantity in terms of an attribute's part-worth and simultaneously groups attributes into benefits so that nonlinear satiating effects are present. Given the prevalence of attribute-based research in marketing, we expect our model to have wide potential application.

Our results indicate that investing in new benefit dimensions is a useful way to escape attribute satiation within benefits, allowing firms to identify truly new product innovations. Our model can be useful in facilitating such analysis. Additional work is needed in testing other shapes of satiation and incorporating consumer needs to better understand potential sources of heterogeneity. It is still unclear which shape of the benefit function is optimal. We leave these and other issues for future research.

Endnotes

[1] Precisely, b is subadditive in terms of the absolute values. If b is monotonic, $b(\mathbf{0}) = 0$, and Equation (10) is satisfied, b is subadditive for B^+ and superadditive for B^-: Since Equation (9) implies that $0 \le b(\mathbf{a}_{S_l^+}) \le b(\mathbf{a}_{B^+})$, and $b(\mathbf{a}_{B^-}) \le b(\mathbf{a}_{S_l^-}) \le 0$, we have $b(\mathbf{a}_{B^+}) \le \sum_{l=1}^{L^+} b(\mathbf{a}_{S_l^+})$ and $b(\mathbf{a}_{B^-}) \ge \sum_{l=1}^{L^-} b(\mathbf{a}_{S_l^-})$.

[2] This condition is not necessary because a monotonically increasing and concave real-valued function can be defined without this condition. However, we include this condition to exploit useful properties of functions under this condition.

[3] This does not imply the opposite of loss aversion (e.g., Tversky and Kahneman 1991). If loss aversion exists, we may need fewer negative attributes. Yet, we still need much more negative attributes as the value of the positive attribute increases.

[4] Because of confidentiality agreements, a full description of the data cannot be disclosed.

[5] We also tried to estimate different versions of the model: (i) the BBC model with a power function, $f(y) = y^\delta$, $0 < \delta < 1$, which is more flexible than the simple logarithmic function that we use in our empirical application; and (ii) the BBC model not assuming that the price is a unique, stand-alone variable. The overall estimation result of these different versions is consistent with that of the model with the simple logarithmic function and unique price variable.

[6] We also estimated an MNL model with two-way interactions. There are 129 interaction terms. Because of sparseness of the data set, each interaction term may not truly capture the interaction effects but serve as a dummy for a specific choice alternative in a specific observation, leading to the estimates of the main effects to be totally different from those from the BBC and standard MNL models. An extra simulation analysis confirms that the interaction terms just make the model extremely overspecified.

[7] We found that the likelihood and parameter estimates converge after the first 5,000 iterations.

References

Allenby GM, Ginter JL (1995) Using extremes to design products and segment markets. *J. Marketing Res.* 32(4):392–403.

Allenby GM, Brazell JD, Howell JR, Rossi PE (2014) Economic valuation of product features. *Quant. Marketing Econom.* 12(4): 421–456.

Ansari A, Jedidi K, Dube L (2002) Heterogeneous factor analysis models: A Bayesian approach. *Psychometrika* 67(1):49–77.

Batra R, Lenk P, Wedel M (2010) Brand extension strategy planning: Empirical estimation of brand-category personality fit and atypicality. *J. Marketing Res.* 47(2):335–347.

Bhat CR (2008) The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Res. Part B: Methodological* 42(3):274–303.

Calantone RJ, Sawyer AG (1978) The stability of benefit segments. *J. Marketing Res.* 15(3):395–404.

Day GS, Shocker AD, Srivastava RK (1979) Customer-oriented approaches to identifying product-markets. *J. Marketing* 43(4): 8–19.

Dixit AK, Stiglitz JE (1977) Monopolistic competition and optimum product diversity. *Amer. Econom. Rev.* 67(3):297–308.

Dubois P, Griffith R, Nevo A (2014) Do prices and attributes explain international differences in food purchases? *Amer. Econom. Rev.* 104(3):832–867.

Gelfand AE, Dey DK (1994) Bayesian model choice: Asymptotics and exact calculations. *J. Royal Statist. Soc. Ser. B (Methodological)* 56(3):501–514.

Green PE, Srinivasan V (1978) Conjoint analysis in consumer research: Issues and outlook. *J. Consumer Res.* 5(2):103–123.

Green PE, Srinivasan V (1990) Conjoint analysis in marketing: New developments with implications for research and practice. *J. Marketing* 54(4):3–19.

Haley RI (1968) Benefit segmentation: A decision-oriented research tool. *J. Marketing* 32(3):30–35.

Johnson EJ, Meyer RJ (1984) Compensatory choice models of noncompensatory processes: The effect of varying context. *J. Consumer Res.* 11(1):528–541.

Kim J, Allenby GM, Rossi PE (2002) Modeling consumer demand for variety. *Marketing Sci.* 21(3):229–250.

Kim J, Allenby GM, Rossi PE (2007) Product attributes and models of multiple discreteness. *J. Econometrics* 138(1):208–230.

Lamberton CP, Diehl K (2013) Retail choice architecture: The effects of benefit- and attribute-based assortment organization on consumer perceptions and choice. *J. Consumer Res.* 40(3):393–411.

Lancaster KJ (1966) A new approach to consumer theory. *J. Political Econom.* 74(2):132–157.

Lee S, Kim J, Allenby GM (2013) A direct utility model for asymmetric complements. *Marketing Sci.* 32(3):454–470.

Louviere JJ (1984) Hierarchical information integration: A new method for the design and analysis of complex multiattribute judgment problems. *Adv. Consumer Res.* 11(1):148–155.

Michalek JJ, Feinberg FM, Papalambros PY (2005) Linking marketing and engineering product design decisions via analytical target cascading. *J. Product Innovation Management* 22(1):42–62.

Myers JH (1976) Benefit structure analysis: A new tool for product planning. *J. Marketing* 40(4):23–32.

Newton MA, Raftery AE (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Royal Statist. Soc. Ser. B (Methodological)* 56(1):3–48.

Oppewal H, Louviere JJ, Timmermans HJ (1994) Modeling hierarchical conjoint processes with integrated choice experiments. *J. Marketing Res.* 31(1):92–105.

Satomura T, Kim J, Allenby GM (2011) Multiple-constraint choice models with corner and interior solutions. *Marketing Sci.* 30(3): 481–490.

Schwartz G (1978) Estimating the dimension of a model. *Annals Statist.* 6(2):461–464.

Stephens M (2000) Dealing with label switching in mixture models. *J. Royal Statist. Soc. Ser. B (Statist. Methodology)* 62(4):795–809.

Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. *Quart. J. Econom.* 106(4): 1039–1061.

Wedel M, Steenkamp JBE (1991) A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *J. Marketing Res.* 28(4):385–396.

Wind Y (1973) A new procedure for concept evaluation. *J. Marketing* 37(4):2–11.