APPENDICES: Technical and Online

MKSC-16-0281.R1

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

APPENDIX A: Superiority of Bi-Level Query Method

Online Only:

APPENDIX B: Accommodating the "No Choice" Option

APPENDIX C: Lagrangian Formulation of Eq. (3)

APPENDIX D: Insensitivity of Individual-Level Learning to Constraint Value

APPENDIX E: Setting Weights (v_1 and v_2) in Eq. 16

APPENDIX F: Simultaneous vs. Sequential Optimization for Eq. (15) and (16)

APPENDIX G: Robustness Checks for Preference Gap Cutoffs, c_j, in Eq. (3)

APPENDIX H: Details on Query Engine and Response Times

APPENDIX A: Superiority of Bi-Level Query Method

A1 Introduction

To provide insights on why the bi-level questionnaire works better, we introduce simplified data models for the single- and bi-level cases so that generalization error can be computed.

Specifically, we consider two linear models: The *form* model maps form features $x \in \mathbb{R}^{d_x}$ to the form score $s \in \mathbb{R}$, with parameters β and i.i.d. random error $\varepsilon_s \sim N(0, \sigma_s^2)$:

$$s = \beta^T x + \varepsilon_s. \tag{A1}$$

The *preference* model maps the form score s and product attributes $u \in \mathbb{R}^{d_u}$ to the utility $y \in \mathbb{R}$, with parameters b_s for the form score and b_u for the product attributes and i.i.d. random error $\varepsilon_y \sim N(0, \sigma_y^2)$:

$$y = b_s s + b_u^T x + \varepsilon_y. \tag{A2}$$

In the following, we discuss two cases corresponding to the two types of questionnaires studied in this paper: In the single-level questionnaire, we directly collect answers from purchase questions; and in the bi-level questionnaire, we collect answers for both form and purchase questions. For both cases, our goal is to build a predictive model so that for new inputs in the form of $\langle x, u \rangle$, we can predict their utility. We investigate the generalization errors in utility prediction using these two types of experiments.

As an overview, our insights are as follows:

- **When the form signal-to-noise ratio $||\boldsymbol{\beta}||_1/\boldsymbol{\sigma_s}$ is sufficiently large, and the form-to-utility noise ratio $\boldsymbol{\sigma_s^2/\sigma_y^2}$ is sufficiently small, the bi-level questionnaire is a superior choice.**
- **In the extreme case where form responses are noiseless ($\boldsymbol{\sigma_s^2 = 0}$), the bi-level questionnaire is always better.**

A2 Case 1: Single-phase Questionnaire

A2.1 Data

In the first case, we consider the existence of data $\mathcal{D} = \{y^i, x^i, u^i\}_{i=1}^{N_1}$, i.e., we can observe the actual utilities, but not form scores. This is a simplification from the actual experiments, where only comparisons between ys rather than their actual values are observable.

We also make the following assumption to smooth the analysis: Let $X \in \mathbb{R}^{N_1 \times d_x}$ and $U \in \mathbb{R}^{N_1 \times d_u}$ be the data matrices where each row of X (U) is a data point $(x^i)^T$ ($(u^i)^T$). Denote $x_{\cdot j}$ ($u_{\cdot j}$) as the jth column of X (U). We assume that

- (**R1**) the matrix $D = [X, U]$ is column-wise orthogonal;
- (**R2**) $1^T x_{\cdot i} = 0$ for all i;

- (**R3**) $x_{\cdot i}^T x_{\cdot i} = N_1$ for all $i \neq j$.

(R1) can be achieved through design-of-experiments when $N_1 \geq d_x + d_u$, and is plausible when N_1 is large and x and u are independently sampled in each dimension. (R2) and (R3) can always be achieved by preprocessing the data to have zero means and unit variance column-wise. To further facilitate some of the detailed proofs, we also assume that each dimension of x and u is bounded (through truncated distributions).

A2.2 Prediction and uncertainty

For a specific dataset $\mathcal{D}$, we can estimate parameters for the following model through ordinary least square (OLS):

$$y = b_s \beta^T x + b_u^T u + \varepsilon_y + b_s \varepsilon_s. \tag{A3}$$

The model can be further simplified as

$$y = b^T w + \varepsilon, \tag{A4}$$

with $b^T := [b_s \beta^T, b_u^T]$, $w^T = [x^T, u^T]$, and the i.i.d. random variable $\varepsilon \sim N(0, \sigma_y^2 + b_s^2 \sigma_s^2)$.

Denote $\hat{b}_{\mathcal{D}}$ as the OLS estimate of b derived from data $\mathcal{D}$. We have $\mathbb{E}[\hat{b}_{\mathcal{D}}] = b$, and the variance-covariance matrix of $\hat{b}_{\mathcal{D}}$ (due to the noise in observations of y) is:

$$Var(\hat{b}_{\mathcal{D}}) = (\sigma_y^2 + b_s^2 \sigma_s^2)(D^T D)^{-1}. \tag{A5}$$

A2.3 Generalization error

The generalization error of $\hat{b}_{\mathcal{D}}$ is defined as

$$\begin{aligned} L(\hat{b}_{\mathcal{D}}) &= \mathbb{E}_w[(\hat{b}_{\mathcal{D}}^T w - b^T w)^2] \\ &= \int_w ((\hat{b}_{\mathcal{D}} - b)^T w)^2 p(w) dw \\ &= \int_w \left(\sum_{i=1}^{d_x+d_u} \Delta\hat{b}_{\mathcal{D},i} w_i\right)^2 p(w) dw \\ &= \int_w \left(\sum_{i=1}^{d_x+d_u} (\Delta\hat{b}_{\mathcal{D},i} w_i)^2 + \sum_{i=1}^{d_x+d_u} \sum_{j=i+1}^{d_x+d_u} (\Delta\hat{b}_{\mathcal{D},i} \Delta\hat{b}_{\mathcal{D},j} w_i w_j)\right) p(w) dw \\ &= \int_w \sum_{i=1}^{d_x+d_u} (\Delta\hat{b}_{\mathcal{D},i} w_i)^2 p(w) dw + \int_w \sum_{i=1}^{d_x+d_u} \sum_{j=i+1}^{d_x+d_u} (\Delta\hat{b}_{\mathcal{D},i} \Delta\hat{b}_{\mathcal{D},j} w_i w_j) p(w) dw \\ &= \sum_{i=1}^{d_x+d_u} \int_{w_i} (\Delta\hat{b}_{\mathcal{D},i} w_i)^2 p(w_i) dw_i + \sum_{i=1}^{d_x+d_u} \sum_{j=i+1}^{d_x+d_u} \int_{w_i,w_j} (\Delta\hat{b}_{\mathcal{D},i} \Delta\hat{b}_{\mathcal{D},j} w_i w_j) p(w_i, w_j) dw_i dw_j \end{aligned} \tag{A6}$$

where $\Delta\hat{b}_{\mathcal{D},i}$ is the ith element of $\hat{b}_{\mathcal{D}} - b$. We assume that each element of the test data point w_i is drawn i.i.d. with $\mathbb{E}[w_i] = 0$ and $Var(w_i) = 1$ (again, this can be done through proprocessing of the data space). Therefore, we have

$$\int_{w_i,w_j} w_i w_j p(w_i, w_j) dw_i dw_j = 0, \tag{A7}$$

and

$$L(\hat{b}_{\mathcal{D}}) = \sum_{i=1}^{d_x+d_u} \Delta\hat{b}_{\mathcal{D},i}^2. \tag{A8}$$

Notice that $L(\hat{b}_{\mathcal{D}})$ is *model-specific* and $\hat{b}_{\mathcal{D}}$ is random (Eq. (A5)). Taking expectation over $\hat{b}_{\mathcal{D}}$ leads to the *data-specific* generalization error

$$\begin{aligned}\mathbb{E}_{\hat{b}_{\mathcal{D}}}[L(\hat{b}_{\mathcal{D}})] &= \int_{\hat{b}_{\mathcal{D}}} \sum_{i=1}^{d_x+d_u} \Delta\hat{b}_{\mathcal{D},i}^2 p(\hat{b}_{\mathcal{D}}) d\hat{b}_{\mathcal{D}} \\ &= \sum_{i=1}^{d_x+d_u} \int_{\hat{b}_{\mathcal{D}}} \Delta\hat{b}_{\mathcal{D},i}^2 p(\hat{b}_{\mathcal{D},i}) p(\hat{b}_{\mathcal{D},-i}|\hat{b}_{\mathcal{D},i}) d\hat{b}_{\mathcal{D}} \\ &= \sum_{i=1}^{d_x+d_u} \int_{\hat{b}_{\mathcal{D},i}} \Delta\hat{b}_{\mathcal{D},i}^2 p(\hat{b}_{\mathcal{D},i}) d\hat{b}_{\mathcal{D},i} \\ &= \sum_{i=1}^{d_x+d_u} Var(\Delta\hat{b}_{\mathcal{D},i}) \\ &= (\sigma_y^2 + b_s^2\sigma_s^2) tr((D^T D)^{-1}).\end{aligned} \tag{A9}$$

Here $\hat{b}_{\mathcal{D},-i}$ are elements of $\hat{b}_{\mathcal{D}}$ other than the ith element. The derivation used the fact that

$$\int_{\hat{b}_{\mathcal{D},-i}} p(\hat{b}_{\mathcal{D},-i}|\hat{b}_{\mathcal{D},i}) d\hat{b}_{\mathcal{D},-i} = 1, \tag{A10}$$

and that $\mathbb{E}[\Delta\hat{b}_{\mathcal{D},i}] = 0$.

Finally, to derive the generalization error for the questionnaire (and the use of OLS), we need to compute

$$\mathbb{E}_{\mathcal{D},\varepsilon_y,w}[L(\hat{b}_{\mathcal{D}})] = \mathbb{E}_{\mathcal{D}}[(\sigma_y^2 + b_s^2\sigma_s^2) tr((D^T D)^{-1})].$$

Notice that from (R1) we have $D^T D = N_1 I \in \mathbb{R}^{(d_x+d_u)\times(d_x+d_u)}$. This leads to the solution

$$L_1 := \mathbb{E}_{\mathcal{D},\varepsilon_y,w}[L(\hat{b}_{\mathcal{D}})] = (\sigma_y^2 + b_s^2\sigma_s^2) \frac{d_x+d_u}{N_1}. \tag{A11}$$

The solution in Eq. (A11) is consistent with intuition: (1) Increasing noise in response (σ_y^2 and σ_s^2) increases the generalization error. In particular, the part-worth of the form score scales the influence of σ_s^2; and (2) The dimensionalities of s and u, and the sample size, influence the generalization error in opposite directions.

A3 Case 2: Bi-phase Questionnaire

A3.1 Data

To analyze the second case, we consider the existence of data $\mathcal{D} = \{y^i, x^i, u^i, s^i\}_{i=1}^{N_2}$, i.e., in addition to what we can observe in Case 1, *now we also observe the form score*. This is again a simplification from the actual experiments, where only comparisons between forms rather than their actual scores are observable. We adopt the same assumptions as in Case 1.

A3.2 Prediction and uncertainty

With the data $\mathcal{D}$, we can build two OLS models. For form score, we have the prediction

$$\hat{s}_{\mathcal{D}} = \hat{\beta}_{\mathcal{D}}^T x, \tag{A12}$$

and for utility

$$\hat{y}_{\mathcal{D}} = \hat{b}_{s_{\mathcal{D}}}^{T}\hat{\beta}_{\mathcal{D}}^{T}x + \hat{b}_{u_{\mathcal{D}}}^{T}u, \tag{A13}$$

where $\hat{\beta}_{\mathcal{D}}$, $\hat{b}_{s_{\mathcal{D}}}$, $\hat{b}_{u_{\mathcal{D}}}$ are OLS estimates of β, b_s, b_u, respectively. For simplicity, we introduce $\hat{\theta}_{\mathcal{D}} := \{\hat{\beta}_{\mathcal{D}}, \hat{b}_{s_{\mathcal{D}}}, \hat{b}_{u_{\mathcal{D}}}\}$.

According to our models, the variance-covariance matrix of $\hat{\beta}_{\mathcal{D}}$ is

$$Var(\hat{\beta}_{\mathcal{D}}) = \sigma_s^2 (X^T X)^{-1} = \frac{\sigma_s^2}{N_2} I. \tag{A14}$$

The variance-covariance matrix of $[\hat{b}_{s_{\mathcal{D}}}, \hat{b}_{u_{\mathcal{D}}}^{T}]$ is

$$Var([\hat{b}_{s_{\mathcal{D}}}, \hat{b}_{u_{\mathcal{D}}}^{T}]) = \sigma_y^2 ([s, U]]^T [s, U])^{-1}. \tag{A15}$$

Here $s := [s^1, \ldots, s^{N_2}]^T$ are the collected form scores.

A3.3 Generalization error

The model-specific generalization error is

$$\begin{aligned} L(\hat{\theta}_{\mathcal{D}}) &= \mathbb{E}_w \left[(\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D}}^{T}x + \hat{b}_{u_{\mathcal{D}}}^{T}u - b_s\beta^T x - b_u^T u)^2 \right] \\ &= \mathbb{E}_w \left[((\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D}} - b_s\beta)^T x + (\hat{b}_{u_{\mathcal{D}}} - b_u)^T u)^2 \right] \\ &= \textstyle\sum_{i=1}^{d_x} (\Delta\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D},i})^2 + \sum_{j=1}^{d_u} (\Delta\hat{b}_{u_{\mathcal{D},j}})^2. \end{aligned} \tag{A16}$$

Here $\Delta\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D},i}$ is the ith element of $\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D}} - b_s\beta$, and $\Delta\hat{b}_{u_{\mathcal{D},j}}$ is the jth element of $\hat{b}_{u_{\mathcal{D}}} - b_u$. Similar to Case 1, the derivation requires the assumptions (R1-3).

The data-specific generalization error is

$$\mathbb{E}_{\hat{\theta}_{\mathcal{D}}}[L(\hat{\theta}_{\mathcal{D}})] = \mathbb{E}_{\hat{\theta}_{\mathcal{D}}} \left[\textstyle\sum_{i=1}^{d_x} (\Delta\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D},i})^2 + \sum_{j=1}^{d_u} (\Delta\hat{b}_{u_{\mathcal{D},j}})^2 \right] \tag{A17}$$

We first compute the second term on the right-hand side (RHS):

$$\mathbb{E}_{\hat{b}_{u_{\mathcal{D}}}} \left[\textstyle\sum_{j=1}^{d_u} (\Delta\hat{b}_{u_{\mathcal{D},j}})^2 \right] = \sum_{j=1}^{d_u} Var(\hat{b}_{u_{\mathcal{D},j}}) = tr(Var(\hat{b}_{u_{\mathcal{D}}})). \tag{A18}$$

From Eq. (A15) and using the analytical inverse of 2-by-2 block matrices, we have

$$Var(\hat{b}_{u_{\mathcal{D}}}) = \frac{\sigma_y^2}{N_2} \left(I + \frac{U^T s s^T U}{N_2 s^T s - s^T U U^T s} \right) \tag{A19}$$

Now we compute the first term on the RHS of Eq. (A17):

$$\begin{aligned} \mathbb{E}_{\hat{b}_{s_{\mathcal{D}}}, \hat{\beta}_{\mathcal{D}}} \left[\textstyle\sum_{i=1}^{d_x} (\Delta\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D},i})^2 \right] &= \mathbb{E}_{\hat{b}_{s_{\mathcal{D}}}, \hat{\beta}_{\mathcal{D}}} \left[\textstyle\sum_{i=1}^{d_x} \left((\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D},i})^2 + (b_s\beta)^2 - 2\hat{b}_{s_{\mathcal{D}}}\hat{\beta}_{\mathcal{D},i} b_s\beta \right) \right] \\ &= \textstyle\sum_{i=1}^{d_x} \left(\mathbb{E}_{\hat{b}_{s_{\mathcal{D}}}}[\hat{b}_{s_{\mathcal{D}}}^2] \mathbb{E}_{\hat{\beta}_{\mathcal{D}}}[\hat{\beta}_{\mathcal{D}}^2] - (b_s\beta_i)^2 \right) \\ &= \textstyle\sum_{i=1}^{d_x} \left(\left(Var(\hat{b}_{s_{\mathcal{D}}}) + b_s^2 \right) \left(Var(\hat{\beta}_{\mathcal{D},i}) + \beta_i^2 \right) - (b_s\beta_i)^2 \right) \\ &= \textstyle\sum_{i=1}^{d_x} \left(Var(\hat{b}_{s_{\mathcal{D}}}) Var(\hat{\beta}_{\mathcal{D},i}) + b_s^2 Var(\hat{\beta}_{\mathcal{D},i}) + \beta_i^2 Var(\hat{b}_{s_{\mathcal{D}}}) \right) \\ &= tr(Var(\hat{\beta}_{\mathcal{D}})) \left(Var(\hat{b}_{s_{\mathcal{D}}}) + b_s^2 \right) + \beta^T \beta Var(\hat{b}_{s_{\mathcal{D}}}). \end{aligned} \tag{A20}$$

From Eq. (A15) and using the analytical inverse of 2-by-2 block matrices, we have

$$Var(\hat{b}_{s_{\mathcal{D}}}) = \sigma_y^2 \left(s^T s - \frac{s^T UU^T s}{N_2}\right)^{-1} \tag{A21}$$

Lastly, we need to compute the generalization error for the questionnaire:

$$\mathbb{E}_{\mathcal{D},\varepsilon_s,\varepsilon_y,w}[L(\hat{\theta}_{\mathcal{D}})] = \mathbb{E}_{\mathcal{D}}\left[tr\left(Var(\hat{b}_{u_{\mathcal{D}}})\right) + tr\left(Var(\hat{\beta}_{\mathcal{D}})\right)\left(Var(\hat{b}_{s_{\mathcal{D}}}) + b_s^2\right) + \beta^T \beta Var(\hat{b}_{s_{\mathcal{D}}})\right] \tag{A22}$$

In the following, we will derive the upper bound of $\mathbb{E}_{\mathcal{D},\varepsilon_s,\varepsilon_y,w}[L(\hat{\theta}_{\mathcal{D}})]$ through those of $\mathbb{E}_{\mathcal{D}}\left[Var(\hat{b}_{s_{\mathcal{D}}})\right]$ and $\mathbb{E}_{\mathcal{D}}\left[tr\left(Var(\hat{b}_{u_{\mathcal{D}}})\right)\right]$. We will then compare the resultant upper bound with the generalization error from Case 1.

A3.4 Upper bound of $\mathbb{E}_{\mathcal{D}}\left[Var(\hat{b}_{s_{\mathcal{D}}})\right]$

We start by studying

$$\mathbb{E}_{\mathcal{D}}\left[Var(\hat{b}_{s_{\mathcal{D}}})\right] = \sigma_y^2 \mathbb{E}_{\mathcal{D}}\left[\left(s^T s - \frac{s^T UU^T s}{N_2}\right)^{-1}\right] \tag{A23}$$

Notice that

$$s^T s - \frac{s^T UU^T s}{N_2} = \beta^T X^T X\beta + \varepsilon_s^T X\beta + \varepsilon_s^T (I - \frac{UU^T}{N_2})\varepsilon_s \tag{A24}$$

Using (R1), we have $\beta^T X^T X\beta = \beta^T \beta$. For the second RHS term of Eq. (A24), we have $\varepsilon_s^T X\beta = \sum_{i=1}^{N_2} \varepsilon_s^i \beta^T x^i$, where ε_s^i and x^i are the error of the ith sample, and the ith input, respectively. Recall that from the data model, $\varepsilon_s^i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_s^2)$ and $x^i \overset{\text{i.i.d.}}{\sim} N(0,1)$. Therefore there exists $q > 1$,

$$|\varepsilon_s^i \beta^T x^i| < q\sigma_s ||\beta||_1 \tag{A25}$$

with high probability, where $||\beta||_1$ is the l_1-norm of β. To make this argument more concrete, we can modify the data model to let ε_s^i and x^i follow truncated normal distributions, so that Eq. (A25) holds almost surely. For example, we can truncate $|\varepsilon_s^i|$ at $3\sigma_s$, and $|x^i|$ at 3, leading to $q = 9$. With these, we have

$$\mathbb{E}_{X,\varepsilon_s}[\varepsilon_s^T X\beta] \geq -N_2 q\sigma_s ||\beta||_1 \tag{A26}$$

For the last term on the RHS of Eq. (A26), we first introduce Lemma 1.

Lemma 1 *Let $U \in \mathbb{R}^{N_2 \times d_u}$ be a random matrix where each element $u_{j,i}$ is i.i.d. standard normal. For sufficiently large N_2 and small d_u, $N_2 I - UU^T$ is p.d. (positive definite) with high probability. In particular, when $(N_2, d_u) = (100,2)$, the probability for $N_2 I - UU^T$ to not be p.d. is less than 0.01; for $(N_2, d_u) = (1000,2)$, the probability is less than 10^{-10}.*

Proof. First we note that

$$N_2 I - UU^T = \sum_{i=1}^{d_u} \left(\frac{N_2}{d_u} I - u_{\cdot,i} u_{\cdot,i}^T\right). \tag{A27}$$

We will show that $A_i = \frac{N_2}{d_u} I - u_{\cdot,i} u_{\cdot,i}^T$ is p.d. with high probability for all i. To this end, we inspect each row of A_i. For the jth row, the sum of its elements (denoted by $a_{i,j}$) is

$$a_{i,j} = \frac{N_2}{d_u} - u_{j,i}^2 - \sqrt{N_2 - 1} \frac{\sum_{k \neq j} u_{k,i}}{\sqrt{N_2 - 1}} u_{j,i}. \tag{A28}$$

Here $\frac{\sum_{k \neq j} u_{k,i}}{\sqrt{N_2 - 1}}$ and $u_{j,i}$ are i.i.d. standard normal. From Lemma 2 (below), we have

$$P\left(\frac{\sum_{k \neq j} u_{k,i}}{\sqrt{N_2 - 1}} u_{j,i} > t\right) \leq \exp\left(-h(t)t - \frac{1}{2}\ln(1 - h(t)^2)\right), \tag{A29}$$

where

$$h(t) = \frac{\sqrt{1 + 4t^2} - 1}{2t}. \tag{A30}$$

We also know that $u_{j,i}^2$ follows a chi-square distribution with 1 d.f. Therefore, we can compute the following probability

$$P(a_{i,j} \leq 0) = \int_0^\infty P\left(\frac{\sum_{k \neq j} u_{k,i}}{\sqrt{N_2 - 1}} u_{j,i} \geq \frac{N_2}{d_u} I - t\right) p(u_{j,i}^2 = t) dt. \tag{A31}$$

The above probability can be calculated numerically; e.g., for $(N_2, d_u) = (100,2)$, $P(a_{i,j} \leq 0) = 2.5 \times 10^{-5}$, and for $(N_2, d_u) = (1000,2)$, $P(a_{i,j} \leq 0) = 1.7 \times 10^{-14}$.

A sufficient condition for $N_2 I - UU^T$ to be p.d. is for A_i to be p.d. for all i. Since $a_{i,j}$ for all js are correlated, we have

$$P(a_{i,j} > 0, \forall i, j) \geq \prod_{i=1}^{d_u} \prod_{j=1}^{N_2} \left(1 - P(a_j \leq 0)\right) \approx 1 - d_u N_2 P(a_j \leq 0). \tag{A32}$$

For $(N_2, d_u) = (100,2)$, $1 - P(a_{i,j} > 0, \forall i, j) \leq 0.01$ and for $(N_2, d_u) = (1000,2)$, $1 - P(a_{i,j} > 0, \forall i, j) \leq 10^{-10}$.

Lemma 2 *Let X and Y be i.i.d. standard normal, and $Z = XY$. We have*

$$P(Z > t) \leq \exp\left(-h(t)t - \frac{1}{2}\ln(1 - h(t)^2)\right), \tag{A33}$$

where

$$h(t) = \frac{\sqrt{1 + 4t^2} - 1}{2t}. \tag{A34}$$

Proof. We first note that the random set $\{X, Y\}$ is equal in distribution to the random set $\{(X - Y)/\sqrt{2}, (X + Y)/\sqrt{2}\}$. Therefore $Z = XY$ is equal in distribution to $(X^2 - Y^2)/2$. So the moment generating function of Z is

$$\begin{aligned} \mathbb{E}_Z[\exp(hZ)] &= \mathbb{E}_{X,Y}\left[\exp\left(\frac{h}{2}(X^2 - Y^2)\right)\right] \\ &= \mathbb{E}_X\left[\exp\left(\frac{h}{2}X^2\right)\right] \mathbb{E}_Y\left[\exp\left(-\frac{h}{2}Y^2\right)\right] \\ &= (\sqrt{1 - h^2})^{-1} \end{aligned} \tag{A35}$$

Then we have

$$\begin{aligned} P(Z > t) &\leq \inf_{h \geq 0} \mathbb{E}_Z[\exp(h(Z-t))] \\ &= \inf_{h \geq 0} \exp\left(-ht - \frac{1}{2}\ln(1-h^2)\right). \end{aligned} \tag{A36}$$

Let $l(\theta) := -ht - \frac{1}{2}\ln(1-h^2)$, its minimum is reached when

$$h = \frac{\sqrt{1+4t^2}-1}{2t}. \tag{A37}$$

From Lemma 1, $\varepsilon_s^T(I - \frac{UU^T}{N_2})\varepsilon_s > 0$ with probability close to 1 for reasonably large N_2 and small d_u. By further truncating the distribution of u^i, we can make this statement certain: $\varepsilon_s^T(I - \frac{UU^T}{N_2})\varepsilon_s > 0$ almost surely for reasonably large N_2 and small d_u. We can now use Jensen's inequality and Eq. (A26) to get

$$\begin{aligned} \mathbb{E}_{\mathcal{D}}\left[Var(\hat{b}_{s_{\mathcal{D}}})\right] &\leq \sigma_y^2 \mathbb{E}_{\mathcal{D}}[(\beta^T X^T X\beta + \varepsilon_s^T X\beta)^{-1}] \\ &\leq \sigma_y^2 (\beta^T X^T X\beta - N_2 q\sigma_s ||\beta||_1)^{-1} \\ &= \frac{\sigma_y^2}{N_2}(\beta^T\beta - q\sigma_s||\beta||_1)^{-1} \end{aligned} \tag{A38}$$

A3.5 Upper bound of $\mathbb{E}_{\mathcal{D}}\left[tr\left(Var(\bar{b}_{u_{\mathcal{D}}})\right)\right]$

In order to derive and upper bound, we now study

$$\mathbb{E}_{\mathcal{D}}\left[tr\left(Var(\bar{b}_{u_{\mathcal{D}}})\right)\right] = \mathbb{E}_{\mathcal{D}}\left[tr\left(\frac{\sigma_y^2}{N_2}\left(I + \frac{U^T ss^T U}{N_2 s^T s - s^T UU^T s}\right)\right)\right]. \tag{A39}$$

The trace on the RHS has the following upper bound

$$\begin{aligned} tr\left(\frac{\sigma_y^2}{N_2}\left(I + \frac{U^T ss^T U}{N_2 s^T s - s^T UU^T s}\right)\right) &= \frac{\sigma_y^2 d_u}{N_2} + \frac{\sigma_y^2}{N_2}\frac{tr(U^T ss^T U)}{N_2 s^T s - s^T UU^T s} \\ &\leq \frac{\sigma_y^2 d_u}{N_2} + \frac{\sigma_y^2 tr(U^T ss^T U)}{N_2^3(\beta^T\beta - q\sigma_s||\beta||_1)}. \end{aligned} \tag{A40}$$

We also have

$$\begin{aligned} \mathbb{E}_{\mathcal{D}}[tr(U^T ss^T U)] &= \mathbb{E}_{\mathcal{D}}[tr(U^T(X\beta + \varepsilon_s)(X\beta + \varepsilon_s)^T U)] \\ &= \mathbb{E}_{\mathcal{D}}[tr(U^T \varepsilon_s \varepsilon_s^T U)] \\ &= \mathbb{E}_{\mathcal{D}}\left[\sum_{i=1}^{d_u}\left(\sum_{k=1}^{N_2} u_i^k \varepsilon_s^k\right)^2\right] \\ &= \mathbb{E}_{\mathcal{D}}\left[\sum_{i=1}^{d_u}\left(\sum_{k=1}^{N_2}\left(u_i^k\varepsilon_s^k\right)^2 + 2\sum_{k=1}^{N_2}\sum_{k'>k}^{N_2}\left(u_i^k\varepsilon_s^k u_i^{k'}\varepsilon_s^{k'}\right)\right)\right] \\ &= \sum_{i=1}^{d_u}\left(\sum_{k=1}^{N_2}\mathbb{E}_{\mathcal{D}}\left[\left(u_i^k\varepsilon_s^k\right)^2\right] + 2\sum_{k=1}^{N_2}\sum_{k'>k}^{N_2}\mathbb{E}_{\mathcal{D}}\left[u_i^k\varepsilon_s^k u_i^{k'}\varepsilon_s^{k'}\right]\right) \\ &= d_u N_2 \sigma_s^2. \end{aligned} \tag{A41}$$

Plug Eq. (A40) and Eq. (A41) into Eq. (A39) to get

$$\mathbb{E}_{\mathcal{D}}\left[tr\left(Var(\hat{b}_{u_{\mathcal{D}}})\right)\right] \leq \frac{\sigma_y^2 d_u}{N_2}\left(1 + \frac{\sigma_s^2}{N_2(\beta^T\beta - q\sigma_s||\beta||_1)}\right) \tag{A42}$$

A3.6 Upper bound of the generalization error

Using results from Eq. (A14), Eq. (A38), and Eq. (A42), we have

$$\mathbb{E}_{\mathcal{D},\varepsilon_s,\varepsilon_y,w}[L(\hat{\theta}_{\mathcal{D}})] \leq \frac{\sigma_y^2 d_u}{N_2}\left(1 + \frac{\sigma_s^2}{N_2\gamma}\right) + \frac{d_x\sigma_s^2}{N_2}\left(\frac{\sigma_y^2}{N_2\gamma} + b_s^2\right) + \beta^T\beta\frac{\sigma_y^2}{N_2\gamma}, \tag{A43}$$

where $\gamma = \beta^T\beta - q\sigma_s||\beta||_1$). Denote the upper bound as $\bar{L}_2 := \mathbb{E}_{\mathcal{D},\varepsilon_s,\varepsilon_y,w}[L(\hat{\theta}_{\mathcal{D}})]$. When N_2 is a large number, the upper bound can be approximated as

$$\bar{L}_2 \approx \frac{\sigma_y^2}{N_2}\left(d_u + \frac{\beta^T\beta}{\gamma}\right) + \frac{b_s^2 d_x\sigma_s^2}{N_2} \tag{A44}$$

A4 Comparison between $\bar{L}_2$ and L_1

Using results from Eq. (A11) and Eq. (A44), and denoting $N_2/N_1 = \alpha$, we can now derive the lower bound of the difference between generalization errors from Case 1 and Case 2:

$$L_1 - \bar{L}_2 = \left(d_x - \frac{\beta^T\beta}{\alpha\gamma} + d_u\left(1 - \frac{1}{\alpha}\right)\right)\frac{\sigma_y^2}{N_1} + \left(d_x + d_u - \frac{d_x}{\alpha}\right)\frac{\sigma_s^2}{N_1}. \tag{A45}$$

A few remarks can be made from Eq. (A45), as follows.

A4.1 Remark 1

When the form responses are noiseless, i.e., $\sigma_s^2 = 0$, $L_1 - \bar{L}_2 \geq 0$ if $d_x \geq 1/\alpha$. Consider the specific case where $N_1 = 2N_2$ ($\alpha = 0.5$). Then $d_x \geq 2$ will suffice for questionnaire 2 to have generalization error no greater than questionnaire 1.

A4.2 Remark 2

When $\sigma_s^2 \neq 0$, $L_1 - \bar{L}_2 \geq 0$ if the form signal-to-noise ratio $||\beta||_1/\sigma_s$ is sufficiently large, and the form-to-utility noise ratio σ_s^2/σ_y^2 is sufficiently small. To show this, we start by rearranging Eq. (A45) to show that $L_1 - \bar{L}_2 \geq 0$ if

$$\sigma_s^2 \leq \left(\frac{\alpha - \frac{||\beta||_2^2}{\gamma d_x}}{1-\alpha}\right)\sigma_y^2. \tag{A46}$$

Because σ_s^2 is positive, the above condition cannot hold when $\alpha\gamma d_x < ||\beta||_2^2$. Using the fact that

$$||\beta||_2^2 \leq ||\beta||_1^2, \tag{A47}$$

we can derive the following sufficient condition when $L_1 - \bar{L}_2 < 0$:

$$
\begin{aligned}
& \alpha\gamma d_x < ||\beta||_2^2 \\
\Rightarrow \quad & \alpha(||\beta||_2^2 - q\sigma_s||\beta||_1)d_x < ||\beta||_2^2 \\
\Rightarrow \quad & \alpha q\sigma_s||\beta||_1 d_x > (\alpha d_x - 1)||\beta||_2^2 \\
\Rightarrow \quad & \alpha q\sigma_s d_x > (\alpha d_x - 1)||\beta||_1 \\
\Rightarrow \quad & ||\beta||_1 < \frac{\alpha q\sigma_s d_x}{\alpha d_x - 1}.
\end{aligned}
\tag{A48}
$$

For the specific case where d_x is large and $\alpha = 0.5$, the above condition can be reduced to

$$||\beta||_1 < q\sigma_s \tag{A49}$$

■

APPENDIX B (Online): Accommodating the "No Choice" Option

Because our purchase question can be viewed as a forced binary choice, we explored the changes required to accommodate a "no choice" option as well. This can be done via a minor extension of Eq. (3) and (4). The primal problem becomes:

$$\min_{\mathbf{w}} \mathbf{w}^T\mathbf{w}$$
$$\text{subject to } \mathbf{w}^T\phi\left(\mathbf{x}_j^{(1)}\right) - \mathbf{w}^T\phi\left(\mathbf{x}_j^{(2)}\right) \geq c_j, \forall \text{ choices } j \text{ other than no choice}, c_j \in \{1,2\}$$
$$1 \geq \mathbf{w}^T\phi\left(\mathbf{x}_{j'}^{(1)}\right) - \mathbf{w}^T\phi\left(\mathbf{x}_{j'}^{(2)}\right) \geq -1, \forall \text{ choices } j' \text{as "no choice"}$$

The dual is then:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - [\mathbf{c}^T, \mathbf{1}^T, -\mathbf{1}^T]\boldsymbol{\alpha}$$
$$\text{subject to: } \boldsymbol{\alpha} \geq \mathbf{0}.$$

The additional constants in the linear term of the objective come from the constraints related to the "no choice" data. Since this problem has exactly the same formulation as the one examined in the paper, the same algorithm applies.

APPENDIX C (Online): Lagrangian Formulation of Eq. (3)

We note again that the SVM formulation does not explicitly rely on a notion of likelihood, Yet it is possible to treat the convex SVM objective as a negative log-likelihood using hinge-loss and a simple i.i.d. Gaussian prior; see, for example, Evgeniou, Pontil, and Toubia (2007).

For the problem formulation in (3), first note that its Lagrangian is:

$$L(\mathbf{w}, \boldsymbol{\alpha}) = \mathbf{w}^T\mathbf{w} + \sum_j \alpha_j \left(c_j - \mathbf{w}^T \left(\boldsymbol{\phi}\left(\mathbf{x}_j^{(1)}\right) - \boldsymbol{\phi}\left(\mathbf{x}_j^{(2)}\right) \right) \right).$$

$L(\mathbf{w}, \boldsymbol{\alpha})$ can be considered as a negative log-likelihood with parameters $\boldsymbol{\alpha}$ balancing the Gaussian prior ($\mathbf{w} \sim N(\mathbf{0}, \mathbf{I})$) and the data ($\mathbf{x}_j^{(1)}$ more preferred than $\mathbf{x}_j^{(2)}$). The likelihood has the following form:

$$L(\mathbf{w}; \{(\mathbf{x}_1, \mathbf{x}_2)_i, c_i\}) = \exp(-\frac{1}{2}\mathbf{w}^T\mathbf{w}) \prod_j \exp\left(-\frac{\alpha_j}{2} \left(c_j - \mathbf{w}^T \left(\boldsymbol{\phi}\left(\mathbf{x}_j^{(1)}\right) - \boldsymbol{\phi}\left(\mathbf{x}_j^{(2)}\right) \right) \right) \right),$$

which can be further simplified as

$$L(\mathbf{w}; \{(\mathbf{x}_1, \mathbf{x}_2)_i, c_i\}) = \exp(-\frac{1}{2}\mathbf{w}^T\mathbf{w}) \prod_j \exp\left(-\frac{\alpha_j}{2} \left(c_j - \mathbf{w}^T\mathbf{z}_j \right) \right),$$

with $\mathbf{w}$ the estimated parameters, $\mathbf{z}_j = \boldsymbol{\phi}\left(\mathbf{x}_j^{(1)}\right) - \boldsymbol{\phi}\left(\mathbf{x}_j^{(2)}\right)$ the data, and $\boldsymbol{\alpha}$ hyperparameters. This indicates that the probability of choosing $\mathbf{x}_j^{(1)}$ over $\mathbf{x}_j^{(2)}$ follows the PMF:

$$\Pr\left(\mathbf{x}_j^{(2)} \succ \mathbf{x}_j^{(2)}; \mathbf{w}\right) = \begin{cases} 1, \mathbf{w}^T\mathbf{z}_j \geq c_j \\ \exp\left(-\frac{\alpha_j}{2} \left(c_j - \mathbf{w}^T\mathbf{z}_j \right) \right), \text{otherwise} \end{cases}.$$

Several differences between our model and Evgeniou et al.'s (2007) should be mentioned:

(1) We use a Gaussian kernel to define $\boldsymbol{\phi}^T\boldsymbol{\phi}$

(2) Due to the infinite dimension of the feature space induced by the Gaussian kernel, we do not model or learn the variance-covariance matrix of the part-worths. This matrix is assumed to be identity (as is indicated by the shrinkage term).

(3) We treat the data as constraints in the form: $\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2) \geq c$ for all $\mathbf{x}_1$ preferred to $\mathbf{x}_2$, while in Evgeniou et al. (2007), data are used to form a quadratic loss in the objective: $\left(1 - \mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2)\right)^2$, for which an analytical solution exists.

[1] Evgeniou, Theodoros, Massimiliano Pontil, and Olivier Toubia. "A convex optimization approach to modeling consumer heterogeneity in conjoint estimation." *Marketing Science* 26, no. 6 (2007): 805-818.

APPENDIX D (Online): Insensitivity of Individual-Level Learning to Constraint Value

The constant in Eq. (10) is set exogenously to 1. Note that, when it's set to 0, a trivial solution ($w = 0$) becomes the optimal one. To prevent this from happening, the hard-margin SVM formulation uses "1" as the minimal gap between the preferred and non-preferred choices.

The following explains why the constant gap can be set arbitrarily. First, similar to the conversion from Eq. (3) to Eq. (4), here the dual of Eq. (10) is:

$$\text{(P1)} \quad \min_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - \mathbf{1}^T\boldsymbol{\alpha}$$
$$\text{subject to: } \boldsymbol{\alpha} \geq \mathbf{0},$$

where the coefficient vector ($\mathbf{1}$) of the linear term in the objective is contributed by the constant gap in the primal. Let its solution be $\boldsymbol{\alpha}_1^*$, and now also let the gap be c. The resulting dual becomes:

$$\text{(P2)} \quad \min_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} - c\mathbf{1}^T\boldsymbol{\alpha}$$
$$\text{subject to: } \boldsymbol{\alpha} \geq \mathbf{0},$$

It is readily seen that the solution of (P2) is $c\boldsymbol{\alpha}_1^*$, and thus the solution to the primal of P2 will also be a scaled version of that of Eq. (10). **This means that the choice of the gap will not affect the preference ranking**, i.e., if one product is preferred to the other under the model derived from (P1), the same is true under that from (P2).

We note, however, that when two gaps are introduced, as is the case in Eq. (4), the choice of the gap values (in the paper, these are 1 and 2) does affect the solution, i.e., choosing other numbers than 2 may lead to a preference model with different rankings. We examine this case in detail in Appendix G.

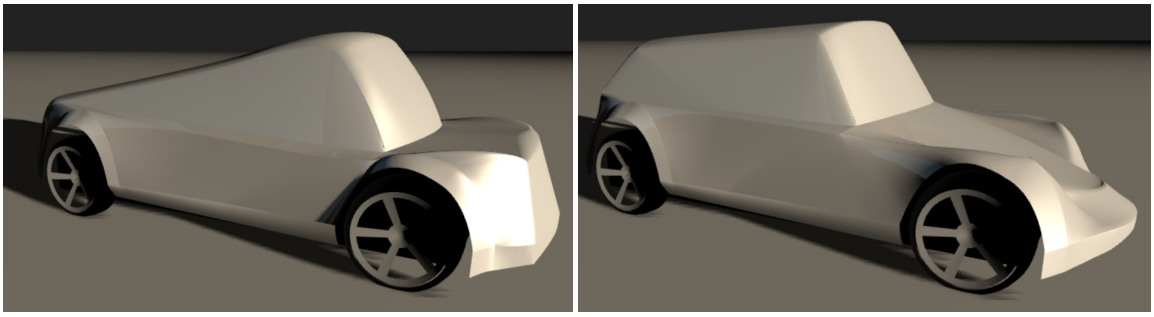
APPENDIX E (Online): Setting Weights (v_1 and v_2) in Eq. 16

As noted in the paper proper, $v_1 = 0.99$ and $v_2 = 0.01$ were selected for the experiments. Here we explore why, as well as the effects of alternative settings.

First, note that the lopsided weighting ($v_1/v_2 = 99$) is a result of the finding that when v_2 is large, the generated design vector (x) tends to be populated by zeros and ones. This is reasonable, since such vectors are the corners of the 19-dimensional design space, and so are truly "apart" from one another. This is consistent with the curse of dimensionality: higher-dimensional Euclidean space has larger corner volumes.

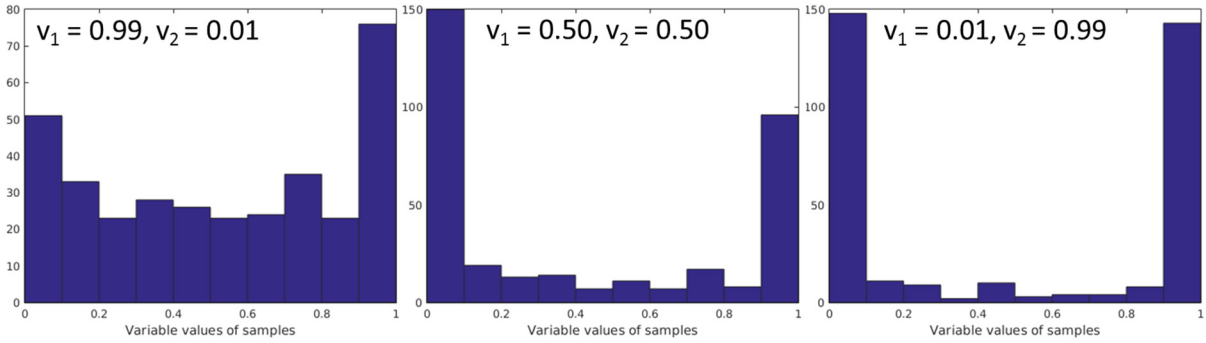
Based on this finding, we empirically tested various settings of the weights. The chosen ones produce designs off the boundaries of the design space, and also achieve reasonable hit-rate on the validation set during an internal preliminary test. [Specifically, we conducted many pilot surveys before the final one, and tuned parameters through those pilot surveys.]

While these settings do satisfy the need to explore the design space, the resultant user comparisons between extreme designs may not lead to effective learning of preference models that can distinguish designs with mild parametric changes. In addition, we hypothesize (but cannot validate in the absence of attitudinal data) that users will be less involved and feel confused when asked to nearly always compare designs with unusual appearances. Here is one such example:



For this reason, we would like the samples to have less extreme values, which motivates the presented settings of $v_1 = 0.99$ and $v_2 = 0.01$.

More concretely, here are three distributions of the sample $\mathbf{x}_2^{new}$ with v_1 set to 0.99 (the value used), 0.50 (equal weighting), and 0.01 (reversed weighting):



These results help justify the empirical choice of $v_1 = 0.99$ and $v_2 = 0.01$. This is because only the first provides a reasonably uniform distribution over the variable values, while even equal weighting is heavily U-shaped.

[**Details on producing these figures**: The style function follows the first simulation scenarios in the paper. The GA algorithm follows the same settings as in the user experiments: We run 100 generations with population size 20 for Eq. (15), and 500 generations with population size 50 for Eq. (16). We draw one pair of samples from $[0,1]^{19}$ initially, and sample the space by solving Eq. (15) and Eq. (16) for 18 iterations while accumulating the samples. The plotted distributions combine values from the 18 samples of $\mathbf{x}_2^{new}$.]

APPENDIX F (Online): Simultaneous vs. Sequential Optimization for Eq. (15) and (16)

Because the problems in Eq. (15) and Eq. (16) are highly nonconvex, they were solved using Genetic Algorithms (GAs); and, due to limited response time by the application engine, search terminated when the "max" number of iterations was reached. This "max" was in fact empirically tuned for the server we used (Google App engine), so that we could obtain responses within the server response limit (30 seconds) without causing glitches or long delays during user interactions.

Under this setting, one would still be able to obtain a response if one solved the following all-in-one problem instead of decomposing it into Eq. (15) and Eq. (16):

$$\max_{\mathbf{x}_1^{new},\mathbf{x}_2^{new}\in[0,1]^{19}} F_{aio}(\mathbf{x}_1^{new},\mathbf{x}_2^{new}) = v_1\exp(-\|S(\mathbf{x}_1^{new})-S(\mathbf{x}_2^{new})\|^2) + v_2(\|\mathbf{x}_1^{new}-\mathbf{x}_2^{new}\|^2 + \min_j\left\|\mathbf{x}_1^{new}-\mathbf{x}_j^{old}\right\|^2 + \min_j\left\|\mathbf{x}_2^{new}-\mathbf{x}_j^{old}\right\|^2)$$

We can compare the "solution qualities", i.e., the values of $F_{aio}(\mathbf{x}_1^{new},\mathbf{x}_2^{new})$ by solutions from solving the all-in-one problem directly and through Eq. (15) and Eq. (16) sequentially. The operative hypothesis is that the latter will tend to yield better solutions; because the problem is decomposable, searching in two smaller spaces is more effective than in the combined space, **given the same budget for searching**.

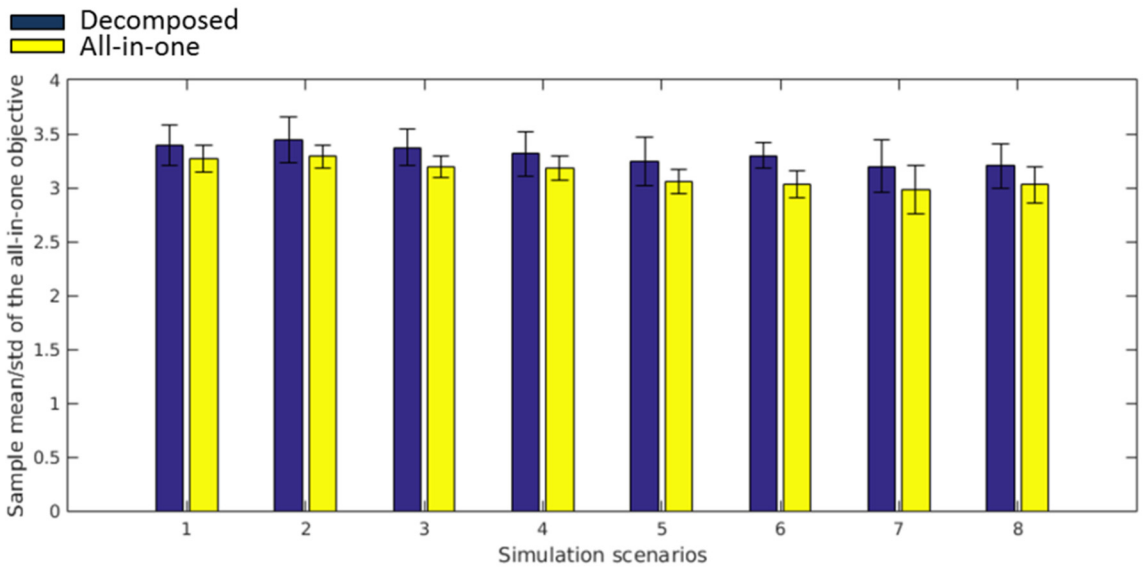
To test this hypothesis empirically, we conducted the following experiment. The style function follows the 8 simulation scenarios of Table 7:

Table 7: Consumer Preference Scenarios

Form importance	Response accuracy	Respondent hetero-geneity	Form score weight (λ)*	Form attribute coefficients: Independent terms (γ)	Form attribute coefficients: Interaction terms (δ)	Functional attribute partworths
Low	Low	Low	0.0043	N(0.5, 0.25)	N(0, 4.80)	N(0.5, 0.25)
Low	Low	High	0.0044	N(0.5, 1.5)	N(0, 13.7)	N(0.5, 1.5)
Low	High	Low	0.0028	N(3.0, 1.5)	N(0, 56.3)	N(3, 1.5)
Low	High	High	0.0057	N(3.0, 9.0)	N(0, 88.4)	N(3, 9.0)
High	Low	Low	0.0173	N(0.5, 0.25)	N(0, 4.80)	N(0.5, 0.25)
High	Low	High	0.0176	N(0.5, 1.5)	N(0, 13.7)	N(0.5, 1.5)
High	High	Low	0.0112	N(3.0, 1.5)	N(0, 56.3)	N(3.0, 1.5)
High	High	High	0.0230	N(3.0, 9.0)	N(0, 88.4)	N(3.0, 9.0)

Specifically, v_1 and v_2 are set to 0.99 and 0.01, respectively. The GA algorithm follows the same settings as in the user experiments: We run 100 generations with population size 20 for Eq. (15), and 500 generations with population size 50 for Eq. (16). For the all-in-one problem, we run 600 generations with population size 50. Note that this setting is slightly in favor of the all-in-one case as it uses the larger population size. We also draw 18 pairs of $\mathbf{x}^{old}$ uniformly from $[0,1]^{19}$ in order to calculate F_{aio}.

With these settings, we report a comparison between the F_{aio} values derived from the two optimization routines:



That the "decomposed" means have uniformly larger objective values provides substantial empirical support for the hypothesis and the pragmatic choice of using Eq. (15) and Eq. (16).

APPENDIX G (Online): Robustness Checks for Preference Gap Cutoffs, c_j, in Eq. (3)

In Eq. (3), c_j, was set to 1 and 2, which can entail substantive effects in the final solution. Here, we examine sensitivity to these choices. The sensitivity of the gap (1 for "better" and 2 for "much better") can be found from the Lagrange multipliers of the solution of the dual problem in Eq. (4). Investigating revealed that all Lagrange multipliers are positive (sample mean = 1.51, sample std = 0.81), i.e., all constraints in the SVM formulation are active. This indicates that the choice of the gaps (1 and 2) can indeed affect the styling preference model. This result, however, can be expected since only a relatively small number of questions are used to form each individual-level preference model in a high-dimensional space.

Note that the choice of the scale of the gap should not affect results substantially, since style preference is **further scaled by a part-worth** when forming the overall purchase preference, i.e., increasing the gap from 2 and decreasing the part-worth for styling will *maintain the landscape of the purchase preference with respect to shape-related variables*. The effects of the Gaussian parameter and the shrinkage parameters on generalization performance usually requires cross-validation, which may not be feasible for real-time interaction. Therefore, we set these parameters according to the default values used in the standard LibSVM package (e.g., www.csie.ntu.edu.tw/~cjlin/libsvm/).

We tested different c_j in Eq (3) for the 'worst' scenario in Table 8: high form importance, low response accuracy, and high heterogeneity. In this simulation, although {100, 200} yields the highest hit rate, the overall performance is relatively unaffected. We present these in two tables: {ratio fixed at ½; scale varies from 0.1 to 1000} and {ratios vary from ½ to 1/10, lower scale point fixed at unity}

<Choosing different c_j in Eq. (3)>

c_j	{0.1,0.2}	{1,2} (base)	{10,20}	{100,200}	{1000,2000}
Form preference hit rate	64.2	65.1	65.5	66	64.5

c_j	{1,2} (base)	{1,3}	{1,5}	{1,7}	{1,10}
Form preference hit rate	65.1	65.1	64.2	64.9	64.6

These values are nearly entirely invariant across a wide range of ratios and magnitudes, suggesting that the final results are not sensitive to this choice, at least for this problem. For others, manual tuning may be inevitable.

We note in closing that it is possible to introduce slack variables to form a soft SVM, which allows the gaps to be flexible rather than fixed to 1 and 2. However, as in any soft SVM formulation, doing so will inevitably introduce a hyperparameter that determines the balance between the training error (the sum of slacks) and the shrinkage. A typical way to tune the hyperparameter is through cross-validation. Thus, doing so during the online survey will roughly increase response time by a factor of N, where N is the number of folds used in cross-validation.

APPENDIX H (Online): Details on Query Engine and Response Times

Here we provide benchmarks regarding runtimes, since rapid query generation is critical for the success of any crowdsourced online survey platform. The live survey is hosted on a paid Google Application Engine site: https://vehiclechoicemodel.appspot.com.

We collected wall-clock response times for queries of the web application, using a lower- and a higher-end instance classes provided by Google App Engine. From the data, we conclude that (1) substantial differences exist between the response times for the two instance classes; and (2) later queries in the survey require longer response times. **Results show that scaling of the presented web app is feasible using existing cloud computing platforms in the market (e.g., Google App Engine).**

Details of the setup: The web application is implemented in JAVA 8 on the server side and javascript on the client side, and deployed to a Standard environment on Google App Engine. Two instance classes for comparison are the first-generation F1 and F4 (for JAVA 8, second-generation instances are not available). The specifications and costs of these instance classes are provided below [spec, price].

Instance class	Memory Limit	CPU Limit	Supported Scaling Types	Cost per hour per instance
F1	128 MB	600 MHz	Automatic	$0.05
F4	512 MB	2.4 GHz	Automatic	$0.20

We adopted the standard environment, but have not tested the response time of the app in a **flexible** environment, which is suitable for consistent traffic [doc]. The source code of the web app is available here [source; redacted for review], and the latest version of the web app can be accessed here [app].

Response time comparison: For each instance class, we record the response time for four *rounds* of queries. Each round consists of the following sequence of queries: Form ("F"), Attribute ("A"), Form and attribute ("F + A"), Form only. Except for the first "F" query which has pre-defined form parameters, each "F" query activates preference learning and the search of a new pair of forms. Each "A" query activates the search of a new pair of attributes for the given forms. Each "F + A" query activates the search of a new pair of forms and attributes. The "Form only" query only sends user choice data back to the server, and does not require any response from the server. Therefore, it has negligible response time.

Since the randomness in response time is insignificant, we conducted three independent surveys for each instance class. The mean and standard deviations of the response time (unit: sec.) are presented below:

F1	Round 1	Round 2	Round 3	Round 4
F + A	4.63 (1.57)	6.35 (0.23)	09.43 (0.35)	11.89 (0.10)
F	5.20 (0.43)	7.98 (0.54)	11.17 (0.42)	14.03 (0.34)
A	0.85 (0.26)	0.95 (0.23)	00.89 (0.08)	00.91 (0.06)
F4				
F + A	1.29 (0.06)	2.79 (0.04)	3.99 (0.14)	4.93 (0.11)
F	2.21 (0.16)	3.42 (0.09)	4.68 (0.19)	5.53 (0.13)
A	0.57 (0.11)	0.84 (0.08)	1.06 (0.24)	0.93 (0.06)

The results show that F4 processes queries effectively faster than F1, as expected, in particular for the heavier computations involved in "F + A" and "F" queries. We consider the average of these – roughly on the order of 3 seconds, and less than half the values for F1 –sufficient for commercial applications, and of course they can be further reduced with more computational power.

As the survey plays out, response times increase. This is because later searches have higher overhead, i.e., they incorporate samples from Form and Preference models learned based on previous rounds of survey questions. Lastly, the "F" queries take a slightly longer time than the "F + A" queries simply because the former are executed after the latter, and thus have slightly higher overhead in computation.

[spec] Google, The App Engine Standard Environment. URL: https://cloud.google.com/appengine/docs/standard/#instance_classes

[price] Google, App Engine Pricing. URL: https://cloud.google.com/appengine/pricing

[doc] Google, Choosing an App Engine environment. URL: https://cloud.google.com/appengine/docs/the-appengine-environments

[source] Source code for the paper. URL: https://github.com/RedactedForReview

[app] Web app for the paper. URL: https://vehiclechoicemodel.appspot.com/