Form + Function: Optimizing Aesthetic Product Design via Adaptive, Geometrized Preference Elicitation

Namwoo Kang
K-School
Korea Advanced Institute of Science and Technology (KAIST)
nwkang@kaist.ac.kr

Yi Ren
Department of Mechanical Engineering
Arizona State University
yiren@asu.edu

Fred M. Feinberg
Ross School of Business and Department of Statistics
University of Michigan
feinf@umich.edu

Panos Y. Papalambros
Department of Mechanical Engineering
University of Michigan
pyp@umich.edu

September 6, 2016

The authors gratefully acknowledge the support of University of Michigan Graham Sustainability Institute Fellowship.

Under review at Marketing Science; please do not quote or distribute.
Form + Function: Optimizing Aesthetic Product Design via Adaptive, Geometrized Preference Elicitation

Abstract

Visual design is critical to product success, and the subject of intensive marketing research effort. Yet visual elements, due to their high-dimensional, holistic, and interactive nature, do not lend themselves well to optimization using extant methods for preference elicitation. Here we present a systematic methodology to incorporate real-time, interactive, 3D-rendered product configurations into a conjoint framework. The method relies on rapid, scalable machine learning algorithms to adaptively update product designs along with standard information-oriented product attributes. At the heart of the method is a parametric decomposition of a product's geometry, along with a novel, adaptive “bi-level” query task that can estimate individuals’ preferences among visual designs as well as their trade-offs against such traditional elements as price and product features. We illustrate the method’s performance through extensive simulations and a field test for the design of a mid-priced sedan, using real-time 3D rendering and an online panel. Results indicate not only substantially enhanced predictive accuracy, but two quantities beyond the reach of standard conjoint methods: trade-offs between form and function overall, and willingness-to-pay for specific design elements.

Keywords: product design optimization; conjoint analysis; machine learning; preference elicitation; visual design
1 Introduction

A product’s visual form has long been acknowledged as a pivotal element of consumer choice (Kotler and Rath 1984, Bloch 1995, Veryzer and Hutchinson 1998, Dahl et al. 1999, Bloch et al. 2003). Firms as diverse as Bang & Olufsen, Apple, and Tesla have not only made visual design an emblematic linchpin of their success, but have also helped propel design, as both discipline and corporate mission, into the public sphere. Even old-line firms have taken note: IBM’s product design team is now among the largest in the world.¹

Academic research in both marketing and engineering has consequently grappled with how to capture, and ultimately optimize, product “form,” with varying definitions, terminologies, and degrees of success. Preference modeling studies incorporating visual design elements have addressed general concepts like “form” (Bloch 1995, Orsborn and Cagan 2009, Tseng et al. 2012), “styling” (Dotson et al. 2016), “design” (Landwehr et al. 2011, Ren and Papalambros 2011), and “appearance” (Creusen and Schoormans 2005), as well as more specific ones like “shape” (Orsborn et al. 2009, Kelly et al. 2011), “silhouette” (Reid et al. 2012), and “profile” (Lai et al. 2005). “Form” – broadly construed – often assumes a central role in real-world preference modeling problems. According to Bloch (1995), form helps attract consumer attention, communicate product information, and stimulate visual pleasure, thereby generating a long-lasting perceptual impression. In practice, sales predictions can be significantly improved by accommodating form (Landwehr et al. 2011). Marketers and designers also find valuable trade-offs between form and functional attributes (Dotson et al. 2016, Sylcott et al. 2013a), and revealing such trade-offs can lead to superior balance between product appeal and functionality (Reid et al. 2012).

Product aesthetics comprises purely visual elements like color and packaging, haptic (Peck and Childers 2003) and sensory (Krishna 2012) aspects that can be altered and optimized independently, and more nuts-and-bolts ‘geometric’ elements that both convey product image and constrain / interact with the internal operations of the product itself. Optimizing these geometrical elements is critical for efficient design of components and production processes, but mapping from the geometry of a product to how much potential consumers might like it is a complex exercise (Michalek et al. 2005). A further challenge is quantifying how elements of form preference are traded off against ‘traditional’ attributes like price and performance. Such trade-offs are critical in allowing designers to determine not only how important “design” is overall to a particular

¹ The Times of London (www.thetimes.co.uk/tto/public/perfectharmony/article4581302.ece) and Wired (www.wired.com/2014/12/disappearing-business-of-design/).
consumer or type of consumer, but whether specific aspects of design – like a low-profile car hood that may constrain the powertrain and require costly amendments to maintain crashworthiness – can be accommodated, subject to supply-side and consumer budget constraints.

Marketers have traditionally employed conjoint-based techniques to assess consumer trade-offs (Green and Srinivasan 1990; Green, Krieger, and Wind 2001), but have struggled to incorporate product form into overall preference modeling. One method of doing so involves the elicitation of adjectival descriptors. For example, the largest survey of its kind, Maritz Corp.’s New Vehicle Car Survey (NVCS), asks 250,000 new car buyers annually to rate statements like “styling is at or near the top of important characteristics in a new vehicle”; to rate the importance of, satisfaction with, and likelihood of rejecting vehicles based on interior and exterior styling; to assess trade-offs with vehicle safety; and to assess 28 “image” elements. Yet this procedure still leaves designers in a quandary as to what to actually design: while it’s helpful to learn a particular consumer likes “sleek” cars, what does such a car actually look like? Reasonable people can differ on the operationalization of adjectival labels: one consumer’s sleek is another consumer’s clunky, and a sleek SUV may have a radically different visual footprint from a sleek sports car. Moreover, potential buyers can express design preferences but be in no position to enact them, due to financial, familial, or other constraints, making it difficult to know how “design,” as an overall attribute – or various aspects of design, like “ruggedness” – is traded-off against more prosaic but measurably important elements like price, safety, and other features.

Another, more effortful, method of implementation involves eliciting desired descriptions from consumers and attempting to pre-render these into appealing designs. For example, consumers wanting a “sophisticated” car will find certain physical shapes closer to this ideal than others. Dotson et al. (2016) discuss the conjoint literature devoted to this approach, and present a method to incorporate a set of images parsimoniously into standard conjoint tasks. Although this is doubtless an advance on using adjectives alone to describe desired visual attributes, it does entail several substantive and implementational shortcomings that characterize much prior research: first, it requires that researchers pre-determine which design elements are desirable for consumers and feasible for manufacturers; second, it requires these design elements to be rendered into suitable stimuli; and third, it requires that these stimuli be manageable in number while

---

2 Specifically, the 28 image elements were: classic, responsive, youthful, bold, luxurious, prestigious, stylish, functional, safe, innovative, economical, simple, conservative, environmentally-friendly, sleek, distinctive, fun to drive, comfortable, well-engineered, rugged/tough, elegant, family-oriented, good value, sporty, powerful, sophisticated, exciting.
incorporating a potentially large set of design elements that could have substantial interactions, both positive (e.g., sleek and sporty) and negative (e.g., sleek and sturdy) in terms of their ability to be jointly achieved in the same design, as well as downstream desirability.

This last shortcoming – the need to avoid large visual stimuli sets – although addressable through split questionnaire techniques that can distribute the econometric burden across many consumers (Adigüzel and Wedel 2008), is exacerbated by the intrinsic nonlinearity of the visual design space. To take a simple example, a consumer who likes yellow cars and loves red ones may strongly dislike orange ones, despite its being “intermediate” in a canonical parameterization of the color space. That is, designers cannot determine a set of stimuli among which a target consumer has modest preference differentials and simply interpolate between them to determine a utility surface. Moreover, keeping target visual stimuli constant across respondents runs the risk of mismeasuring heterogeneity: ideally, each respondent should be able to veer into the region of the design space containing her most preferred product configuration, as opposed to being imprisoned within the convex hull or simplex edge bordered by the pre-configured designs. Indeed, a key (and achieved) objective for Dotson et al. (2016) was keeping the number of stimuli to a manageable level, given that most respondents exhibit substantial fatigue in choice-based conjoint designs within 20-25 queries.³ Lastly, or screen images are also two-dimensional and fixed, mere projections of the 3D designs that respondents must visualize and integrate to fully experience. Although this reduces both complexity in presentation and latency in administration, it substantially reduces realism and respondent involvement.

We later review research in preference modeling, efficient algorithms, and design optimization (see section 2); yet to the authors’ knowledge, few studies combine a form preference model with one for overall preference, and even these few rely on decomposition via disjoint measurement. For example, both Dotson et al. (2016) and Sylcott et al. (2013) conducted separate surveys, one each for modeling form and overall preferences, which need to be deployed on populations with similar characteristics. This is reasonable for lab studies where strict demographics controls can be exacted; it is more problematic when applied to online “crowdsourced” groups that can differ in crucial characteristics. Combining or “fusing” data across

³ Dotson et al. (2016) “include 32 distinct images that can be paired with any combination of the other attributes in the choice survey... respondents answered 20 choice tasks each with 4 alternatives, which means that each image appeared in the choice tasks only 2-3 times for each respondent. This survey design... is insufficient to estimate effects for all images, particularly when accounting for heterogeneity in preferences for images.” It “rapidly becomes infeasible in commercial conjoint studies.”
such groups presents serious impediments to accurately capturing individual-level preference (see, for example, Feit et al. 2010, Feit et al. 2013, Qian and Xie 2013).

To overcome these challenges, we propose a new methodology that disentangles form preference from overall preference and coordinates them adaptively, in real time, while allowing respondents to manipulate 3D renderings of product geometry. The method makes use of both Bayesian methods for preference measurement and machine learning tools for rapidly adapting a multidimensional design space whose consumer-oriented characteristics (e.g., boxy, retro, etc.) do not need to be determined in advance, or indeed at any point via verbal protocols. We apply the proposed “bi-level adaptive conjoint” analysis to model both form preference (for vehicles) based on underlying product geometry and overall preference by revealing tradeoffs between form (e.g., hood length, windshield pitch) and functional attributes (here, price and fuel efficiency). Results – via both Monte Carlo simulation and a crowdsourced experiment – suggest that the proposed method not only zeroes in on each consumer’s preferred visual design, but more generally, can elicit more accurate individual-level preferences than conventional conjoint alone. These benefits in turn allow, we believe for the first time, explicit measurement of trade-offs among willingness-to-pay (WTP) and design variables, e.g., whether price-sensitive consumers have marked preferences for certain styles, whether sports car buyers are less concerned about fuel economy, etc.

The paper is structured as follows: Section 2 reviews related approaches from both the marketing and engineering design literatures. Section 3 proposes the “bi-level adaptive conjoint” method, whose advantages over traditional conjoint are illustrated both by extensive simulations in Section 4 and web-based panelists in Section 5. Section 6 concludes by discussing findings and potential extensions.

2 Prior Literature on Product Design Preference Optimization

The literature on product design, both as a stand-alone field and within cognate disciplines, is vast. We therefore provide concise integrative discussions of three lines of prior research directly bearing on the subsequent development: Section 2.1 addresses methods for eliciting form preference only; Section 2.2 discusses eliciting both form preference and overall preference; and Section 2.3 reviews optimization and machine learning algorithms for preference elicitation generally and conjoint analysis specifically.
2.1 Form Preference Modeling

Form preference has been addressed in engineering design using a variety of approaches, mainly differing in terms of the parametric nature of the preference function, and secondarily in terms of how these parameters are estimated (although we will discuss estimation extensively for the real-time adaptive portion of our implementation, we will be largely agnostic regarding estimation technologies otherwise). Table 1 summarizes previous research focused on eliciting form preference via parametric models, primarily from the engineering design field.

[Table 1, “Parametric Models for Eliciting Form Preference”, about here^4]

Among the first approaches to optimizing the visual design space was Lai, Chang, and Chang (2005), who tested 2D designs for a passenger car by having three professional product designers develop appropriate initial candidates, which were then broadened using images of 125 existing passenger cars, and subsequently culled by a panel of experts to 27 combinative designs for a (parametric) Taguchi experiment. Despite its pioneering nature in calibrating form preference, their study would be of limited interest to marketers, due to its lack of parametric heterogeneity, the non-adaptive nature of its querying strategy, and to a lesser extent its reliance on 2D models and a ratings-based conjoint approach. Lugo et al. (2012) designed a wheel rim, and Reid et al. (2013) a vehicle shape, using similar methods, with a linear preference function estimated via standard regression techniques, but also adaptively at the aggregate level for 2D designs. By contrast, quadratic preferences that allowed for internal extrema were applied by Orsborn et al. (2009) to 2D shape design, using a choice-based instrument at the individual level. Sylcott et al. (2013) included interaction terms among attributes, and Kelly et al. (2011) allowed both quadratic preferences and potential interactions, although both were limited to aggregated inferences. Interaction terms are especially important for form optimization, since some consumer-valued qualities depend on multiple geometrical elements: for example, a product is only “compact” through an interaction among its dimensions.

There are several limitations in this line of engineering design research. First, as noted earlier, most has sidestepped preference heterogeneity, so that results would suggest a single “optimal” design suitable for the population as a whole. Second, with few exceptions this line of work has relied on non-adaptive (and sometimes non-choice-based) query designs, which are demonstrably outperformed by adaptive techniques (Toubia et al. 2003, Toubia et al. 2004, Abernethy et al. 2008), which we take up again in Section 2.3 and in our proposed method. Third,
nearly all prior research in the area has relied on 2D product representations, although there are exceptions that did not explicitly calibrate a form preference model (e.g., Ren and Papalambros 2011, Reid et al. 2013). In the marketing literature, Kim, Park, Bradlow, and Ding (2014) emphasized the importance of different kinds of attributes and ran conjoint studies that involved product designs for both hedonic (e.g., sunglasses) and utilitarian (e.g., coffeemakers) products, using 2D designs from a candidate list of 20 possibilities, in a manner similar to Dotson et al. (2016). To control 3D rendering parametrically is a challenge for optimization algorithms (e.g., Hsiao and Liu 2002) and conjoint interface designers, as well as for participants who may find such representations cumbersome to navigate, despite their being perceptually essential.

2.2 Form and Overall Preference Modeling

Engineers, industrial designers, and marketers often face conflicting design choices due to trade-offs between form and function attributes. For example, consumers typically want products—cell phones, clothing, automobiles, etc.—that are trim (form) yet durable (function), or sophisticated (form) yet moderately priced (function), attributes with conflicting design imperatives. There is presently little formal research to guide this trade-off. Two papers other than the present one have specifically addressed it, all centering on vehicle design; critical differences in approach and data requirements among the three studies are summarized in Table 2.

[Table 2, “Approaches to Relating Form and Function Preferences”, about here]

Dotson et al. (2016) conducted two separate surveys: a preliminary one for form preference, and a subsequent choice survey for overall preference, using a pre-selected set of 2D visual stimuli. Form preference was accommodated into the overall preference model via a novel error covariance structure specification, based on the Euclidean distance between form alternatives in the design space of physical dimensions. An attractive benefit of this formulation is that the covariance matrix can be readily updated to accommodate new (visual) alternatives simply by calculating its distance to existing alternatives. However, this relies on a presumption that (Euclidean) distance between form alternatives can serve as a measure of form preference dissimilarity, and the form preference function applied is not parametric in the sense of engineering design. Moreover, although it accomplished the proximate goal of quantifying trade-offs between form and function, the covariance matrix cannot be used directly to locate optimal designs on an individual or grouped basis. Sylcott et al. (2013) conducted three separate conjoint surveys, potentially exacerbating problems stemming from demographic differences among sampled groups: the first for form preference; the second for function preference (without price); and the
last for overall preference using two meta-attributes (e.g., form and function, each with three levels like low, medium, and high). Owing to this overall approach, there is no real way to incorporate form preference into an overall preference model, nor can the meta-conjoint analysis be linked with the form and the function preference models, so the relationship between form variable and overall preference is indeterminate. Despite the clear advances of each of these studies, they necessarily relied on estimating appropriate (vehicle) designs and overall preferences separately. Zeroing in on the individual-level sweet spot in the joint space of product designs and traditional conjoint attributes therefore wasn’t possible, even if a “best in set” design could have been located for each participant, based on the set stemming from participants in the pilot design study.

2.3 Optimization and Machine Learning Algorithms

The approach developed here leverages advances in computing power, web-based query design, machine learning, and optimization to build the underlying choice model efficiently enough to allow real-time form optimization, as opposed to simply choosing among a pre-determined set of (typically 2D) alternatives. For purposes of comparison, we summarize research according to two dimensions: estimation methods / shrinkage properties (Table 3), and adaptive query design (Table 4). Our coverage is again selective and deliberately concise. The interested reader is referred to Toubia, Evgeniou, and Hauser (2007; their Table 1 especially) for extensive background on methods; to Netzer et al. (2008) for general challenges in preference elicitation; to Toubia, Hauser, and Garcia (2007) for both simulation results and empirical comparisons among traditional and polyhedral methods; to Halme and Kallio (2011) for an overview of choice-based estimation specifically; and to Chapelle et al. (2004) for machine learning techniques and Dzyabura and Hauser (2011) specifically in reference to adaptive questionnaire design.

[Table 3, “Estimation Methods for Preference Elicitation Models”, about here]

Hierarchical Bayesian (HB) methods have long been popular for estimating partworths in conjoint, and serve to overcome sparse individual-level information by shrinkage towards the population mean (Lenk et al. 1996, Rossi and Allenby 2003). Due to stability and its near-ubiquitous use in applications (e.g., Sawtooth) we adopt HB for estimating individual-level partworths of the overall preference model. Toubia et al. (2003) proposed a polyhedral method especially well-suited to metric paired-comparison query design; observation-based constraints translate into hyperplanes, so the interior of the associated polyhedron represents the set of feasible estimates, the center of which maximizes distances to the hyperplanes. The method has
been extended to adaptive choice queries using classical and Bayesian approaches (Toubia et al. 2004; Toubia and Flores 2007). In a similar vein, Evgeniou et al. (2007) proposed a method for shrinking individual-level estimates towards population-level ones, but using a different shrinkage method (compared with HB), minimizing a convex loss function.

Cui and Curry (2005) and Evgeniou et al. (2005) extended the Support Vector Machine (SVM) – a popular machine learning algorithm used in classification problems – to conjoint applications. Specifically, Evgeniou et al. (2005) proposed an SVM mix that can accommodate parametric heterogeneity by shrinking individual-level partworths toward population means using a linear sum, lowering computational cost dramatically compared to HB at a comparable level of accuracy. We therefore adapt this method for form preference as well as adaptive design for both form and overall queries. As detailed in section 3, we couple this with a Gaussian kernel rank SVM mix to handle non-linear form preference.

Adaptive question design methods for conjoint analysis typically employ “utility balance,” wherein profiles in each choice set have similar utilities based on partworths estimated from previous answers (Toubia et al. 2007a, p. 247); the approach is comparable to “uncertainty sampling” for query strategy in the machine learning field (Settles 2010). Previous research on adaptive querying, outlined in Table 4, has demonstrated that it generally outperforms non-adaptive designs, especially when response errors are low, heterogeneity is high, and the number of queries is limited (Toubia et al. 2007a).

Polyhedral methods (e.g., Toubia et al. 2003, 2004, Toubia and Flores 2007) typically select a next query by minimizing polyhedral volume longest axis length, thereby finding the most efficient constraints (i.e., queries) to reduce the uncertainty of feasible estimates (i.e., the polyhedron). By contrast, Abernethy et al. (2008) measure uncertainty as the inverse of the Hessian of the loss function and select the next query by maximizing its smallest positive eigenvalue. Here, we use the utility balance concept as well, but extend it from an individual’s prior responses alone to incorporate those of other (previous) respondents. This adaptive query design strategy, based on a single respondent's data, may not be efficient in the early steps of sampling, due to paucity of data. We therefore use an SVM mix for adaptive query design with modest computational cost for shrinkage, which samples more efficiently despite insufficiency of individual response data, as explained in detail in section 3.
3 Proposed Model

3.1 Overview

Here we propose and develop a new method aimed at adaptively measuring the “utility” associated with both design elements and traditional product (conjoint) attributes. At the heart of the method are iterative “bi-level” questions. A bi-level question consists of two sequential sub-questions, as shown in Figure 1a: one for form alone, the other for both form and function. Before delving into specifics of implementation, we must stress that they are exactly that: details that enable the method to work quickly and reliably with real subjects. Bayesian estimation with data augmentation has provided the insight that, for example, asking for a binary choice can be viewed as a truncation or transformation mechanism from a less diametric scale, such as one with five ordinal points, or even one that is interval-scaled. The formal properties of the responses used here have been chosen to be amenable to scalability, for ease of respondent use, and due to the availability of well-vetted algorithms, and not because the method would not “work” with other scale types.

That said, for the form question, we utilize (as per Figure 1a) a standard anchored scale task. Specifically, we present two 3D vehicle renderings and ask “Which of the following styles do you prefer more?” Responses are indicated on an ordered 4-point scale: “left one is much better,” “left one is better,” “right one is better,” or “right one is much better.” Four points were used to allow a moderate degree of preference expression over a binary choice task, but without exact indifference, which would provide little ‘traction’ for the forthcoming adaptive algorithm. Next, for the purchase question, we presented the previous 3D vehicle renderings again with “functional attributes,” such as price and MPG. The respondent was then asked “Which car would you be more likely to buy?” and made a binary choice between the two presented vehicles. Such bi-level questions are repeated a specific maximum number of times, set by the analyst. The potential tendency for respondents to maintain their choice on the form question for the purchase question, irrespective of the newly supplied functional attribute information, was controlled for by counterbalancing, that is, by switching the order of the two sub-questions from round to round, as indicated in Figure 1a.5 After a choice was made by the respondent, the two designs were updated for the next round of (maximally-informative) comparisons, as shown in Figure 1b.

---

5 The actual interactive interface used for this study can be accessed at [vehiclechoiceexample.appspot.com](http://vehiclechoiceexample.appspot.com). Identifying information required by IRB has been removed for journal review.
3.2 Scoring, Utility, and Updating Algorithm

Form preference of individual $i$ stems from the form preference model:

$$ s_i = S_i(x) + \epsilon_i $$  

(1)

where $s_i$ is the form score, $S_i$ is a non-linear preference function, $x$ is a vector of design features representing the form, and $\epsilon_i$ is random error. Based on the form score, the overall preference for individual $i$ is then given by the following linear utility model:

$$ U_i(s_i, a) = \lambda_i s_i + \beta_i^T a + \epsilon_i. $$  

(2)

The vector $a$ consists of binary dummy variables for function attribute levels (i.e., a three-entry binary vector for a four-leveled attribute); $\lambda_i$ is the weight of the form score; $\beta_i$ is the partworth vector for functional attribute levels; and $\epsilon_i$ is associated error. In other words, Eq. (2) is a standard conjoint utility function with a form score ($\lambda_i s_i$) appended; it must be calibrated via Eq. (1) and weighted via $\lambda_i$, using Eq. (2).

The two preference models, Eqs. (1) and (2), were updated iteratively in real time by bi-level questioning, as per Figure 1a. The process for the odd-numbered rounds was as follows:

- **Form question:** an individual makes a metric paired-comparison between two forms created by design variables, $x^{(1)}$ and $x^{(2)}$. The preference model $S_i(x)$ in Eq. (1) is trained; then form scores, $s_i^{(1)}$ and $s_i^{(2)}$, are estimated. Finally, two function attributes (price and MPG in our application), $a^{(1)}$ and $a^{(2)}$, are sampled for the subsequent purchase question.

- **Purchase question:** an individual makes a binary choice between two bundles of form and functions $[s_i^{(1)}, a^{(1)}]$ and $[s_i^{(2)}, a^{(2)}]$. The weight of the form score $\lambda_i$ and the partworths for functions $\beta_i$ in Eq. 2 are estimated. Finally, two forms, $x^{(1)}$ and $x^{(2)}$, were sampled for the subsequent form question.

The overall process for querying, sampling, and learning is summarized visually, in flow chart form, in Figure 2, and verbally as follows:

**Start.** A new questionnaire is initialized when an individual accesses the website.

**Step 1:** Sampling form pair. A pair of vehicle renderings is generated from the design space based on the current form preference model. The pair is such that their expected form preference is roughly equal, but their shapes differ maximally from one another and from all forms used before. If this is the first question for the current subject, a fixed form pair is used.

**Step 2:** Querying form question. A metric paired-comparison response is received from the subject.

**Step 3:** Learning form preference. A form preference model is trained based on previous form responses from this subject. Once the form model is learned (or updated if not the first round), the
During each survey, we used the rank SVM mix algorithm (Evgeniou et al. 2005) for fast training and an uncertainty sampling scheme similar to Settles (2010) and Tong and Koller (2002) for real-time generation of comparison pairs. After all user data were collected, we estimated individual partworths using standard hierarchical Bayesian (HB) techniques (Lenk et al. 1996, Rossi and Allenby 2003). We emphasize again that we are agnostic on specific machine learning algorithms, which can be replaced with alternatives suited to the analyst’s specific survey environment and research goals. The rest of this section will elaborate on how these algorithms are applied in our proposed model: Section 3.2 discusses learning methods for both the form and overall preference models, and Section 3.3 on sampling methods to generate pairs for comparison.

3.2 Learning Preferences

3.2.1 Form Preference Learning

As mentioned earlier, the form preference model in Eq. (1) is trained using a rank SVM mix algorithm. The idea is to fit a model consistent with the metric paired-comparison results, i.e., if one form of the pair is much preferred to the other, the preference gap between the two should also be larger than a pair that is less differentiated, i.e., one is preferred to the other. Following the
treatments in Joachims (2002) and Chapelle and Keerthi (2010), the training problem can be formulated as follows:

\[
\min_{w} \quad w^T w \\
\text{subject to} \quad w^T \phi(x_j^{(1)}) - w^T \phi(x_j^{(2)}) \geq c_j, \forall j = 1, \ldots, m
\]  

(3)

where \( c_j \in \{1,2\} \).

Here \( x_j^{(1)} \) and \( x_j^{(2)} \) are the design features for the chosen and unchosen forms in the \( j \)-th questions, respectively. User responses are represented by \( c_j \): when \( x_j^{(1)} \) is “better” than \( x_j^{(2)} \), \( c_j \) is set to 1; when \( x_j^{(1)} \) is “much better”, \( c_j \) is set to 2. The objective in Eq. (3) follows a hard-margin SVM formulation where \( w^T w \) represents the model complexity.

One can project \( x \) to a high-dimensional space, i.e., \( x \rightarrow \phi(x) \), so that the constraints for \( \forall j = 1, \ldots, m \) can be satisfied. The dual problem of Problem 3 can be expressed as follows:

\[
\min_{\alpha} \quad \frac{1}{2} \alpha^T Q \alpha - \mathbf{c}^T \alpha \\
\text{subject to} \quad \alpha \geq 0
\]

(4)

where \( \alpha \) are Lagrangian multipliers and \( Q \) is an \( m \) by \( m \) matrix with each element, \( Q_{ij} \), being the inner product \( \langle \phi(x_i), \phi(x_j) \rangle \). Following Chang and Lin (2011), a common choice for this inner product relies on the Gaussian kernel:

\[
\langle \phi(x_i), \phi(x_j) \rangle = \exp(-\gamma \|x_i - x_j\|^2),
\]

(5)

where the Gaussian parameter \( \gamma \) is set at \( \gamma = 1/(\text{number of design features}) \). The dual problem can then be solved efficiently following the algorithm of Fan et al. (2005): based on the resultant Lagrangian multipliers, \( \alpha \), user preferences on a given form \( x \) can be quantified as

\[
S(x) = \sum_{j=1}^{m} \alpha_j (K(x, x_j^{(1)}) - K(x, x_j^{(2)})),
\]

(6)

where \( x_j^{(1)} \) are all chosen forms during the survey and \( x_j^{(2)} \) the unchosen ones. For stability and comparability, the design features \( x \) are normalized to have zero mean and unit standard deviation before being used in training and prediction. Note that a soft-margin SVM formulation (Cortes and Vapnik 1995, Cristianini and Shawe-Taylor 2000) could be used in place of Eq. (4) to deal with ‘noisy’ user responses.

Due to the limited data collection from individuals, it is desirable to leverage data collected from all participants to improve the robustness and accuracy of individual-level preference models, similar to the shrinkage underlying HB methods. As discussed earlier, we use the modest computational cost method of Evgeniou et al. (2005), with partworth of participant \( i \) given as
\[ w = \sum_j \alpha_j \left( \phi(x^{(1)}_j) - \phi(x^{(2)}_j) \right), \]  
and the population-wise partworth as
\[ \bar{w} = \frac{1}{N} \sum_{n=1}^N \sum_j \alpha^{(n)}_j \left( \phi(x^{(1,n)}_j) - \phi(x^{(2,n)}_j) \right), \]
where \( N \) is the total number of participants, \( \alpha^{(n)}_j \) is the Lagrangian multiplier for the \( j \)-th question for participant \( n \), and \( x^{(1,n)}_j \) and \( x^{(2,n)}_j \) are the chosen and unchosen forms in that question, respectively. With a weighting factor \( \eta_i \in [0,1] \), the individual form preference for participant \( i \) and a given form \( x \) can be calculated as:
\[
S^*_i(x) = (\eta_i w + (1 - \eta_i) \bar{w})^T \phi(x) \\
= \eta_i \sum_j \alpha_j \left( \langle \phi(x^{(1)}_j), \phi(x) \rangle - \langle \phi(x^{(2)}_j), \phi(x) \rangle \right) \\
+ (1 - \eta_i) \frac{1}{N} \sum_{n=1}^N \alpha^{(n)}_j \left( \langle \phi(x^{(1,n)}_j), \phi(x) \rangle - \langle \phi(x^{(2,n)}_j), \phi(x) \rangle \right) \\
= \eta_i S_i(x) + (1 - \eta_i) \frac{1}{N} \sum_{n=1}^N S_i(x)
\]

If \( \eta_i \) is small, the function of individual \( i \) shrinks strongly toward the population-level function. For active learning during the survey, \( \eta_i \) can be selected at the discretion of the analyst: when there are few prior respondents, a large value of \( \eta_i \) can be used; otherwise, a smaller \( \eta_i \) value is appropriate. Optimal \( \eta_i \) for the final estimation can be determined by cross-validation; in our experiments, we used \( \eta_1 = 1 \) and \( \eta_N = 0.7 \) for the first respondent and the last respondent, respectively, with intermediate respondents interpolated linearly between these values.

### 3.2.2 Overall Preference Learning

The coefficients in the overall preference model of Eq. (2) can be estimated analogously to those of the form preference model, that is, using a rank SVM mix algorithm for active learning during surveys and HB for population-level modeling. The problem formulation for individual-level learning is as follows:
\[
\min_{W_i} W_i^TW_i \\
\text{subject to } W_i^T x^{(1)}_{ij} - W_i^T x^{(2)}_{ij} \geq 1, \forall j = 1, \ldots, m,
\]
where \( W_i^T = [\lambda_i, \beta_i^T] \) are the linear coefficients, and \( x^{(1)}_{ij} = [s_{ij}, a_{ij}] \) consists of the form score and the binary dummy variables of the function attributes for the \( j \)-th comparison for individual \( i \). All constraints are set to be greater than or equal to one, as the comparison in this case is binary rather than metric. The dual problem of Eq. (10) is solved using a linear kernel, i.e., \( \phi(X_{ij}) = X_{ij} \), and the
resultant individual-wise partworth vector $\mathbf{W}_i$ can be expressed as:

$$\mathbf{W}_i = \sum_{j=1}^{m} (\mathbf{X}_{ij}^{(1)} - \mathbf{X}_{ij}^{(2)})^T \alpha_{ij},$$

(11)

In order to leverage population-level data, the linear shrinkage method of Evgeniou et al. (2005) is again used for individual-level partworths, $\mathbf{W}_i^*$; specifically,

$$\mathbf{W}_i^* = \eta_i \mathbf{W}_i + (1 - \eta_i) \frac{1}{N} \sum_{n=1}^{N} \mathbf{W}_n,$$

(12)

where $\eta_i$ is the weight for individual $i$, and $N$ is the number of individuals. For population-level preference modeling, a hierarchical binary logit model (Rossi et al. 2005), with weakly-informative and zero-centered priors, is used for estimation. Specifically, at the upper level of the Bayesian model, we assume $\mathbf{W}_i$ to have a multivariate normal distribution,

$$\mathbf{W}_i \sim \mathcal{N}(\mathbf{0}, \Lambda).$$

(13)

At the lower level, choice probabilities take binary logit form:

$$\Pr(y_{ij} = 1) = \frac{1}{1 + \exp[\mathbf{W}_i^T (\mathbf{X}_{ij}^{(2)} - \mathbf{X}_{ij}^{(1)})]}^{-1},$$

(14)

where $\Pr(y_{ij} = 1)$ and $\Pr(y_{ij} = 0)$ denote the probabilities of selecting $\mathbf{X}_{ij}^{(1)}$ and $\mathbf{X}_{ij}^{(2)}$, respectively, for the $j$-th question of individual $i$.

### 3.3 Sampling Questions

We adaptively sample the next pair of forms or functional attributes based on two criteria. First, to obtain utility balance (Abernethy et al. 2008) or uncertainty sampling (Settles 2010), a profile pair comprising a question should have as near to the same utility as possible according to the current model. The second criterion is to maximize the minimum distance from existing data points, from both the current participant and all previous ones. The implementation is as follows: among the pair, the first form (or function attribute set) is sampled solely by the second criterion:

$$\max_{\mathbf{x}_1^{new}} \quad \sum_{j=1}^{m_{x^{old}}} \|\mathbf{x}_1^{new} - \mathbf{x}_j^{old}\|^2$$

subject to \(lb \leq \mathbf{x}_1^{new} \leq ub\),

(15)

where $\mathbf{x}_1^{new}$ is the first form (or function attribute set) alternative for the next question, $\mathbf{x}_j^{old}$ are the $j$-th form alternatives used in previous questions, and $m_{x^{old}}$ is the number of form alternatives used previously.

Once $\mathbf{x}_1^{new}$ is sampled, its form preference value (or utility) can be calculated based on the
current model. The second sample, $x_2^{\text{new}}$, is obtained via optimizing a weighted sum of the two criteria:

$$\min_{x_2^{\text{new}}} \quad v_1 \exp \left( \| S(x_2^{\text{new}}) - S(x_1^{\text{new}}) \|^2 \right) - v_2 \left( \| x_2^{\text{new}} - x_1^{\text{new}} \|^2 + \sum_{j=1}^{m} \| x_j^{\text{old}} - x_j^{\text{old}} \|^2 \right)$$

subject to $lb \leq x_2^{\text{new}} \leq ub$, (16)

where $S(x)$ is the form preference model and $v_1$ and $v_2$ are the weights. By construction, $x_2^{\text{new}}$ should be far away from not only previous samples, but also from the current first sample; $v_1$ and $v_2$ are chosen by the experimenter to balance the two criteria. Due to high potential nonconvexity, locating each successive form pair is accomplished via genetic algorithms, by enumerating all combinations of attribute levels using Eqs. (15) and (16).

4 Application: Geometric Set-Up and Model Simulation

We demonstrate the benefits of the proposed approach by applying it to the design of a highly multiattribute, visually complex durable: a passenger sedan. As discussed in our overview of the literature, vehicle design has been among the main application domains of form design optimization models in the engineering discipline, and was proposed as a canonical application of the presentation of pictorial information in conjoint by Green and Srinivasan (1990).

To apply any method for form optimization requires a way to explore the space of designs. Both Green and Srinivasan (1990) and Dotson et al. (2016), as well as the overwhelming bulk of real-world applications in the marketing discipline, rely on a candidate set of images, which as discussed previously make interpolation or extrapolation precarious, along with parsimony challenges for heterogeneity estimation. An alternative approach common in engineering and some prior marketing research (e.g., Michalek et al. 2005) is an explicit product topology model, that is, a geometric representation of the external form and internal workings of the product. Here, our goal is less onerous, as form optimization only requires parameterizing the external (3D) shape of the product in question, not ensuring that, for example, it is possible to engineer internal components to conform with cost and safety constraints.

For vehicle form representation, we therefore developed a 3D parametric vehicle shape model (Ren and Papalambros 2011 provide additional technical detail; see as well Orsborn et al. 2009). This parametric model generates 3D renderings using two-leveled structures, as follows. First, nineteen design variables $x$ (ranging from 0 to 1; see Figure 3) were determined to be sufficient to realistically set the coordinates of control points, which in turn generate (Bezier) surfaces of the 3D model. Examples of the underlying parameterization include the distance from
the front grille to the midpoint of the hood, the elevation and pitch of the center of the windshield’s highest point, etc. The full space of potential sedan shapes would doubtless require additional parameters – for example, door and window shape are not explicitly optimized – but the 19 variables used here provide a very wide array of configurations, covering a broad swath of tested models currently in the North American sedan market, and thereby provide a reasonable trade-off between fidelity and parametric complexity / dimensionality. Figure 3 illustrates locations of all control points, some coordinates of which are determined directly by the design variables, whereas others were either fixed or adjusted automatically to maintain surface smoothness. During training, the set of 19 variable values translate to 26 control points that map to some $325 = 26 \times 25/2$ design features, each representing the distance between a pair of control points.

[Figure 3, “19 Design Variables and Their Control Points”, about here]

There is obviously a cornucopia of functional attributes important to potential car buyers. Here, we focus on two of the most critical, price and gas mileage (MPG), in terms of their trade-offs to one another and to the various form attributes embedded in the design model. In order to avoid presumptions of linearity of response, especially to price, we included five discrete attribute levels for vehicle price and MPG, selected based on sales data of so-called “CD cars” (passenger sedans) in the US. Specifically, we chose discrete points based on the 10th, 25th, 50th, 75th, and 90th percentiles of both price and MPG, as shown in Table 5.

<table>
<thead>
<tr>
<th>Level</th>
<th>Price (MSRP)</th>
<th>MPG (city/highway)</th>
<th>Percentile (market data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$23K</td>
<td>23/27</td>
<td>10th</td>
</tr>
<tr>
<td>2</td>
<td>$25K</td>
<td>23/29</td>
<td>25th</td>
</tr>
<tr>
<td>3</td>
<td>$26K</td>
<td>24/30</td>
<td>50th</td>
</tr>
<tr>
<td>4</td>
<td>$29K</td>
<td>25/31</td>
<td>75th</td>
</tr>
<tr>
<td>5</td>
<td>$31K</td>
<td>26/32</td>
<td>90th</td>
</tr>
</tbody>
</table>

Because the bi-level adaptive technique is novel, it was important to test it in theory before doing so in practice. Consequently, we first present an extensive series of simulations designed not only to demonstrate parametric recovery, but the fit and hit rate performance of each component of the bi-level querying technique.

---

6 The full parametric model and mapping of the 19 variables to design features is available from the authors.
4.1 Design for Simulation and Empirical Application

As proposed at the outset, we presume that the analyst wishes to model both form and overall preferences at the individual level, and can pose but a limited number of questions via a single-shot survey instrument. Previous research (e.g., Sylcott et al. 2013a, Dotson et al. 2016) carrying out similar analyses do not lend themselves to such one-shot form vs. function preference assessments, due to the need for time-intensive analysis between the separate form and function survey instruments. To explore the benefits of overcoming this limitation, we simulate three possible modeling options, also estimated in the forthcoming empirical application, as shown in Table 6: Model 1 is the “base,” Model 2 is the “half version” of the proposed model, and Model 3 the “full version.” The “half version” allows assessment of the bi-level question structure and adaptive design separately.

<table>
<thead>
<tr>
<th>Models</th>
<th>Querying</th>
<th>Learning</th>
<th>Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td>Single level:</td>
<td>Form and overall:</td>
<td>Non-adaptive (DOE)</td>
</tr>
<tr>
<td>(Base: single-level)</td>
<td>20 purchase questions</td>
<td>HB (linear)</td>
<td></td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td>Bi-level:</td>
<td>Form: Rank SVM mix</td>
<td>Non-adaptive (DOE)</td>
</tr>
<tr>
<td>(Half: bi-level)</td>
<td>10 form questions &amp;</td>
<td>(nonlinear)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 purchase questions</td>
<td>Overall: HB (linear)</td>
<td></td>
</tr>
<tr>
<td><strong>Model 3</strong></td>
<td>Bi-level:</td>
<td>Form: Rank SVM mix</td>
<td>Adaptive</td>
</tr>
<tr>
<td>(Full: bi-level &amp;</td>
<td>10 form questions &amp;</td>
<td>(nonlinear)</td>
<td></td>
</tr>
<tr>
<td>adaptive)</td>
<td>10 purchase questions</td>
<td>Overall: HB (linear)</td>
<td></td>
</tr>
</tbody>
</table>

To accord with typical real-world implementations, all three models are assumed to be informed by the survey responses of 100 subjects, each with a total of 20 questions, including form and purchase questions; an online pilot study suggested that 20 questions sufficed for this particular application. For validation, we used 100 hold-out questions each for form and purchase questions (i.e., 200 total) to compute hit rates. As described earlier, “form” consisted of 19 continuous design variables, whereas “function” comprised the five levels each for price and MPG.

In line with conventional (non-adaptive) conjoint analysis techniques, Model 1 ("base") used purchase questions only, and accommodated both form and function within a single linear preference model, which was estimated using standard HB methods. For DOE (design of experiments), a Latin hypercube sampling method was used to generate questionnaire designs for
both continuous and discrete variables.\textsuperscript{7}

Model 2 is the “half version” of the proposed model, and allows testing of the bi-level structure. The bi-level structure makes it possible for form and overall preference to be modeled using different specifications (i.e., a nonlinear model for form preference and a linear model for overall preference), after which the form model can be nested into the overall model.\textsuperscript{8} Form preference was estimated by a rank SVM mix (Gaussian nonlinear), overall preference using HB (linear), and a Latin hypercube sampling method was used for DOE. Compared to Model 1, Model 2 sacrifices 10 purchase questions while adding 10 separate form questions.

Model 3 is the full version of the proposed model, and tests the bi-level structure, non-linear specification, and adaptive questionnaire design effects together. The querying and learning structures are the same as Model 2, but Model 3 uses adaptive sampling so that form and function profiles are generated in real time and that each question has different form profiles (i.e., a potentially limitless number of forms).

Broadly speaking, the simulation design adapted those used widely in academic research applying conjoint methods (e.g., Arora and Huber 2001, Toubia et al. 2004, Evgeniou et al. 2005). A mainstay of previous research is that response accuracy is controlled by the magnitude of an individual’s parameters (part-worths), while respondent heterogeneity is controlled by the variance of parameters (across respondents). We operationalized accuracy and respondent heterogeneity by setting each to two levels, “low” and “high.” For example, the magnitudes of parameters were set to $\beta=0.5$ and $\beta=3$ for low and high response accuracy, respectively. On a logit scale, these represent deviations in log-odds of 0.5 and 3.0 from a baseline of zero (i.e., $\beta=0$); or, in terms of probability, according to $(1 + \exp(-\beta))^{-1}$, which translates into 0.62 and 0.95, respectively, on a probability baseline of 1/2. The parameter variances were set relative to the level of $\beta$, to $\sigma^2 = 0.5\beta$ and $\sigma^2 = 3\beta$ for low and high respondent heterogeneity, respectively. Based on these parameters, four normal distributions were defined: $\beta$ was drawn from each distribution, and then four partworth levels for each function attribute, ($-\beta, -\beta/3, \beta/3, \beta$), were generated, keeping constant differences set to $2\beta/3$.

For creating individual form preference functions, 19 continuous design variables

\textsuperscript{7} Throughout, for Latin hypercube sampling, we used the lhsdesign Matlab library; for HB, the hierarchical binary logit model in the rhierBinLogit R package (Rossi et al. 2005); and for rank SVM, we implemented the rank SVM algorithm based on the LIBSVM package (Chang and Lin 2011).

\textsuperscript{8} For the forthcoming online experiment (Section 5), subjects may be more easily able to trade-off between form and function, because they are shown a pair of vehicle forms first, followed by price and MPG with the same forms, an empirical issue not easily addressed by simulation alone.
generated a complex form preference model via main-effects ("independent") parameters $\gamma$ and "interaction" parameters $\delta$. The independent term of the $k$-th design variable, $\gamma_k$ was drawn from four pre-defined distributions (analogous to the method used for the function attributes). Specifically, for $k = 1, 2, ..., 19$, four points $(-\gamma_k/3, \gamma_k, -\gamma_k, \gamma_k/3)$ were generated, then cubic spline interpolation (denoted $\Phi(\gamma_k, x_k)$) was applied to create a continuous function with respect to the $k$-th design variable, $x_k$. We then drew $19 \times 18 / 2 = 171$ interaction terms $\delta_{ij}$, representing the relationship between the $i$-th and $j$-th design variables (for $i \neq j$). The form function, nonlinear but continuous, is therefore:

$$ S(x) = \sum_{k=1}^{19} \Phi(\gamma_k, x_k) + \sum_{i=1}^{19} \sum_{j=1}^{i-1} \delta_{ij} x_i x_j $$  \hspace{1cm} (17)

The distributions of $\delta_{ij}$ were balanced in the sense that the independent and interaction terms were set to a 2:1 ratio. [Specifically, following Evgeniou et al. (2005), we randomly generated 1000 independent terms and 1000 interaction terms, then compared the ratios of absolute values of independent and interaction terms, stopping when the standard deviation of the normal distribution for $\delta$ accorded with the 2:1 ratio.] Form score weight, $\lambda$, in Eq. (2) represents the importance of form preference, and was selected to make the ratio of absolute values of form score $s$ and function attribute preference $\beta^T \mathbf{a}$ to be 1:2 for the “low” and 2:1 for the “high” form importance cases. To do so, we generated 10000 random product profiles and 10000 consumer preference models, examined the ratio of absolute values of form score and function preference, then selected the values that allowed for the 1:2 and 2:1 ratios.

Consequently, we created eight consumer preference scenarios, as defined in Table 7. To check hit rate, we generated five sets of all eight scenarios, so that 40 total scenarios were used for the simulation.

[Table 7: "Consumer Preference Scenarios," about here]

4.2 Simulation Results

Table 8 shows the results of the various simulation scenarios, where hit rates were taken as the mean across the five sets. An asterisk (*) indicates the best, or not significantly different from best at $p < 0.05$, across the three models.

[Table 8: “Simulation Hit Rates,” about here]

Except for one case (low form importance, low response accuracy, and high respondent heterogeneity), the “full” Model 3 outperformed Model 1 for both form and overall hit rates. For the
form hit rate, every case suggests that Model 2 (bi-level structure and non-linear modeling) offers sizable improvements over Model 1 (base). Every case also shows Model 3 performing as well as or better than Model 2 (adaptive vs. non-adaptive questionnaire design), significantly outperforming Model 2 in 5 out of 8 cases. For overall hit rate, half the cases favor Model 2 over Model 1. Model 3 performed as well as or better than Model 2 in all but one case, significantly outperforming Model 2 in 3 out of 8 cases. These simulation results suggest that the proposed bi-level adaptive method (Model 3) can handily outperform the conventional one (Model 1), even with the sacrifice of 10 purchase questions. Notably from the perspective of the goals of the present study, form preference accuracy can be improved substantially (increasing to 65.7% from a base of 52.1%, or a 26% improvement on average), enabling marketers to pass along more reasonable target design values to industrial designers and engineers.

[Figure 4, “Sensitivity to the Number of Questions”, about here]

We conducted a post-analysis, testing sensitivity to the total number of questions (total number of form and purchase questions from 10 to 60 in increments of 10) on hit rate. Results appear in Figure 4, using the results of Model 1 and Model 3 with what is arguably the most difficult scenario in Table 8: high form importance, low response accuracy, and high heterogeneity. Except for the 10 questions case (i.e., 5 form and 5 purchase questions for Model 3 vs. 10 purchase questions for Model 1), Model 3 consistently outperformed Model 1 in overall preference accuracy. This owes to the fact that the form preference accuracy (for hit rate) of Model 3 was always significantly better than for Model 1, even though half of the purchase questions are sacrificed. More overall questions enabled better form preference accuracy for Model 3, whereas the performance of Model 1 did not improve substantially after 30 questions.

5 Online Experiment

The simulation spoke clearly to the advantages of bi-level adaptive querying. But, as the saying goes, what works well in theory may not do so in practice. To assess real-world performance of the bi-level adaptive technique for form preference optimization, we conducted three online surveys that correspond with the models simulated in Table 8. Three online groups were recruited through ClearVoice Research (ClearVoice 2014), a prominent online panel provider, and, to accord with the simulation scenarios, each comprised 100 subjects. Demographics were specified

---

to match with the general US adult population of car-owning households; post-analysis confirmed
the accuracy of recruitment.\textsuperscript{10} A total of 20 questions were used for learning and 10 holdout
questions (i.e., 5 form and 5 purchase holdout questions) used to check hit rates. Respondents
completed the online task of their own volition, with no time limits, on devices of their choosing.
The survey mechanism was implemented as follows. On the client side, JavaScript, WebGL and
ThreeJS were used to enable real-time 3D model rendering and interaction through mainstream
web browsers, with no additional software requirements. Critically, users were able to rotate the
real-time-generated 3D images for each presented form before deciding on their responses. On the
server side, Google App Engine was used for both executing real-time machine learning algorithms
and for data storage.

5.1 Model Performance

All three models were estimated as described in Sections 2 and 3. We examine implications
of parameter estimates after comparing relative performance quality. Hit rates for the three models
in the online experiments are shown in Table 9, with the proposed model performing best across
the board.

<table>
<thead>
<tr>
<th>Table 9: Hit Rates in Online Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Model 1 (Base: single-level)</td>
</tr>
<tr>
<td>Form preference hit rate</td>
</tr>
<tr>
<td>Overall preference hit rate</td>
</tr>
<tr>
<td>54.4%</td>
</tr>
<tr>
<td>57.2%</td>
</tr>
<tr>
<td>Model 2 (Half: bi-level)</td>
</tr>
<tr>
<td>Form preference hit rate</td>
</tr>
<tr>
<td>Overall preference hit rate</td>
</tr>
<tr>
<td>62.0% (+7.6%)</td>
</tr>
<tr>
<td>62.0% (+4.8%)</td>
</tr>
<tr>
<td>Model 3 (Full: bi-level &amp; adaptive)</td>
</tr>
<tr>
<td>Form preference hit rate</td>
</tr>
<tr>
<td>Overall preference hit rate</td>
</tr>
<tr>
<td>64.0% (+9.6%)</td>
</tr>
<tr>
<td>68.2% (+11.0%)</td>
</tr>
</tbody>
</table>

Figures in parentheses show percentage improvement over “base” Model 1

Model 2 (in Table 9) clarifies the effect of the bi-level structure, which entails substantial
improvements in prediction, an increase of 7.6% and 4.8%, for form and overall preferences hit
rates, respectively, compared to Model 1 (or 14.0% and 8.4% of their respective baselines). Model 3
(again in Table 9) shows the effect of the bi-level structure as before, but also of adaptive sampling.

\textsuperscript{10} Averages were as follows: 50.2 years age; 82% Caucasian; 81% suburban or small town; 95% high school;
69% some college; 58% working; 4% student; 15% homemaker; 23% retired; $58767 household
income; 55% married; 4.3 family size; 2.6 children; 65% spouses employed. Full cross-classified
categorical breakdowns are available from the authors.
These results further suggest that Model 3 offers an increase of 9.6% and 11.0% for form and overall preferences hit rates, respectively, compared to Model 1 (or 17.6% and 19.2% of their respective baselines) and an increase of 2.0% and 6.2% compared to Model 2 (or 3.2% and 10.0% of baseline). The overall pattern of results suggests that adaptive sampling is useful to elicit both non-linear form preferences and linear overall preferences. Specifically, the bi-level structure appears to have affected predictive accuracy for form preference more than for overall preference; and adaptive sampling affected overall preference predictions more than those for form.

Although the overarching purpose of this study is to model both form and function preferences together, within the confines of a one-shot survey, and to measure the tradeoffs among specific design variables and functional ones, we did test another model that did not incorporate form. Specifically, we removed form attributes from Model 1 to check overall preference prediction based on functional attributes alone. In Model 1a, we trained the overall preference model using only the function attributes, price and MPG, then re-checked hit rate. The results were dramatic: the hit rate increases to 64.6%, from the 57.2% of Model 1 (or 12.9% of baseline). This suggests that predicting overall preference by incorporating form design variables and function attributes within a single linear model may be suboptimal as a general approach. Model 2 in fact shows slightly poorer performance in overall preference hit rate, as it sacrifices 10 purchase questions and instead models form preference. The proposed method, Model 3, by contrast, affords significantly better prediction (68.2%) for overall preference than Model 1a (64.6%).

### 5.2 Trade-offs between Specific Form and Function Attributes

Output from the proposed model can provide several novel quantities of use to academic researchers and marketing practitioners. Two stand out: the relative importance of form and function overall, and the relative importance of various components of form and function. The distinction is roughly akin to that in ordinary conjoint between the overall impact of entire attributes, and the relative importance of various levels: it’s possible to find price critical over its entire range, but be relatively indifferent among prices well within one’s budget range.

We first examine whether there appear natural groupings – that is, a segmentation – within the data in terms of the overall “weight” placed on form and on the two functional attributes (price and MPG). Recall that, in Eq. (2), the deterministic part of utility for respondent \(i\) is given by \(\lambda_i S_i + \beta_i^T \mathbf{a}\). Price and MPG importances are calculated by the difference between the highest and lowest partworth. These values and \(\{\lambda_i\}\) are averaged across MCMC draws to compute three deterministic values for each of the 100 subjects, who can then be clustered using (first)
hierarchical and (subsequently) K-means methods according to their form ($\lambda$), price, and MPG importances. Standard metrics suggested four clusters fit the data best; both raw and standardized averages appear in Table 10.

Table 10: Clustering Based on Form, Price, MPG

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Raw Differences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form</td>
<td>4.45</td>
<td>2.92</td>
<td>5.42</td>
<td>2.26</td>
<td>5.62</td>
</tr>
<tr>
<td>Price</td>
<td>1.92</td>
<td>4.09</td>
<td>2.69</td>
<td>0.71</td>
<td>0.55</td>
</tr>
<tr>
<td>MPG</td>
<td>1.63</td>
<td>2.24</td>
<td>2.00</td>
<td>2.00</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>Standardized Differences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form</td>
<td>0</td>
<td>-0.92</td>
<td>0.58</td>
<td>-1.31</td>
<td>0.70</td>
</tr>
<tr>
<td>Price</td>
<td>0</td>
<td>1.33</td>
<td>0.47</td>
<td>-0.73</td>
<td>-0.83</td>
</tr>
<tr>
<td>MPG</td>
<td>0</td>
<td>0.67</td>
<td>0.40</td>
<td>0.40</td>
<td>-0.79</td>
</tr>
</tbody>
</table>

The four clusters can be thereby be roughly interpreted as:

**Group 1:** Price and MPG are important (relative to Form)

**Group 2:** All three (Price, MPG, Form) are valued in balance

**Group 3:** MPG is important (relative to Price especially)

**Group 4:** Form is very important (relative to Price and MPG)

Of the four groups, the fourth is far more concerned with vehicle aesthetics than the other three, while the first group doesn’t appear to value Form very strongly. In other words, willingness to pay for vehicle form is high in group 4 and low in group 1. Group 3, however, has relatively high WTP, since both Form and MPG are valued relative to Price.

These groupings are, of course, illustrative, since our parametric model does not incorporate every element of car design (e.g., door and window shape, metal vs. plastic exterior panels, etc., are not included) nor common attributes (e.g., color, audio, interior configuration). However, once the model is run and both form and function parameters are obtained, it is possible to extract much more than purportedly homogeneous segments that parcel consumers on their valuation of design overall (vs. price and MPG). Rather, the analyst can use individual-level price coefficients and weights on the underlying control points to infer which specific design elements individuals are willing to pay for. Product designers can then compare the cost of provision of those design elements – say, a more sloped profile that would provide less space for the engine compartment but could enhance aerodynamics – with WTP for that element, as well as any
available demographics that could serve as a classic discriminant function or hierarchical model covariate. We conclude our discussion of the empirical conjoint data with an analysis of this sort: which design elements are associated with relatively high WTP, and do these vary substantially across the respondent pool?

5.3 Trade-offs between Specific Form and Function Attributes

To examine the relationship between functional attributes like price and specific form attributes is somewhat more complex than the usual trade-off computations that typically follow conjoint. This is because, while most functional attributes in conjoint are “vector” type – e.g., all else equal, it is better to get higher mileage and haggle for a lower price – this is seldom the case with design attributes. For example, one might like a large and highly angled windshield, but both size and pitch are self-limiting: no matter how much one might like more of them, each eventually veers into dysfunctionality. That is, design attributes tend to be of the U-shaped ideal-point type, with a respondent-specific internal “Goldilocks” maximum. Because we have calibrated an individual-preference model, it is not difficult to calculate, for each respondent, four quantities: a maximum (what degree of that element is liked best, contingent on all the other form elements being jointly optimized), its form “score” (as per Eqs. 1 & 2), and the associated gradient and Hessian. The latter two can be quickly computed using numerical techniques, and for each respondent; and stable quantities were obtained for all four design variables.

Sensitivity to Design Variables. The diagonal components of the Hessian – which will all be negative for internal maxima – correspond to curvature or sensitivity: how (un)willing the respondent would be to give up one unit of that form attribute? [Recall that all form elements were normalized before optimization, rendering them comparable on a dimensionless scale.]

[Table 11: “Form vs. Function Trade-offs,” about here]

For each of the original 19 design variables (see Figure 3), summary statistics for these sensitivities appear in Table 11, where it is apparent that there is a wide variance in such sensitivity across both design attributes and people. For example, design variable x8 – the elevation of the central point where the back windshield meets the roof – appears to be the least sensitive design element: its median value was -0.169 (tabled value = -0.2), compared to the average for all design variables of -17.2. In simple terms, respondents were, on average, 100 times less sensitive to this design parameter than the others. However, sensitivity to the x8 design variable was quite heterogeneous, as evidenced by its mean value of -36.8; evidently, some respondents are extremely sensitive to changes in the location of this control point, judging by the skewness of the distribution
of sensitivities (i.e., half of respondents are below 0.5% of the mean value). By contrast, six design attributes stand out in terms of high sensitivity to change, listed with their medians and means (see Figure 3): x1 (-33.2, -48.3) and x2 (-36.9, -61.6), the horizontal and vertical position of the midpoint the hood/windshield join; x9 (-36.3, -52.5) and x10 (-38.3, -141.2), the horizontal and vertical position of the midpoint of the hood front; x14 (-34.4, -49.7), the lateral displacement of the hood/windshield join point directly in front of the driver; and x19 (-32.8, -53.1), the outward displacement of the driver’s side hood front. Because distance between mean and median provides a measure of skewness – that is, heterogeneity across respondents in sensitivity – it is clear that x10 (which affects not only shape but also driver visibility) is both important to most people (high median) as well as exceedingly important to some (even higher mean).

Trade-offs Against Design Variables. However, examining sensitivities alone is merely suggestive: perhaps the consumers who are most sensitive to specific elements of design also have the lowest overall value for design, relative to price (or MPG). It is even possible that nearly all respondents, while responding to design changes in and of themselves, have no trade-off against “functional” attributes, especially price. As such, we wish to construct metrics for “Willingness To Trade-off” (WTT) for each design attribute vs. the two functional attributes (price and mileage) in our study. This is made more complex by our having allowed for nonlinear response to the five levels of both price and MPG. So, for simplicity of presentation, we compute two values, akin to the interquartile range, for each consumer’s utility function: the difference in the 25th and 75th percentile values. That is, the partworth differences for $29K vs. $25K and 23/29 MPG vs. 25/31; or, more simply $4000 and 2MPG, which can then be standardized into willingness-to-trade-off $1000 and 1MPG (by dividing by 4 and 2, respectively). Finally, the deterministic part of the utility model is given as \( \lambda s_t + \beta_t^T a \), so that we can unambiguously answer “what % change in each design variable within \( s_t \) translates into maintaining utility if price were either $1000 higher (WTT Price, WTTP) or mileage 1 MPG better (WTT Mileage, WTTM)?” For each consumer, this entailed both the “form” model for \( s_t \) and the estimated value of \( \lambda_t \), which measures overall design importance.

As a check, we first computed a typical quantity in conjoint studies, the WTP for MPG, which is a standard trade-off between conjoint attributes. The literature reports a wide range of values, depending on which sort of cars are studied, the method of task (e.g., conjoint, purchase data, or field experiment), demographic composition of the respondent pool, and, critically, the range of prices and MPG values studied. Greene’s (2010) review of this literature highlights findings from Gramlich (2010), who found that WTP in $/mile (calculated based on an increase from 25 to 30
MPG and a gas price of $2/gal.), was approximately $800 for luxury cars, $1480 for compact cars, and $2300 for SUVs; but that this rose $1430, $2580, and $4100, respectively, for a gas price of $3.50/gal (in 2008 prices). In our study, for mid-priced sedans, median WTP in $/mile was $2410, well within the latter reported range, lending some degree of external validity to our results.

Our goal was to quantify trade-offs between our 19 design variables and both price and MPG. To do so, we computed median and mean values for willingness to trade off a .01 deviation (from optimum) in each design variable against price (WTTP, in $1000) and mileage (WTTM, in 1 MPG); these were multiplied by 100 to reflect the relative size of the .01 deviation to the normalized unit scaling for each of the 19 design variables. Results appear in the last four columns of Table 11. Because the trade-off between price and MPG was $2410, we would expect the values for MPG to be about 40% of those for price (empirical values based on the means in the last row of Table 11 were calculated as 35.6% and 32.0% for medians and means, respectively). [We refer henceforth to median values, since the distribution across consumers is quite skewed for some design elements; .01 deviations from design optimum are given by 1/100th the values in the table.]

It is apparent that some design variables were far more valued than others, roughly tracking with the results for the Hessian (columns 2 and 3 of Table 11). For example, as before, x1 and x2 (horizontal and vertical position of the hood/windshield join midpoint) were each valued by approximately $50 for each .01 deviation from optimum (tabled values, 5.36 and 4.86, respectively, in $1000 units/100), and would correspond with mileage losses of .0152 and .0176 MPG each. These two (x1, x2) were not alone, as several of the design variables showed similar substantial sensitivity, with .01 changes corresponding to approximately $50 in price (e.g., x9, x10, x14, x19).

Holding aside potential interaction effects between the design variables accounted for in the form model, and working off just the main effects of Table 11, one can tally up the median values (i.e., summing columns 4 and 6, then dividing by 100) to calculate an “omnibus” value for design changes overall, based on .01 deviations from each consumer’s optimum design along all 19 dimensions. Doing so translated into approximately $439 in WTTP and .157 miles in WTTM, both substantial values, given the rather small 1% deviations. Of course, the ‘demand’ model implied by both the underlying choice model and form submodel is highly nonlinear, so one must be careful in interpreting such “local” results and extrapolating them to the entire design space, where whole regions are likely to show little slope due to their being non-viable for particular consumers (i.e., designs they actively dislike). Furthermore, even with substantial heterogeneity, the

---

11 Specifically, in this case, Δ(partworths of $25K and $29K)/Δ(partworths of 25/31MPG and 23/29MPG), rendered in $1/mile. Sonnier, Ainslie & Otter (2007) provide additional detail on such calculations.
availability of individual-level results means that – given the cost of production of different design elements and demographics for respondents – a manufacturer could roughly compute whether a certain kind of consumer would be willing to pay the added cost of a proposed design alteration, or whether it might be worth reduced fuel efficiency or trade-off against any traditional (functional) conjoint attribute.

6 Conclusion and Future Directions

Preference elicitation is among the great success stories of the application of experimental and statistical methodology to address perpetual, central problems in marketing. As evidenced by widespread adoption throughout the world over the last four decades (Sattler and Hartmann 2008), conjoint methods in particular can currently be deployed, using web-based tools, by practicing managers, with a low upfront burden in selecting optimal stimuli sets and backend estimation technologies. For example, as of the time of writing, Sawtooth’s Discover allows product designers to specify attributes and levels, with subsequent “heavy lifting” – fashioning orthogonal designs, choice-based stimuli sets, and Bayesian estimation – handled seamlessly in the background. Yet even the best current implementations of conjoint founder on the shoals of visual design: while adjectival labels (e.g., sporty, bold, posh, etc.) and pre-generated 2D imagery can easily be included as categorical stimuli and covariates, and undoubtedly help directionally identify “what consumers are looking for,” they neither allow consumers to converge on specific designs that hold particular appeal for them nor designers to focus solely on the design space, rather than pre-rendered descriptions or depictions of that space.

This paper proposes what we believe to be the first comprehensive approach to the visual design problem, one that leverages both the sort of product topology modeling common in engineering and rapid, scalable machine learning algorithms to interweave with state-of-the-art preference elicitation methods developed in marketing. The resulting hybrid, using bi-level adaptive techniques and manipulable, real-time rendered imagery, can be deployed using standard web-based protocols to zero in on each consumer’s preferred product design along with the sorts of attributes used traditionally in conjoint. The approach eschews descriptions or pre-set depictions of any sort, allowing post-hoc processing of individual-level data to determine trade-offs between common attributes like product price and visual design elements, as well as against design overall.

Our empirical analysis focuses deliberately on automotive design, since this is among the most complex durables that a consumer ever purchases: it is high involvement, requires many trade-offs, and choices are highly deliberative. And it is among the most design-intensive of all
products. To take but one ill-fated example, Ford lavished over $6Bn and several years on the
design of its Mondeo "world car." Product topology models are intensively devised and
scrutinized for such products, making them particularly amenable to being geometrized in the
manner deployed here (although we again stress that doing so is far less complex than for full-scale
product design, which requires harmonizing external visual shape with internal component
configuration; Michalek et al. 2005). Because the overwhelming majority of widely-deployed
durable products involve computer-aided design (CAD), engineers can readily provide
low-to-moderate dimensional product form representations for use in generating real-time 3D
models for use within the method, with scant additional input from or mediation by marketers or
econometricians.

Despite providing means to overcome practical limitations on the number of questions
asked of respondents, the 2D nature of prior product representations, and the ability to connect
form and function attributes, other limitations and their attendant challenges remain. First, it is
often difficult to cover the entirety of the design space with parametric design variables. Although
this can be (and has been) accomplished for the space of passenger sedans, what if, for a highly
heterogeneous consumer pool, a designer wished to include all manner of automobiles, or all
manner of personal transportation? Such design spaces are not only impractically large, they
contain never-before-seen hybrids that might be imaginative and desirable, but impossible or
impossibly costly to produce. Methods to somehow account for the feasibility of generated designs
would help further constrain the space of generated surfaces to those that a firm might reasonably
pursue; there is no point having a consumer convey her most preferred design, only for the firm to
reply that she cannot have it. Because the machine learning algorithms here can be tuned to
account for constrained spaces, a designer could, in principle, designate the subspace of “buildable”
configurations, and candidate 3D images would remain within its boundaries.

Another consequence of the curse-of-dimensionality is that active learning with respect to a
large number of design variables and all their interactions entails ever greater computational costs.
Here, including two-way interaction effects greatly enhanced predictive performance, but there’s
no reason to believe that interactions stop there (Dzyabura and Hauser 2011). For example, an
attribute as simple as “large capacity” depends on an interaction of three spatial dimensions, and
there are doubtless customers who especially value a car that is sporty and sleek and mid-priced
and offers superb acceleration, etc. Because we use the responses of former subjects for active

---

learning, the computational burden grows with many subjects, many attributes, many levels, and many interactions. This suggests that the selection of machine learning methods may need to be judicious and individually tailored for large-scale applications to complex durables. Fortunately, the proposed bi-level adaptive structure can be fruitfully deployed in conjunction with a variety of machine learning algorithms, and its flexibility and modularity should allow ready deployment in a multitude of visual product design contexts.
References


Feit, Eleanor McDonnell, Pengyuan Wang, Eric T Bradlow, Peter S Fader. 2013. Fusing aggregate and disaggregate data with an application to multiplatform media consumption. Journal of
Marketing Research 50 (3) 348-364.


Green, Paul E, Abba M Krieger, Yoram Wind. 2001. Thirty years of conjoint analysis: Reflections and prospects. Interfaces, 31 (3) S56-S73.


Hsiao, S.W., Liu, M.C., 2002. A morphing method for shape generation and image prediction in product design. Design studies, 23 (6) 533-556.


Lenk, Peter J, Wayne S Desarbo, Paul E Green, Martin R Young. 1996. Hierarchical bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs. Marketing


### Table 1: Parametric Models for Eliciting Form Preference

<table>
<thead>
<tr>
<th>Research</th>
<th>Parametric preference function</th>
<th>Parameter estimation</th>
<th>Heterogeneity</th>
<th>Survey</th>
<th>Query design</th>
<th>Product</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lai et al. (2005)</td>
<td>S/N ratio</td>
<td>Taguchi</td>
<td>Aggregate</td>
<td>Rating</td>
<td>Non-adaptive</td>
<td>Vehicle</td>
<td>2D</td>
</tr>
<tr>
<td>Orsborn et al. (2009)</td>
<td>Quadratic</td>
<td>BTL</td>
<td>Individual</td>
<td>Choice</td>
<td>Non-adaptive</td>
<td>Vehicle</td>
<td>2D</td>
</tr>
<tr>
<td>Kelly et al. (2011)</td>
<td>Quadratic w/ interaction term</td>
<td>PREFMAP</td>
<td>Aggregate</td>
<td>Rating</td>
<td>Non-adaptive</td>
<td>Vehicle</td>
<td>2D</td>
</tr>
<tr>
<td>Lugo et al. (2012)</td>
<td>Linear</td>
<td>Regression</td>
<td>Aggregate</td>
<td>Rating</td>
<td>Non-adaptive</td>
<td>Wheel</td>
<td>rim</td>
</tr>
<tr>
<td>Reid et al. (2012)</td>
<td>Linear</td>
<td>Regression</td>
<td>Aggregate</td>
<td>Rating</td>
<td>Non-adaptive</td>
<td>Vehicle</td>
<td>2D</td>
</tr>
<tr>
<td>Tseng et al. (2012)</td>
<td>Linear</td>
<td>Regression</td>
<td>Aggregate</td>
<td>Rating</td>
<td>Non-adaptive</td>
<td>Vehicle</td>
<td>2D</td>
</tr>
<tr>
<td>Reid et al. (2013)</td>
<td>[Linear]</td>
<td>BTL</td>
<td>Aggregate</td>
<td>Choice</td>
<td>Non-adaptive</td>
<td>Vehicle &amp; carafe</td>
<td>3D</td>
</tr>
<tr>
<td>Sylcott et al. (2013)</td>
<td>Linear with interaction term</td>
<td>MNL</td>
<td>Aggregate</td>
<td>Choice</td>
<td>Non-adaptive</td>
<td>Vase &amp;</td>
<td>vehicle</td>
</tr>
</tbody>
</table>

### Table 2: Approaches to Relating Form and Function Preferences

<table>
<thead>
<tr>
<th>Survey</th>
<th>Dotson et al. (2016)</th>
<th>Sylcott et al. (2013a)</th>
<th>This study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two separate surveys</td>
<td>Three separate surveys</td>
<td>Bi-level questions in single survey</td>
<td></td>
</tr>
<tr>
<td>(1) form: rating</td>
<td>(1) form: choice</td>
<td>(1) form: metric paired-comparison</td>
<td></td>
</tr>
<tr>
<td>(2) overall: choice</td>
<td>(2) function: choice</td>
<td>(2) overall: choice</td>
<td></td>
</tr>
<tr>
<td>(3) overall: pairwise comparison</td>
<td>(3) overall: pairwise comparison</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time delay between surveys</td>
<td>Yes</td>
<td>Yes</td>
<td>No (real time)</td>
</tr>
<tr>
<td>Query design</td>
<td>Non-adaptive</td>
<td>Non-adaptive</td>
<td>Adaptive</td>
</tr>
<tr>
<td>Preference function</td>
<td>Form: covariance structure</td>
<td>Form: quadratic function</td>
<td>Form: radial basis Overall: linear</td>
</tr>
<tr>
<td>Overall: linear</td>
<td>Overall: linear</td>
<td>Overall: linear</td>
<td></td>
</tr>
<tr>
<td>Estimation</td>
<td>Form: Euclidian distance</td>
<td>Bradley-Terry-Luce (BTL)</td>
<td>Form: Rank SVM mix Overall: Hierarchical Bayesian</td>
</tr>
<tr>
<td>Overall: Bayesian</td>
<td>Overall: Bayesian</td>
<td>Overall: Bayesian</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>Individual</td>
<td>Individual</td>
<td>Individual</td>
</tr>
<tr>
<td>Product</td>
<td>Vehicle</td>
<td>Vehicle</td>
<td>Vehicle</td>
</tr>
</tbody>
</table>
### Table 3: Estimation Methods for Preference Elicitation Models

<table>
<thead>
<tr>
<th>Research</th>
<th>Method</th>
<th>Shrinkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenk et al. (1996)</td>
<td>Hierarchical Bayes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rossi and Allenby (2003)</td>
<td>Metric paired-comparison analytic-center estimation</td>
<td>No</td>
</tr>
<tr>
<td>Toubia et al. (2003)</td>
<td>Metric paired-comparison analytic-center estimation</td>
<td>No</td>
</tr>
<tr>
<td>Cui and Curry (2005)</td>
<td>Support Vector Machine (SVM)</td>
<td>No</td>
</tr>
<tr>
<td>Evgeniou et al. (2005)</td>
<td>SVM mix</td>
<td>Yes</td>
</tr>
<tr>
<td>Toubia and Flores (2007)</td>
<td>Adaptive choice-based analytic-center estimation</td>
<td>No</td>
</tr>
<tr>
<td>Evgeniou et al. (2007)</td>
<td>Heterogeneous partworth estimation with complexity control</td>
<td>Yes</td>
</tr>
<tr>
<td>This study</td>
<td>Form preference: Rank SVM mix Overall preference: Hierarchical Bayes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Table 4: Adaptive Query Design Methods

<table>
<thead>
<tr>
<th>Research</th>
<th>Method</th>
<th>Sampling</th>
<th>Data used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toubia et al. (2003)</td>
<td>Adaptive metric paired-comparison polyhedral question design</td>
<td>Minimize polyhedron volume and length of longest axis</td>
<td>Individual’s prior responses</td>
</tr>
<tr>
<td>Toubia et al. (2004)</td>
<td>Adaptive choice-based polyhedral question design</td>
<td>Minimize polyhedron volume and length of longest axis</td>
<td>Individual’s prior responses</td>
</tr>
<tr>
<td>Toubia and Flores (2007)</td>
<td>Adaptive choice-based polyhedral question design</td>
<td>Minimize polyhedron volume and length of longest axis</td>
<td>Individual’s prior responses</td>
</tr>
<tr>
<td>Abernethy et al. (2008)</td>
<td>Hessian-based adaptive choice-based conjoint analysis</td>
<td>Maximize smallest positive eigenvalue of loss function Hessian</td>
<td>Individual’s prior responses</td>
</tr>
<tr>
<td>This study</td>
<td>Adaptive metric paired-comparison SVM mix question design</td>
<td>Minimize difference between utilities of new pairs and maximize Euclidean distance among all profiles</td>
<td>Both individual’s and others’ prior responses</td>
</tr>
</tbody>
</table>
### Table 7: Consumer Preference Scenarios

<table>
<thead>
<tr>
<th>Form importance</th>
<th>Response accuracy</th>
<th>Respondent heterogeneity</th>
<th>Form score weight (λ)</th>
<th>Form attribute coefficients</th>
<th>Functional attribute partworths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>0.0043</td>
<td>N(0.5, 0.25) N(0, 4.80) N(0.5, 0.25)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>0.0044</td>
<td>N(0.5, 1.5) N(0, 13.7) N(0.5, 1.5)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>0.0028</td>
<td>N(3.0, 1.5) N(0, 56.3) N(3, 1.5)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>0.0057</td>
<td>N(3.0, 9.0) N(0, 88.4) N(3, 9.0)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>0.0173</td>
<td>N(0.5, 0.25) N(0, 4.80) N(0.5, 0.25)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>0.0176</td>
<td>N(0.5, 1.5) N(0, 13.7) N(0.5, 1.5)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>0.0112</td>
<td>N(3.0, 1.5) N(0, 56.3) N(3, 1.5)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>High</td>
<td>0.0230</td>
<td>N(3.0, 9.0) N(0, 88.4) N(3, 9.0)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 8: Simulation Hit Rates

<table>
<thead>
<tr>
<th>Simulation design</th>
<th>Form preference hit rate</th>
<th>Overall preference hit rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form importance</td>
<td>Response accuracy</td>
<td>Respondent heterogeneity</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

*Best, or not significantly different from best at p<0.05, across all models.
Table 11: Form vs. Function Trade-Offs

<table>
<thead>
<tr>
<th>Control Points</th>
<th>Hessian at Max</th>
<th>WTTP: Interquartile Price</th>
<th>WTTM: Interquartile MPG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>x1</td>
<td>-33.2</td>
<td>-48.3</td>
<td>5.36</td>
</tr>
<tr>
<td>x2</td>
<td>-36.9</td>
<td>-61.6</td>
<td>4.86</td>
</tr>
<tr>
<td>x3</td>
<td>-12.4</td>
<td>-27.5</td>
<td>2.10</td>
</tr>
<tr>
<td>x4</td>
<td>-6.1</td>
<td>-27.3</td>
<td>0.86</td>
</tr>
<tr>
<td>x5</td>
<td>-2.5</td>
<td>-13.4</td>
<td>0.36</td>
</tr>
<tr>
<td>x6</td>
<td>-3.6</td>
<td>-165.9</td>
<td>0.67</td>
</tr>
<tr>
<td>x7</td>
<td>-6.4</td>
<td>-9.3</td>
<td>0.80</td>
</tr>
<tr>
<td>x8</td>
<td>-0.2</td>
<td>-36.8</td>
<td>0.02</td>
</tr>
<tr>
<td>x9</td>
<td>-36.3</td>
<td>-52.5</td>
<td>4.20</td>
</tr>
<tr>
<td>x10</td>
<td>-38.3</td>
<td>-141.2</td>
<td>5.30</td>
</tr>
<tr>
<td>x11</td>
<td>-1.9</td>
<td>-3.4</td>
<td>0.17</td>
</tr>
<tr>
<td>x12</td>
<td>-11.3</td>
<td>-12.0</td>
<td>1.41</td>
</tr>
<tr>
<td>x13</td>
<td>-18.2</td>
<td>-27.8</td>
<td>1.94</td>
</tr>
<tr>
<td>x14</td>
<td>-34.4</td>
<td>-49.7</td>
<td>4.92</td>
</tr>
<tr>
<td>x15</td>
<td>-10.3</td>
<td>-26.2</td>
<td>0.84</td>
</tr>
<tr>
<td>x16</td>
<td>-17.7</td>
<td>-26.4</td>
<td>2.25</td>
</tr>
<tr>
<td>x17</td>
<td>-6.2</td>
<td>-19.4</td>
<td>0.80</td>
</tr>
<tr>
<td>x18</td>
<td>-17.8</td>
<td>-45.9</td>
<td>2.64</td>
</tr>
<tr>
<td>x19</td>
<td>-32.8</td>
<td>-53.1</td>
<td>4.45</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>-17.2</td>
<td>-44.6</td>
<td>2.31</td>
</tr>
</tbody>
</table>

Reported values for Willing To Trade-off Price (WTTP) and Willing To Trade-off Mileage (WTTM) refer to units of $1000 or miles, respectively, for a .01 change in the design variable, multiplied by 100. These thus provide a “local” approximation of the full range of the design variables, each of which was normalized to lie on a unit scale.
Figure 1: Iterative bi-level Queries and Design Changes

Figure 1a: Iterative bi-level questions

Figure 1b: Design Query Update Resulting from Prior Round Choice

Note: The “Round 2” designs result from answers to the “Round 1” bi-level queries; that is, both sub-questions of Round 1 lead to the design pair used in Round 2, which in turn lead to those generated for Round 3, etc.
Figure 2: Overall process for querying, sampling, and learning
Figure 3: Nineteen design variables and Their Control Points

Legend: Control Points

\(x_1, x_2\): forward (X) and vertical (Z) displacement of windshield / hood mid-join point

\(x_3, x_4\): forward (X) and vertical (Z) displacement of windshield / roof mid-join point

\(x_5, x_6\): forward (X) and vertical (Z) displacement of windshield curvature determination point

\(x_7, x_8\): forward (X) and vertical (Z) displacement of hood / rear-window mid-join point

\(x_9, x_{10}\): forward (X) and vertical (Z) displacement of hood front midpoint

\(x_{11}, x_{12}\): forward (X) and vertical (Z) displacement of hood centroid point

\(x_{13}, x_{14}\): forward (X) and lateral (Y) displacement of windshield / hood driver centerpoint

\(x_{15}\): vertical (Z) displacement of external hood edge curvature point

\(x_{16}, x_{17}\): forward (X) and lateral (Y) displacement of front bumper lowest edge point

\(x_{18}, x_{19}\): forward (X) and lateral (Y) displacement of hood / front bumper join point
Figure 4: Hit Rate Sensitivity to Total Number of Questions

Overall, Model 3
Overall, Model 1
Form, Model 3
Form, Model 1