Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

Exploring the effectiveness of interactive preference learning for adapting designs to abstract semantic attributes

Ananya Nandy
Dept. of Mechanical Engineering
University of California, Berkeley
Berkeley, CA, USA
ananyan@berkeley.edu

Kosa Goucher-Lambert
Dept. of Mechanical Engineering
University of California, Berkeley
Berkeley, CA, USA
kosa@berkeley.edu

ABSTRACT

Abstract semantic attributes of designs (e.g., comfortable, luxurious, durable) play a significant role in the assessment of user-facing products, capturing intangible factors that people may consider aside from performance requirements. However, due to the difficulty of mapping highly subjective and varying perceptions to specific design features, it remains a challenge to quickly and accurately translate these qualities into designs using computational design tools. Seeking to align computational and human representations of subjective design information, we investigate the utility of adapting representations of semantic attributes to designers' perceptions through interactive models. A study is conducted in which users evaluate parameterized drinking mugs, indicating their perceptions of how comfortable each is to hold. Interactive Bayesian optimization is used to adaptively arrive at a design that optimizes this subjective quantity for each participant individually. Participants (N = 31) guide the model by providing their own decisions or building off of empirical data from a prior group of participants (N = 25). The resulting designs are evaluated across different scenarios, demonstrating the extent to which outputs of non-interactive models can be used to represent a subjective, semantic attribute and how interactive models may improve perceived alignment between human intent and computionally-generated outputs.

1 INTRODUCTION

Successful design often involves creating products or systems that, in addition to satisfying functional requirements, are desirable in their perceptions or ability to evoke emotional response [1, 2, 3]. In the words of Krippendorff and Butter, "To design artifacts for use by others is to design them to be or to have the chance to become meaningful to these others" [4]. Thus, designers have the challenging task of translating high-level, meaningful concepts into low-level features (e.g., geometry), providing value that goes beyond function. Representing abstract semantic information within designs is challenging due to its highly subjective nature. Designers must balance factors such as expressing attributes desired by consumers while also preserving brand recognition or styling [5, 6]. Culture and context can also impact how designers embed visual qualities into a consumer-facing design to align with consumers' expectations [7]. Therefore, automated methods for this type of semantic-to-visual translation must be able to account for such variability in designers' perceptions, enabling them to effectively fulfill their design intent, often in an iterative manner.

The task of embedding subjective qualities within a design (i.e., mapping abstract semantic information to design parameters) has often been left to a designer's intuition or achieved through methods such as conjoint analysis, where the value of different product features are elicited from participant groups and analyzed. These methods have been used to quantify subjective information for design, often preferences over a product's form or functional attributes,

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

by collecting and aggregating data from consumers (e.g., through pairwise decisions or ratings) and subsequently developing preference models [8, 9, 10, 11, 12, 13, 14]. Conjoint analysis ensures outcomes that are specific to the domain and attributes of interest, but is not suited to numerous attribute levels (which is a characteristic of continuous geometries) or incorporating many interaction effects [15]. Furthermore, its use may be deferred to later stages of the design process, such as to focus on marketing and pricing after design has been finalized, because the approach is data and time-intensive [16].

The emergence of powerful AI-driven text-to-image models has further brought the challenge of accurately translating between semantic to visual representations to the forefront, as these models can increasingly be used to support design activities [17, 18]. These models enable the generation of common artifacts using semantic expression directly [19, 20]. However, the models may not be trained on data from specific design domains and often require concrete, rather than abstract, semantic specification to produce desired outputs [21]. A complementary approach can address these challenges by leveraging interactive, human-in-the-loop methods (i.e., methods where an individual can guide the outcome generation process) rather than relying solely on static data or pre-trained models. Interactive, human-in-the-loop models have been developed to collect preference-related data in contexts like product marketing and user personalization [22, 23]. While the prior studies have focused on visually-grounded (e.g., color, material, form) or physically-assessed attributes, this study focuses on learning perceptions of more intangible abstract qualities, which can vary across individuals, in real time. This research therefore evaluates the usefulness of interactive preference learning to enable flexible expression of these types of abstract semantic attributes within design outcomes, based on the alignment of the outcomes with individuals' perceptions and characterization of how users explore and evaluate the design space. Specifically,

1. To what extent do interactive models guided by iterative user feedback improve alignment with individuals' perceptions compared to non-interactive models?
2. How does prior data affect the efficiency and outcome quality of interactive optimization for individual users?
3. What behavioral patterns and interaction strategies (e.g., design space exploration and decision-making) emerge during interactive optimization, and how do they shape satisfaction with the final designs?

To investigate these questions, a human subjects study was conducted where participants were able to adapt a generic design (a parameterized drinking mug) to individually express an abstract semantic attribute (comfortable to hold). Alignment between human perceptions and computationally optimized results from Bayesian preference learning was empirically evaluated. The study results in (1) an evaluation of outcomes (with respect to an abstract semantic attribute) from preference learning models where users exert varying amounts of influence through prior data and/or interactivity and (2) a characterization of decision making during the optimization process to understand designers' use of the interactive method.

The primary contribution of this work is new empirical and methodological insight into the viability and utility of interactive Bayesian optimization (BO) in the domain of human-centered design, specifically with subjective and embodied criteria. The user interaction protocol and empirical findings extend the application of BO in non-trivial ways, especially given the difficulty of expressing and modeling subjective criteria in physical product design. While prior applications of preference learning have largely focused on arbitrary or highly individualized preferences, our findings highlight the importance of capturing both shared perceptual structures and individual variation when navigating complex, subjective design goals. Consequently, this research offers user-centric insights into the benefits and costs of embedding computational design methods with interactivity to achieve subjective design goals that are challenging to formalize or quantify.

2 RELATED WORK

2.1 Mapping Subjective Attributes to Designs

There have been many efforts to quantify the perceptual space of user needs ("product semantics") to translate into designs. Within computer graphics, several approaches have been taken to map subjective semantics to 3D geometries. For example, geometric elements that preserve stylistic similarity between 3D shapes have been used for transferring styles to functionally compatible shapes [24]. Another approach has used crowdsourced pairwise comparisons to map subjective attributes (e.g. comfortable, sporty, etc.) directly to geometry using continuous deformation shape

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

editing [25]. Kansei engineering is a popular approach to extracting the desired emotion from a product [26]. Rating-based semantic differentials, such as those used in Kansei engineering, are one way to quantify subjective attributes and related preference. For instance, Reid et al. use ratings to quantify and generate new designs that better reflect a specific semantic attribute, perceived environmental friendliness [27]. Our work also uses a singular subjective attribute, perceived comfort, as an example. However, as ratings (or methods such as ranking or clustering) can require higher effort [28], we use pairwise decisions, as in [25]. Another approach to capturing product semantics has been to utilize multidimensional scaling to build similarity-based perceptual embedding spaces and relate this space to vectors of various semantic attributes [8,29,30]. Prior work has also used perceptual embedding spaces for difficult-to-capture quantities [31, 32]. However, a challenge is that these perceptual embeddings do not capture individual differences in how people make their judgments, which has been found to impact the construction of quantitative psychological spaces [33]. At the same time, a simulated experiment shows that when crowd-level preferences form, heuristics from the crowd information can increase the efficiency of eliciting preferences [34]. Therefore, an important consideration for addressing these attributes is the balance of aggregation and individualization. We address this balance in our study by considering the impact of combining iterative feedback and prior data vs. relying fully on individual feedback.

2.2 Capturing Subjective Evaluations through Offline Preference Modeling

In this work, "preference" is captured along a specific subjective dimension rather than more broadly. However, preference modeling has more generally been applied extensively to engineering design. Early work in using preference modeling techniques for product design uses a lottery question-based framework to create utility functions that reflect a designer's priorities [35]. Utility functions numerically represent, broadly, how a person values a given option. Though the use of utility analysis has its limitations in engineering design, a major benefit is its ability to "model subjective tradeoffs, particularly those that are nonlinear and/or that must be made under uncertainty," which can be of particular use for adapting to individuals [36].

A common method to model preference is to determine the expected form of a utility function and then estimate weights via a discrete choice experiment, where participants make decisions between a number of choices (often pairwise). In many studies of preference, the choices to be presented to participants are determined ahead of time based on methods such as random sampling, D-efficiency, or Latin square design among others [15, 12, 13, 27]. Approaches have been developed to incorporate form (generally, aesthetics) into these utility functions [9, 14, 37], with preference models used to analyze tradeoffs between function and form [38, 10, 13]. The methods used in the above studies allow the analysis of attribute weightings to determine their impacts separately, but the presence of interaction effects can be a challenge [15].

Prior work has also considered methods that can more flexibly capture nonlinearities inherent to subjective evaluation. For example, support vector machines, Markov chains, genetic algorithms, or artificial neural networks have been used to non-parametrically map subjective attributes to design variables [11, 39, 40, 41]. Additionally, research has successfully demonstrated feature learning for predicting preferred designs [42]. Feature learning is a promising approach to quantifying subjective semantic attributes at an individual level, but requires the collection of large amounts of data (e.g., social media sentiments or user demographics) for offline training [43, 42, 44]. Furthermore, these approaches are often more useful when eliciting insights from consumers, after a clear design direction has been established, rather than during design generation. Like many of the approaches mentioned, the preference learning method applied in our study accounts for potential nonlinearities and interaction effects. However, it additionally allows for real-time interactions (though it does not require it), allowing a comparison of how the interactions impact alignment with perceptions of abstract semantic attributes.

2.3 Interactive Optimization and Active Learning

Interactive and personalized approaches have potential for improving how subjective qualities can be expressed in design. Focusing on the product semantic of "elegance," Poirson et al. use interactive evolutionary computing (IEC) to move towards incorporating individualized perceptions [40]. While interactive evolutionary computing can be used to find outcomes optimized for the subjective attribute, it can be helpful to learn a function that represents the perception of the attribute, like a reward function in human-robot interaction [45]. Such a function can then be utilized for further decisions the individual makes, such as optimizing that attribute for a different design with the same high-level features or considering tradeoffs with other subjective attributes, via reinforcement learning.

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

An alternative method to IEC that has become popular for its flexibility and viability with smaller amounts of data is Bayesian optimization. Bayesian optimization (closely related to Kriging models from geostatistics [46]) is a method that allows a blackbox function to be optimized. Bayesian optimization methods have been explored in many domains, including visual parametric design, to tackle target-oriented cases when high-level feedback is easier to provide than tweaking parameters [47,48,49]. Using Bayesian optimization for individual users is particularly promising, for example, in the case of assistive technology such as exoskeleton gaits or hearing aids [50, 51]. Within engineering design, Bayesian optimization has been used in cases when high-fidelity simulations are computationally expensive to run or when feasibility constraints have to be specified based on domain knowledge [52, 53, 54]. Bayesian optimization can reach an optimal outcome much like interactive evolutionary computing, but it does so through a surrogate function that approximates the hard-to-evaluate, unknown function [55,56]. Surrogate functions are particularly useful for the evaluation of subjective attributes since it is difficult to assume a functional form that will be appropriate for a person's judgments. Specifically, using Gaussian processes (GPs) allows non-parametric estimation of a person's utility function (for example, in Fig. 5), where the form of the function does not have to be specified ahead of time but smoother functions are preferred [55]. Using GPs and their associated uncertainty quantification with active learning can address some of the challenges associated with design of experiments and individual differences. Gaussian Process-based models offer significant advantages over traditional conjoint analysis preference models by providing a more flexible, non-linear representation of subjective preferences, capturing complex interactions between design attributes and continuous geometries.

Active querying has been used for preference elicitation in engineering design, allowing quick convergence to a true utility function for a multi-objective problem [57]. Active preference learning has also been applied to finding the best product concepts when systematically assigning weights to product attributes is difficult, finding rankings for concepts that align with what an experienced designer selects manually [58]. Active learning has also been applied to quantify form [56] and form and function tradeoffs [22]. Although active learning is a challenge itself due to the high dimensionality of design spaces, this study uses Bayesian optimization and active learning in order to allow efficient and adaptive data collection with as few evaluations of the quantity of interest as possible. We build on prior approaches to implement an interactive optimization process based on pairwise decisions and active guidance, using it to investigate if and how interactive models might be used to better align designers' perceptions of subjective attributes with computational representations.

3 METHODS

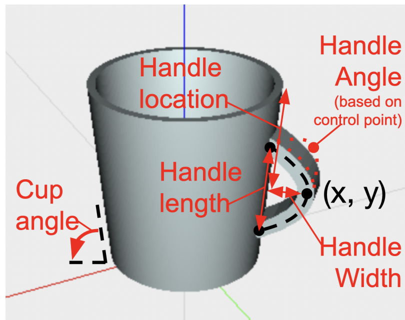
Actively-generated pairwise queries were used to find designs that optimize a subjective attribute. These outcomes were evaluated and compared across models with varying amounts of prior data and interactivity. The study procedure and interactive optimization method are outlined in the following section.

3.1 Design Example

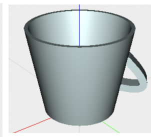
The chosen design example was a drinking mug, while the subjective attribute of interest was how comfortable the mug was to hold. The subjective dimension of "comfortable-to-hold" (vs. an even more abstract consideration like "elegance" found in [40]) was used because, though "comfortable" is abstract, there are known aspects of the design, related to variations in how the mug can be held, that are more likely to be associated with perceptions. This allows the high-dimensional design space to be narrowed to include fewer, but relevant, features, allowing for faster real-time computation. We acknowledge that "comfortable" as an attribute is not strictly based on visual perception, yet humans commonly must judge this quality using vision (e.g., in e-commerce) when the real object is unavailable. Since a mug is a simple, everyday object that most participants have picked up and used, it was assumed that decisions based on visual information were sufficient for the purpose of this study. The 3D representation of the mug (shown in Fig. 1) consisted of a cup with a fixed height, thickness, and bottom radius and a fixed thickness handle created from a Bezier curve with three control points. The mug had five variable parameters: the taper of the cup (cup angle), the distance between the first and last handle control points along the cup surface (handle length), the location of the center of first and last control points along the cup surface (handle location), and the x (handle width) and y (handle angle) positions of the middle control point. Bayesian optimization methods applied in a design or engineering design context have primarily been applied to 2 to 7-dimensional design spaces [40, 50, 54, 47]. These specific parameters were selected to

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

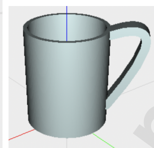
directly map to how the mug can be held using the handle and the outside of the cup because of the necessity to limit the dimensionality of the design space. The parameter bounds were set to extremes that were manually determined to be perceptually reasonable for mugs that exist in reality (shown in Fig. 1b and 1c) and these bounds were used to range normalize the design space to a five-dimensional unit hypercube. The variables were treated as continuous within the hypercube and only discretized for query selection.



(a) Parameter types and labels



(b) Min. bound



(c) Max. bound

Fig. 1: Design space of the parameterized mug presented to participants with corresponding feature labels.

3.2 Participants

Data from 56 participants (29 women, 24 men, 3 nonbinary) were collected with approval from an Institutional Review Board in two stages: Group A (N = 25) and Group B (N = 31). These participants were recruited from university mailing lists that primarily consist of engineering undergraduate or graduate students (32 undergraduate, 19 graduate, 5 graduated/working) and as such, do not necessarily represent a general population, but may represent a population that would desire to incorporate subjective attributes into design. The participants consisted of the following backgrounds: 34 mechanical engineering, 1 product design, 1 architecture, 2 psychology, 11 other engineering or science (biology/bioengineering, industrial/nuclear/chemical engineering, computer science, math), and 7 unspecified.

3.3 Study Procedure

The study was conducted in two stages where the first stage (Group A) was primarily used to collect a set of human-generated comparisons to use in the second stage. Fig. 2 shows the outline of the study.

3.3.1 Conditions (varying interactivity and prior data)

Participants made decisions over two sets of 30 pairwise trials, where they were asked to select the option that they perceived as more comfortable to hold, with the option to provide a small modification based on directional change of parameters. At each new trial after the first, one new design was shown (based on active querying, which is outlined later in Section 3.4) in comparison to the participant's previously selected or modified design. The number of trials was determined based on prior work [50], though it is possible fewer queries could be used [47, 48]. The difference between the two sets of trials was solely the data used to initialize the model. In one condition, the model had no initial data and was updated using only decisions that the participant provided in the moment (referred to as Baseline). In the other condition, the model was initialized with simulation-generated or human-generated comparisons, and then updated using the participant's in-the-moment decisions (called Initialized). No time discounting or weighting of more recent answers was included. The condition order (Baseline or Initialized first) was randomly determined when the participant started the study. The third condition (called Aggregate) involved a model that was updated offline with the simulation or human-generated comparisons, but not updated with any participant-specific data, representing a case where participants are unable to directly interact with the model.

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

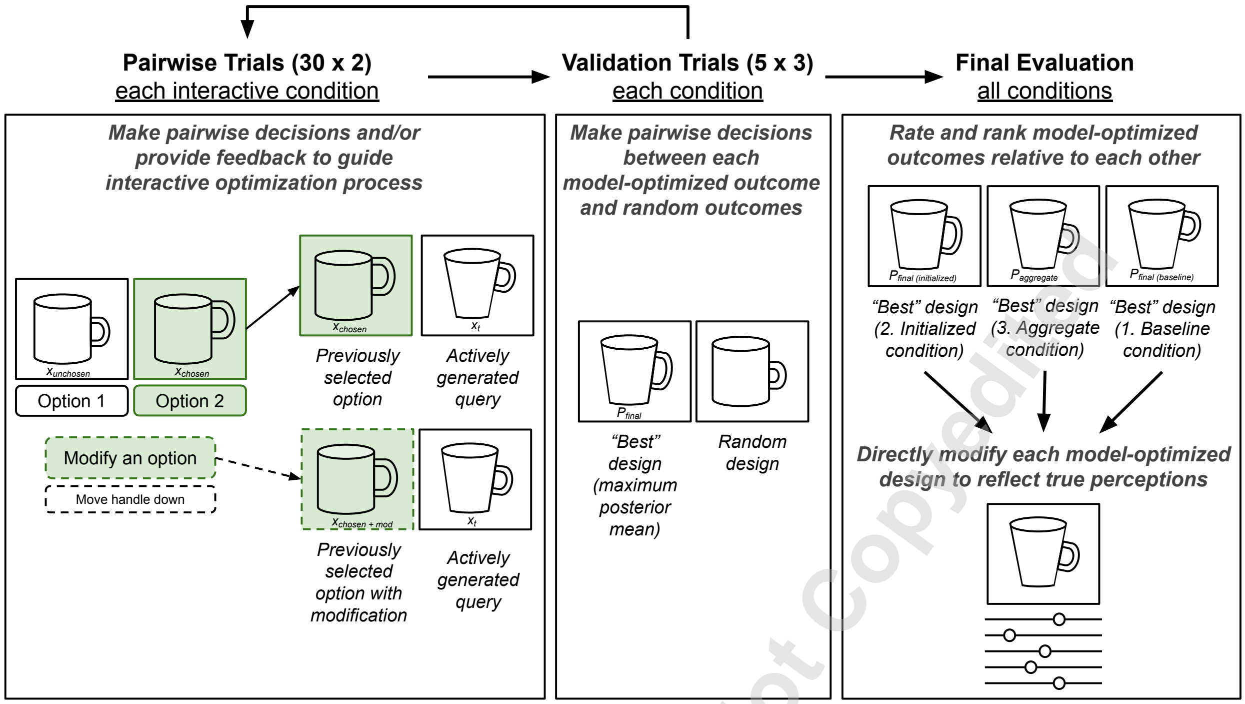


Fig. 2: Study procedure followed by each participant (corresponding to Alg. 1). The pairwise trials were repeated twice in a random order, once with a model containing only individual data and once with a model containing both aggregated and individual data. After each set of pairwise trials, participants completed a set of validation trials, and then a third set of validation trials for the outcome of the non-interactive model with only aggregated data. Finally, three outcomes were compared: the result of the interactive models (1. Baseline and 2. Initialized) and the result of a non-interactive model (3. Aggregate).

3.3.2 *Evaluation*

After each set of pairwise trials, during which the model was updated and queries were actively presented, participants were asked to choose between the comfort-optimized design (P_{final}) and a random design for a set of 5 validation trials. For all validation trials, these random designs were generated based on random numbers from a uniform distribution on the interval 0 to 1 for each normalized parameter. Participants were not explicitly aware of the transition between model updating phase and validation phase. At the end, participants also completed 5 validation trials for the third condition (Aggregate) where they selected between the maximum posterior mean design from the Aggregate model (evaluated over a line L and observed data D), $P_{aggregate}$, and a random design.

After all trials were completed (75 in total), participants compared the comfort-optimized designs (P_{final} for the interactive models and $P_{aggregate}$ for the offline model) from each condition to each other, providing a rating and ranking, with ties allowed. Then, they were allowed to indicate any changes they would make to each design using sliders along each parameter. Finally, they were directed to a survey where they answered several questions about their decision making.

3.3.3 *Groups (initializing with human-generated or simulation-generated data)*

Since human-generated comparisons were not available prior to collecting data from Group A, the Aggregate and Initialized conditions (which rely on baseline aggregated data) were initialized with simulation-generated comparisons. In this simulation, first, two perceptually different mug designs that humans may deem "comfortable" were manually selected. For each simulated participant, one of these two "comfortable" designs was randomly chosen as the participant's optimal design. The preference comparisons (i.e., the result of the simulated participant's utility func-

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

Table 1: Data used to initialize and update the models in each condition (1, 2, 3) and for each group (A, B). 1 and 2 result in varying comfort-optimized designs across participants and groups while 3 only results in varying comfort-optimized designs across groups.

Model (Condition)		Initializing Data (Group)	Updating Data
1. Baseline	Interactive	(A) None	Pairwise trials
		(B) None	
2. Initialized	Interactive	(A) Simulated	Pairwise trials
		(B) Real (from Group A Baseline)	
3. Aggregate	Offline	(A) Simulated	None
		(B) Real (from Group A Baseline)	

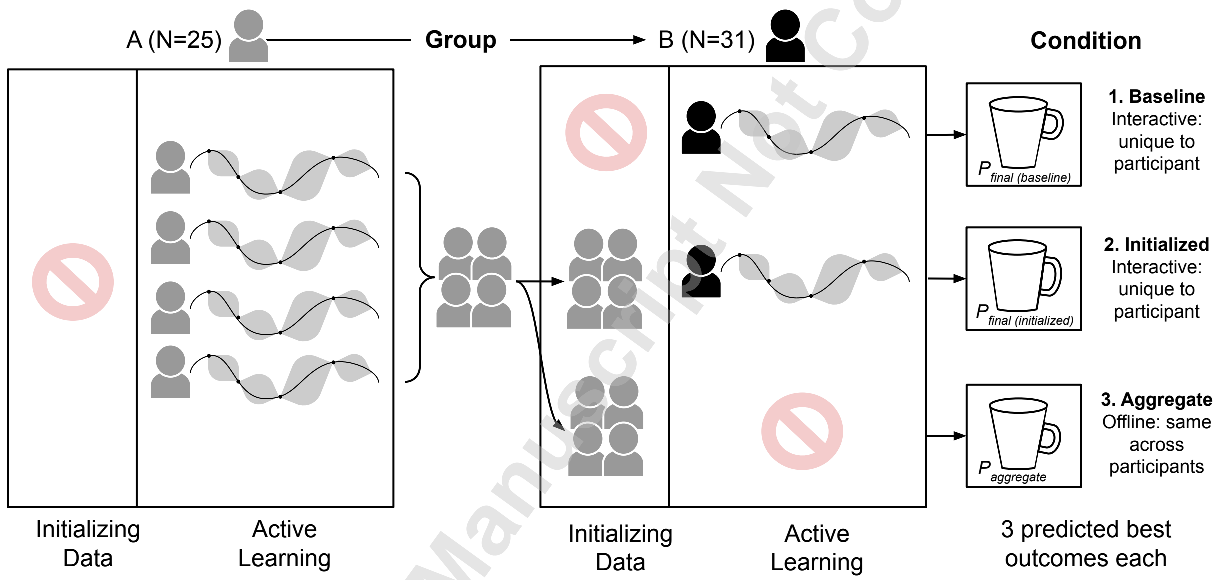


Fig. 3: Procedure for incorporating real human data from Group A to initialize models for Group B. The preference comparisons from the Baseline condition of Group A are combined to generate the Aggregate outcome ($P_{aggregate}$) and as the initializing data for the Initialized condition in Group B.

tion) were generated based on a minimizing the distance to that chosen design, with some amount of noise (i.e., the simulated participants' decisions do not always perfectly represent their utility functions, as is the case with humans). Data from 25 participants was collected in this stage (Group A). Although a group of 25 may not be sufficient to represent a true "crowd", increasing the amount of initial data in the interactive models increases the computation time per trial and therefore, investigation of crowd size is left to future consideration. The data from the Baseline condition of Group A (pairwise decisions that were not guided by any simulated data) was used to initialize models in the second stage with real human-generated data before online learning. Then, data from 31 additional participants was collected using the real data (Group B). The three conditions are summarized in Table 1 and the use of data from Group A to initialize the models for Group B is shown in Fig. 3.

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

3.3.4 Interface

The custom web interface (developed using Flask and hosted on a Google Compute Engine virtual machine with 8vCPU and 16GB memory) for the pairwise trials is shown in Fig. 4. Inspired by the interface in [28], participants could see instructions, followed by side-by-side 3D representations of the two designs being compared. Each 3D representation was dynamically rendered (using OpenJSCAD and three.js) during the data collection, similar to [22], and could be rotated and viewed from any angle if desired. There were two buttons to select either option and a third button to provide a design modification. The third button revealed 12 higher-level options for this feedback (corresponding to increases or decreases in parameter values), shown in Fig. 4. Before the pairwise trials, participants had the opportunity to explore the design space to better understand the meaning behind these feedback options. After the pairwise trials, participants conducted a simultaneous rating/ranking and sequentially, in a random order, indicated any modifications they would make to the optimal designs. Finally, participants were directed to a survey to answer several questions about themselves and their decisions.

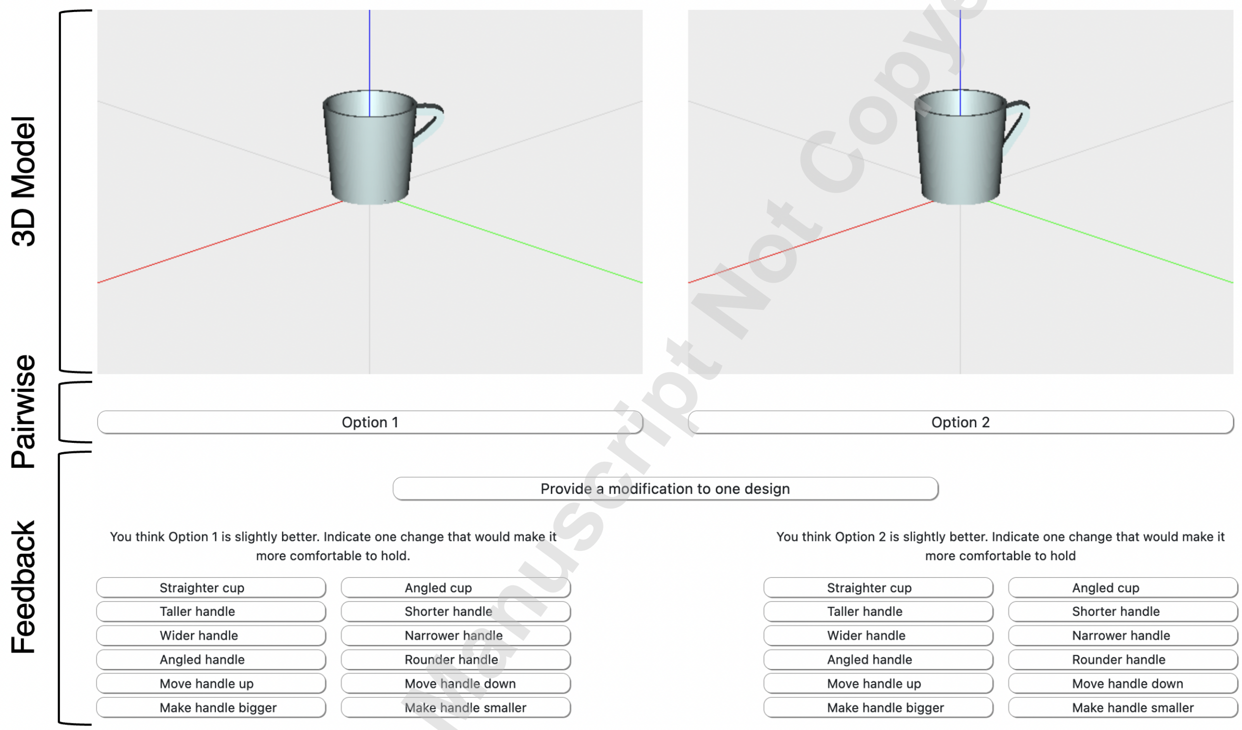


Fig. 4: Participants made pairwise choices using the interface above. If neither design was perceived as more comfortable to hold, a third button revealed choices to suggest a modification to one design. Instructions, including how to suggest the modification and the task (*"Select the one you perceive as more comfortable to hold"*), were presented at the top of the screen throughout the study.

3.4 Interactive Optimization

3.4.1 Gaussian Process model

A Gaussian Process model is a surrogate model, specifically a multivariate Gaussian, specified by a mean function and a covariance kernel. The GP used to model the pairwise queries in our study was specified by Chu and Ghahramani, and has been commonly applied to preference learning tasks [59]. Using a probit likelihood, binary observations can be used to infer a latent function (in our case, the participant's perception of the subjective attribute related to the

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

design). Based on Bayes' theorem, the posterior probability function, which is the probability of a function (f) given the data (D), is

$$P(f|D) = \frac{P(f)}{P(D)} P(D|f). \tag{1}$$

In this case, the data is in the form of pairwise preferences ($D = v_k \succ u_k : k = 1, ..., n$) where $v_k \succ u_k$ refers to instance v_k being preferred to u_k. $P(D|f)$ is then

$$P(D|f) = \prod_{k=1}^{n} P(v_k \succ u_k | f(v_k), f(u_k)). \tag{2}$$

The probability in the likelihood above is 1 if $f(v_k) \geq f(u_k)$ and 0 otherwise, in the ideal case, but a more tolerant formulation assumes that the latent functions are contaminated with noise that follows a Gaussian distribution. Therefore, at each pairwise decision, the model can maintain an estimate of the participants' utility function, with uncertainty, over a set of points. The maximum posterior mean, the point that maximizes the mean of the estimated functions, can be used to approximate the "best" point throughout the optimization process. An implementation from the BoTorch Python library was used (with default prior parameters) to fit the model and sample from its posterior at each step[1]. The BoTorch implementation uses a Laplace approximation of the posterior and a radial basis function kernel (also known as the squared exponential kernel) as the covariance function [60]. This kernel is commonly used because of its flexibility when modeling smooth and continuous functions with the assumption that points close to each other have similar values and has been used, along with its variants, in various domains including preference learning in engineering design [55, 50, 56]. This assumption is reasonable for the given task, as sharp discontinuities are not expected over small design variations (i.e., it is unlikely for two designs with very similar parameters to have a drastic difference in value to the participant).

3.4.2 Active query generation

There are several options for actively determining the next query to present to users in order to efficiently model their preference decisions. Our approach (Alg. 1) was adapted from the algorithm used by Tucker et al. [50], which is based on Thompson sampling and one-dimensional subspaces. Similar line-search approaches have been utilized in other domains such as visual design [47]. While other common acquisition functions (e.g., expected improvement or upper confidence bound) were considered, this particular approach was chosen due to its tractability for balancing exploration and exploitation at higher dimensions and variable levels, which is particularly important when considering continuous geometry. The hyperparameter (m) for discretizing the design space was set to be as small as possible while maintaining a reasonable evaluation time for each trial ($m = 0.005$, resulting in 200 possible values for each normalized parameter from 0 to 1). Furthermore, the approach can be easily augmented by co-active feedback [23], which provides an alternative to direct pairwise selection to improve data quality by mitigating cases when people are unable to perceive small visual differences. This type of feedback was incorporated in our study through eliciting design modifications based on directional change (straighter or more angled cup, taller or shorter handle, wider or narrower handle, move handle up or down, make handle bigger or smaller), though these descriptions may be difficult to specify in more complex cases. The modification enacted by these feedback options was a 10% increase or decrease in the single parameter value (or two parameter values in the case of two feedback options: increasing or decrease "handle size"). Participants were instructed to make modifications with reference to both designs ("If providing a modification, please select it on the side of the option you prefer more"). Thus, if feedback was given, it was incorporated as a preference over both of the designs, leading to a slightly increased impact of feedback (adding two preference comparisons to the data instead of one). No additional weighting was included to more heavily weight direct feedback. Bounds were included in this study for simplicity of normalizing inputs for the model. In our specific implementation, if the feedback was out of bounds, the preference was recorded as the reverse with reference to the

[1] PairwiseGP from https://botorch.org/

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

side that was selected. The feedback mechanism could remove the need for the design space to be bounded strictly if alternative approaches are found for normalization. There are a couple limitations of the active querying method implemented. First, it requires the variables to be continuous, which is not always the case for complex design spaces. Second, there is the possibility for repeat comparisons if the model does not find a better query point along the randomly selected line, which was not accounted for in our study.

Algorithm 1 Interactive optimization via pairwise decisions based on [50] (notation kept as similar as possible).

$D = \emptyset$ or $D_{aggregate}$ ▷ D: Preference comparison data
$P = \emptyset$ or $P_{aggregate}$ ▷ P: “Best” design based on model
$W = \emptyset$ or $W_{aggregate}$ ▷ W: Observed designs

Present the first comparison using random points ▷ Initialization
if modification feedback provided **then**
    Add modification comparisons ($x^{chosen+mod} \succ x^{chosen}, x^{unchosen}$) to D ▷ Exceptions at bounds $[0,1]^5$
    Set $W = W \cup x^{chosen+mod} \cup x^{chosen} \cup x^{unchosen}$
    $x_0 = x_{chosen+mod}$
else
    Add pairwise preference ($x^{chosen} \succ x^{unchosen}$) to D
    Set $W = W \cup x^{chosen} \cup x^{unchosen}$
    $x_0 = x_{chosen}$
end if
if $P = \emptyset$ **then**
    $p_1 = x_0^{chosen}$
else
    $p_1 = P$
end if

for $t = 1, 2, ...n$ **do** ▷ Interactive trials
    L_t = random line through p_t, discretized via m ▷ $m = 0.005$
    $V_t = L_t \cup W$ ▷ Points to update posterior over
    (μ_t, Σ_t) = posterior over points in V_t, given D
    Sample utility function $f_t \sim GP(\mu_t, \Sigma_t)$
    Show new design $x_t = argmax_{x \in V_t} f_t(x)$
    if modification feedback provided **then**
        Add modification comparisons between x_{t-1}, x_t, and $x^{chosen+mod}$ to D
    else
        Add pairwise preference between x_t and x_{t-1} to D
    end if
    Update W with observed designs
    Set $p_{t+1} = argmax_{x \in V_t} \mu_t(x)$
end for

$P_{final} = p_{n+1}$ ▷ Optimized outcome after n trials

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

4 RESULTS

The human subjects study resulted in a single "comfort-optimized" design from aggregated data (outlined in Section 4.1), as well as several individualized "comfort-optimized" designs, produced from each participant's interaction with the optimization model. These interactive optimization models were provided different initial information, either none, simulated aggregate data (N = 25), or real aggregate data (N = 31), to guide the optimization process. Outcomes from both the interactive optimization and the non-interactive aggregate model were compared to understand if and how computational representations of subjective attributes were aligned with participants' perceptions. Interaction data from Group A was analyzed in Section 4.1 and Section 4.2.1 only. Data from one participant in Group A was removed for the analysis in Section 4.2.1 due to a data recording error during the validation trials only. Only the interaction data from Group B (N=31) was considered in all other analysis with no participants removed.

4.1 Collecting the human-generated comparisons to construct the aggregate outcome

The first stage of the study (Group A) was used as a way to collect real data from participants that could be used to both initialize the models in the second stage and generate, offline, the aggregate comfort-optimized design ($P_{aggregate}$). The diagonal of Fig. 5 shows the parameters of the comfort-optimized outcomes generated from each participant's individual data in Group A. Although the pairwise comparison data, not these final designs, were used in the second stage, the outcomes show a tendency towards lower values for the cup angle, higher values for handle length, and lower values for handle location. Each Group A individual's pairwise comparison data from the Baseline condition (not influenced by any prior data) was combined to determine the estimate of participants' aggregated utility function regarding comfort (with respect to the design parameters). Though this is a 5-dimensional function in reality, it is visualized as 2D and 3D surfaces for pairwise combinations of the design parameters in Fig. 5, where the z-axis (surface plots) and lighter color (contour plots) indicates a higher utility for that parameter combination. The overall outcome based on this utility function, and used for the Aggregate condition in Group B is shown in Fig. 6.

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

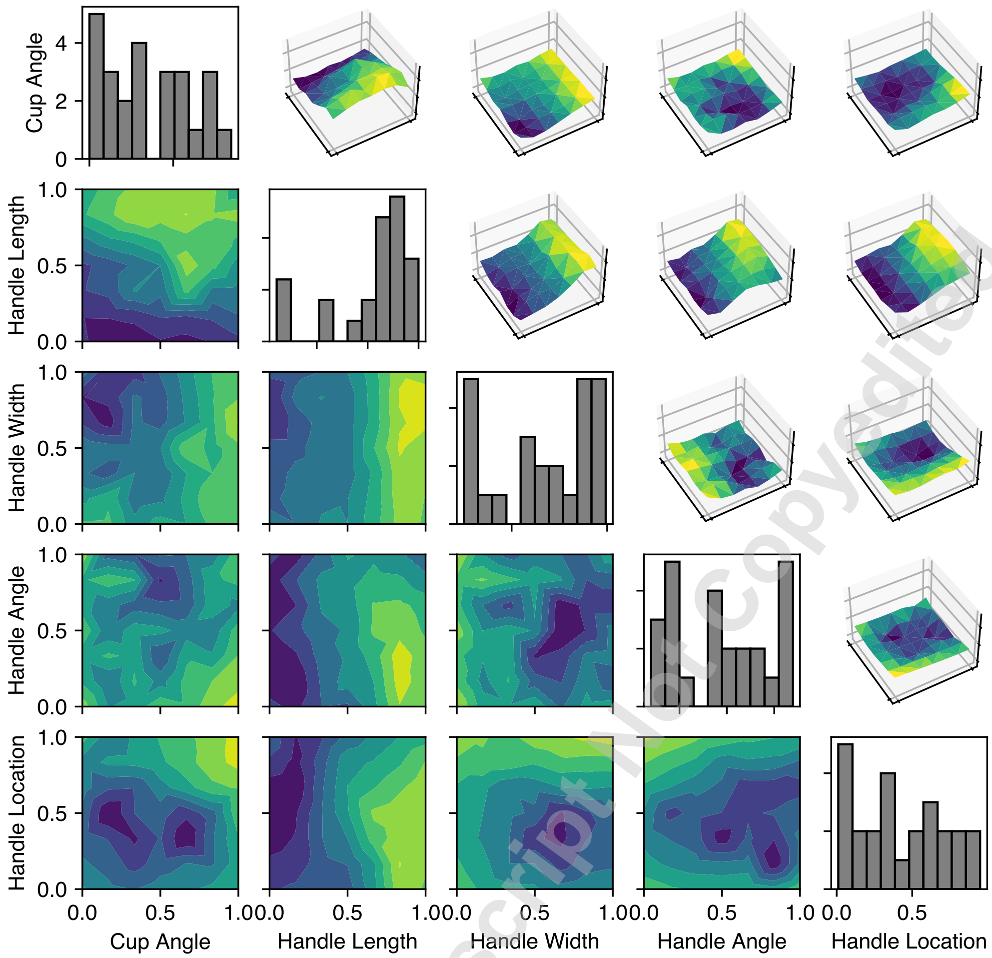


Fig. 5: Visualization of the parameters of the optimized designs from the Group A (1) Baseline condition (diagonal) and the utility function generated from aggregating the data, used for the (2) Initialized condition and (3) Aggregate condition in Group B

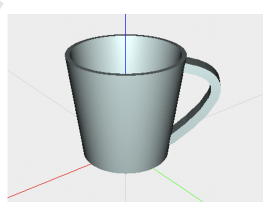


Fig. 6: Aggregate comfort-optimized design generated based on data from Group A with normalized parameter values: [Cup Angle: 0.326, Handle Length: 0.960, Handle Width: 0.776, Handle Angle: 0.508, Handle Location: 0.547]

4.2 Evaluating the quality of and differences between model-generated outcomes

The outcomes – each participant's unique comfort-optimized design from the two interactive models and the aggregate comfort-optimized design for the one non-interactive model – were evaluated based on how they satisfied the task sufficiently (via a hit rate obtained separately for each condition) and how they differed across participants and conditions (via distance comparison).

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

4.2.1 Hit rate reveals success finding perceptually aligned outcomes using interactive and non-interactive models

The hit rate refers to how often a participant selected the model-predicted comfort-optimized design vs. a random design for the five validation comparisons. Therefore, the hit rate can help demonstrate whether a model can achieve a comfort-optimized outcome relative to the rest of the possible design space. A generalized linear mixed model (GLMM) was fit to evaluate the hit rate, with an outcome predicting whether the model's optimal design was chosen or not during the validation trials. The predictors are each condition (Aggregate, Initialized, Baseline) and group, which specify the type of aggregate data the optimization models were initialized with (Simulated or Real), as well as the distance between the model-optimized design and the random design that was presented at each trial.

The GLMM reveals that in the non-interactive Aggregate condition, when real data is used, the model-optimized design is selected more than the alternate ($b = 1.40, p = 0.008$). This indicates that the non-interactive model is able to successfully capture baseline perceptions of the subjective attribute. There are two significant negative predictors for selecting the model-optimized design which show that the model-optimized design is selected less for the non-interactive model with simulated data vs. real data ($b = -1.28, p = 0.005$) and in the interactive Initialized condition seeded with simulated data compared to the non-interactive Aggregate condition seeded with real data ($b = -2.08, p < 0.005$). These results indicate that participants were less likely to select the optimized outcomes when interacting with a model containing our simulated preference comparisons. The Baseline condition is not a significant predictor ($b = -0.11, p = 0.770$) of selecting the model-optimized outcome compared to the non-interactive Aggregate condition (i.e., no sign of participants treating outcomes from these models differently). However, the model-optimized design is selected more in the Initialized condition (using real data and interactive) than in the Aggregate condition (using real data but non-interactive) ($b = 2.87, p < 0.001$).

At the participant-level, the median hit rate, a proportion of the 5 validation trials, is 1 for all conditions in Group B (Real). Only the minimum hit rate differs, with the lowest proportion in the Aggregate condition and the highest proportion in the Initialized condition (Baseline: $Range = [0.4, 1.0]$, Initialized: $Range = [0.6, 1.0]$, Aggregate: $Range = [0.2, 1.0]$). These results demonstrate that the model is able to output designs that are aligned with perceptions generally, with some conditions leading to even better outcomes.

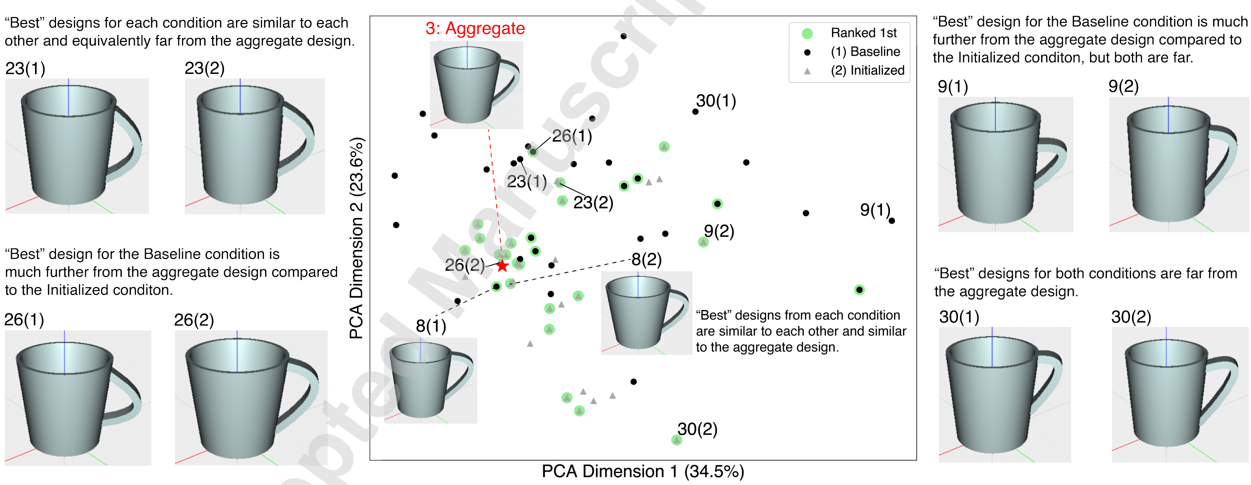


Fig. 7: Visualization and examples of comfort-optimized ("best") design outcomes from participants (8, 9, 23, 26, and 30) in Group B. Outcomes are shown for both sets of pairwise trials: the (1) Baseline condition and the (2) Initialized condition (2) vs. the (3) Aggregate condition. The highlights in green mark if that design was rated by the participant as the design that most aligned with their perceptions of comfort, with ties allowed.

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

Table 2: GLMM modeling the selection of model-optimized designs vs. random designs for the validation trials of each condition and group.

	Selecting Model-optimized Design (vs. Random)	
Predictor	*Log-Odds (CI)*	*p*
(Intercept)	1.40** (0.37 - 2.42)	**0.008**
group [real]	*Reference*	
group [sim.]	-1.28** (-2.16 - -0.39)	**0.005**
condition [aggregate]	*Reference*	
condition [initialized]	1.16* (0.17 - 2.16)	**0.022**
condition [baseline]	-0.11 (-0.87 - 0.65)	0.770
condition [initialized] * group [sim.]	-2.08*** (-3.25 - -0.92)	$<$ **0.001**
condition [baseline] * group [sim.]	1.67** (0.55 - 2.79)	**0.003**
distance between choices	1.31** (0.32 - 2.29)	**0.009**
Random Effects		
σ^2	3.29	
τ_{00pnum}	0.91	
ICC	0.22	
N_{pnum}	55	
Observations	825	
Marginal R^2 / Conditional R^2	0.232/0.399	$^{*}p < 0.05,^{**}p < 0.01,^{***}p < 0.001$

4.2.2 Outcomes from interactive models reflect individual differences in perceptions of "comfortable-to-hold"

The "comfort-optimized" designs were further analyzed to understand how the interactive model could lead to different types of outcomes. The final outcomes that result from the two interactive conditions (1 and 2), compared to the Aggregate outcome (3), which involves no intervention of the individual participant, are shown in Fig. 7 as a 2D projection of the 5-dimensional design space (only for visualization). The highlighted points are those that were ranked as most-aligned to participants' perceptions when they were completing their final evaluation, while the examples demonstrate the type of visual diversity found in participants' final outcomes.

As participants were not informed that they were experiencing different conditions, with all trials presented in the exact same way, it is expected that differences are influenced by the initialization data provided to the model or the difference of an individual's preference (e.g., an "extreme" vs. not) from the Aggregate outcome. The Euclidean distance between each individual outcome and the Aggregate outcome demonstrates a measure of this difference for both the Baseline ($M = 0.67, SD = 0.21, Range = [0.19, 1.14]$) and Initialized ($M = 0.40, SD = 0.29, Range = [0.08, 1.01]$) conditions. As expected based on the nature of the model and the data it is provided in each condition, the average distance from the Aggregate is higher for the Baseline condition than for the Initialized condition. There is no evidence of a correlation between this Euclidean distance and hit rates or ratings for the outcomes of the interactive conditions. Thus, there is no evidence that distance from a baseline outcome (i.e., reaching outcomes that are "far" from the aggregate-level outcome) is associated with decisions about or perceptions of interactively-optimized outcomes. While Euclidean distance was selected for simplicity in comparing the designs in the normalized parameter space, distances in the perceptual space, which are non-trivial to determine, may differ and therefore, could potentially

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

influence the hit rate or alignment ratings.

4.2.3 Parameter-level analysis to reveal driving factors of individual differences

The optimization process modeled the outcomes considering combinations of parameters, but the designs can also be examined at the parameter level to investigate where variations were most prevalent and whether certain features drove individual differences. Fig. 8 shows the parameter distribution of the optimized outcomes of Group B while Table 3 shows the standard deviations of the normalized parameters to demonstrate which parameters vary across participants. The parameter that differs the least is handle length ($SD = 0.21$ and $SD = 0.05$) in both cases, indicating that this feature (i.e., a longer handle) was common across the final outcomes for all participants. The parameter with the highest standard deviation among best designs is handle location ($SD = 0.36$) for the Baseline condition and handle width ($SD = 0.28$) for Initialized condition, demonstrating that these parameters show more variety across participants.

Table 3: Standard deviation of each parameter across Group B participants' best designs.

Baseline (1)		Initialized (2)	
Parameter	SD	Parameter	SD
Handle Location	0.36	Handle Width	0.28
Cup Angle	0.30	Handle Location	0.22
Handle Width	0.28	Cup Angle	0.18
Handle Angle	0.23	Handle Angle	0.18
Handle Length	0.21	Handle Length	0.05

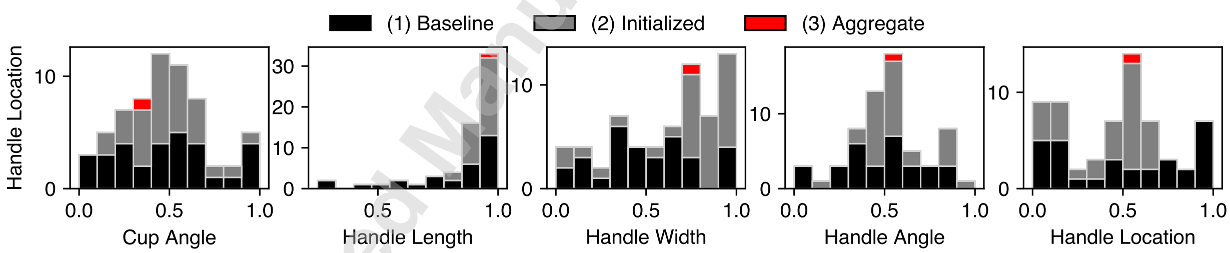
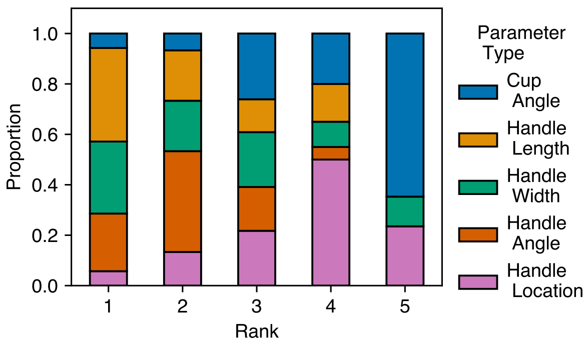


Fig. 8: Parameters of the optimized designs for Group B in the (1) Baseline condition and (2) Initialized condition vs. the (3) Aggregate condition, generated from real data from Group A

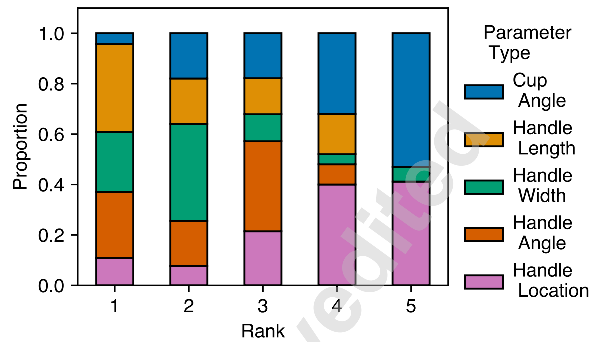
Self-reported rankings of how important each parameter was to the participant when making their decisions (ties allowed) are shown in Fig. 9, including both Group A and Group B. Participants demonstrate variety in their importance rankings. The parameter that is considered the most important by a plurality of participants is handle length. Correspondingly, the smallest standard deviation in outcomes for both conditions in Group B is the handle length, which demonstrates a common feature across the group. The handle width is considered the second most important by a plurality of participants in Group B. This feature also exhibits the highest standard deviation among participants' best designs from the Initialized condition. Thus, it is possible that varying the handle width may have driven many

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

of the individual differences, though again it is important to note that interaction effects can signify that the parameter combinations should be considered holistically.



(a) Group A



(b) Group B

Fig. 9: Self-reported rankings on the importance of each parameter for participants' decision making (1 = Most important).

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

4.3 Evaluating perceptions of the optimization process

The model-generated outcomes were compared relative to each other based on self-reported alignment with perceptions and compared to an outcome directly specified by participants, representing their "ground truth" perceptions. Additionally, the optimization process was qualitatively and quantitatively assessed based on information from the survey and how much the design space was "explored".

4.3.1 Outcomes from an interactive optimization approach are generally best-aligned with human perceptions

Ratings of design outcomes in each condition are shown in Fig. 10 for Group B, based on participants' answers to how well aligned each design was to their perception of a mug that is comfortable to hold. It should be noted that the ratings were completed by comparing each different condition directly and therefore also constitute a ranking. Ratings are highest for the comfort-optimized design from the Initialized (1: $Mdn = 5$, 2: $Mdn = 6$, 3: $Mdn = 5$). There are differences between the outcome ratings of the Baseline and Initialized conditions ($Mdn = -1, Range = [-5, 2]$) and Baseline and Aggregate ratings ($Mdn = -1, Range = [-5, 3]$), where Initialized and Aggregate tend to be rated higher Baseline. The ratings reveal that on average, removing any initializing data and starting the optimization process from scratch results in the worst alignment with perceptions. The difference between Initialized and Aggregate ratings ($Mdn = 1, Range = [-2, 3]$) shows that Initialized also tends to be rated higher than Aggregate.

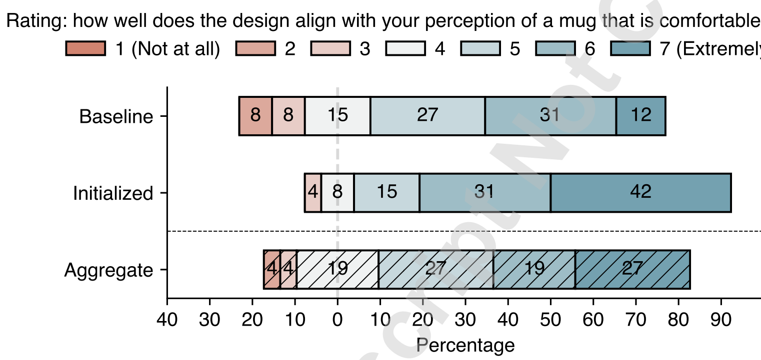


Fig. 10: Ratings for the Baseline ($Mdn = 5$), Initialized ($Mdn = 6$), and Aggregate outcomes ($Mdn = 5$) in Group B.

The validation trials and the ranking provide a relative comparison of how well each optimized design aligned with perceptions, while the rating provides an estimate of the absolute alignment with perceptions. To further investigate whether interactive models can lead to outcomes that align with "true" perceptions (at least within the scope of the limited design space in this study), participants were enabled to specify modifications to each of the three optimized designs to result in a design that would best align with their perceptions. As shown in Fig. 2, the participants could modify each optimized design using sliders that corresponded to the five parameters considered in the study. There was a reset button available (which reset the parameters to the optimized parameters) to ensure that participants' modifications reflected a design that they considered to be better than the optimized design. Fig. 11 shows a Euclidean distance between the participants' model-optimized designs and their manually modified designs, summarizing the overall changes participants decided to make to the optimized design to reflect their "true" perceptions. While participants indicated a relatively high level of satisfaction with the model-optimized outcomes in aligning with their perceptions, participants took the chance to modify their designs to exactly fit what they intended across all conditions, indicating the inability for the models to fully represent participants' perceptions. In line with the rating results, on average, participants appeared to make larger changes to the Baseline outcome to reach their intended result ($M = 0.36, SD = 0.24$) compared to the Aggregate outscome ($M = 0.21, SD = 0.21$) and Initialized outcome ($M = 0.19, SD = 0.17$). The largest changes were made in the Baseline condition ($Range = [0, 1.06]$), followed by the Aggregate condition ($Range = [0, 0.85]$), compared to the Initalized condition ($Range = [0.004, 0.59]$).

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

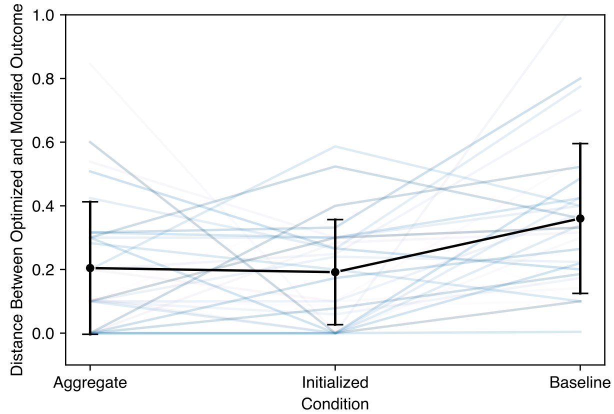


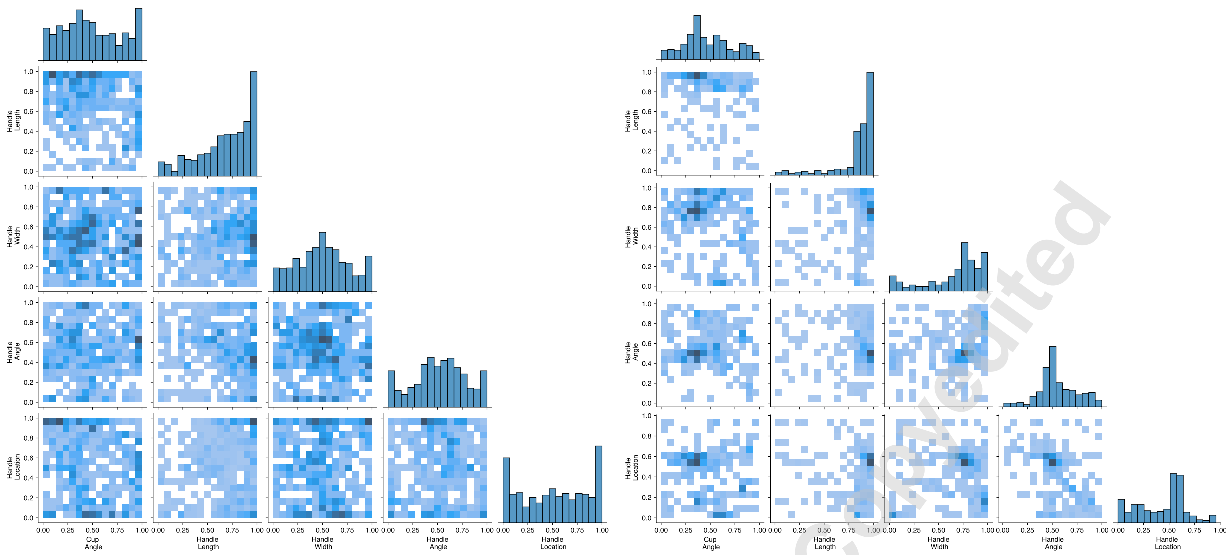
Fig. 11: The average amount of modification made to each of the model-optimized outcomes during final evaluation to reflect "true" perceptions, summarized as the normalized distance between the parameterized designs. The individual lines show participant-level data.

4.3.2 Exploration of the design space during interactive optimization

For each participant, the designs seen during the interactive trials (Baseline and Initialized) vary because of the active querying. Fig. 12 shows the design space visited by participants in Group B throughout the optimization process in the two interactive conditions (Baseline and Initialized). The generalized variance (determinant of the covariance matrix) of all designs that were visited throughout the pairwise trials, excluding any validation trials, provides quantitative insight into the extent the design space was explored during the optimization process. The generalized variance for designs visited by participants in the Baseline condition ($Mdn = 1.07 \times 10^{-8}$) is greater ($W = 84.0, p = 0.0013$) than that of the Initialized condition ($Mdn = 5.34 \times 10^{-11}$), using a Wilcoxon signed-rank test. The generalized variance of participants' final outcomes is similarly greater for the Baseline condition (9.72×10^{-7}) than for the Initialized condition (1.33×10^{-9}), as visualized in Fig. 7. It is unsurprising that the Baseline condition tends to query from a wider range of designs due to a lack of bias towards a specific area of the design space from prior data. Overall, since the Initialized condition allows participants to start from a similar "group-level" design, it follows that the spread of designs visited and the corresponding diversity of outcomes for this condition is lower.

The impact of this broader exploration can be explained by looking more closely at the decisions made during the optimization process. Specifically, the nature of the optimization process involved the presentation of a new design at each iteration that was the model's estimation of "most preferred," under uncertainty. However, because each pairwise comparison was made between a previously selected option and the new option, anchoring may impact decisions (i.e., whether to switch or stay). Fig. 13 shows, on average, the distance between comparisons that result in switching to a new design vs. staying (not considering feature modification). Fig. 13 indicates that larger distances (i.e., the new option is more different in the parameter space) tend to precede a stay by the participant. This observation is statistically significant for both conditions, but more pronounced in the Baseline interactive condition ($W = 21.0, p = 8.65 \times 10^{-6}$) than the Initialized interactive condition ($W = 147.0, p = 0.048$). Participants may be biased towards staying with a previously selected option particularly when new alternatives are very different from their current design. In the Baseline condition, the lack of prior data can additionally lead to actively queried designs that are perceptually unsuitable, particularly in earlier trials.

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687



(a) Baseline

(b) Initialized

Fig. 12: The parameter values visited during the optimization process by Group B in the interactive conditions. The generalized variance of designs visited in the Baseline condition ($Mdn = 1.07 \times 10^{-8}$) is greater ($W = 84.0, p = 0.0013$) than that of the Initialized condition ($Mdn = 5.34 \times 10^{-11}$).

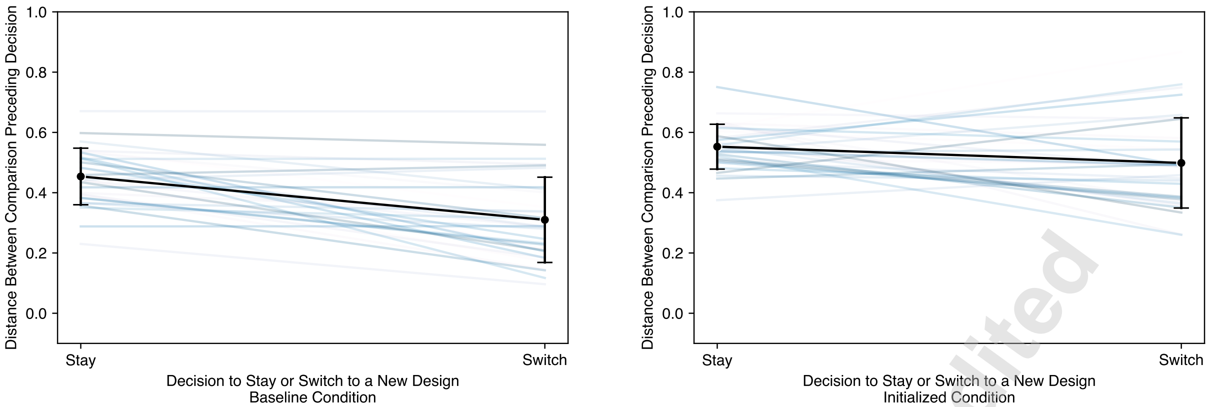
4.3.3 *Impressions of the interactive optimization process demonstrate benefits and challenges*

Free response answers after the study reflect broad impressions of the interactive optimization process, which may have affected satisfaction during the optimization process and subsequently satisfaction with model-optimized results. For example, one participant notes that their preferences were reflected by the updating model in some aspects, but that changing their mind part of the way through made it difficult for the model to capture their final intent exactly:

> *"Having a specific handle shape in mind made selections of Option 1 vs. 2 really easy. It was hard to see the small changes that were made when features were controlled with the "taller handle," etc. buttons. I feel that I might have confused the model a little bit into making it believe that I wanted a larger handle after I changed my mind about that 1/4th of the way through the study. For the most part, the model was pretty spot-on with my preferences for cup angle and handle angle but kept on giving me cups with strange handle lengths for the duration of the study."*

Additionally impressions are shown in Table 4. Positive perceptions of the optimization process appeared to result from being able to visually identify that the active querying method was providing alternatives that were more closely aligned with perceptions. In particular, allowing modifications to provide co-active feedback was a mechanism that allowed participants to immediately see the impact of their decisions on future choices. Participants struggled with being provided choices that were too visually similar, representing a potential mismatch between the data that interactive models need to perform optimally compared to what humans are able to provide. At the same time, participants also expressed confusion towards instances where they were presented with seemingly random, completely different designs. In this study, part of this can be attributed to the validation trials that participants were unaware of. However, this leads to important considerations regarding how interactive computational processes would have to be developed in order to properly capture human perceptions (in this case, considering perceptual similarity). Additional factors such as preference reversal and consistency remain a challenge that is not addressed by the interactive models utilized here, although there have been efforts to try to incorporate preference change over time [61].

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687



(a) The median distance (0.44) preceding a stay is greater than the median distance (0.29) preceding a switch ($W = 21.0, p = 8.65 \times 10^{-6}$)

(b) The median distance (0.54) preceding a stay is greater than the median distance (0.49) preceding a switch ($W = 147.0, p = 0.048$)

Fig. 13: The average distance between comparisons that resulted in switching to a new design vs. selecting a previously selected design again for the Baseline condition (left) and the Initialized condition (right). The individual lines show participant-level data.

5 DISCUSSION

To allow computational methods to better align with human perception, it is important to understand how human perception can be best embedded into optimization processes. In this work, we explore how interactive optimization methods lead to outcomes that are well-aligned (or not) with humans' perceptions of the subjective attribute of comfort. We also identify how different outcomes might arise when aggregating this perception across a group of individuals. We find that commonalities in human perception within a group can be leveraged to generate outcomes that are well-aligned with the intended subjective attribute and that addressing subtle differences of perceptual decisions through interactivity can be beneficial for even better alignment. However, factors related to the optimization approach, such as queries that are too visually similar or confusion around why the model presents very dissimilar designs (e.g., to balance exploration and exploitation), demonstrate remaining challenges of leveraging human-in-the-loop systems for design preference elicitation.

5.1 Interactive optimization can capture human perceptions of a subjective attribute and improve satisfaction with an optimized outcome

The first question addressed is: to what extent do interactive models guided by iterative user feedback improve alignment with individuals' perceptions compared to non-interactive models? The results of this study indicate that the interactive models are able to produce outcomes that capture perceptions of the subjective attribute considered here reasonably well, based on the tendency to select model-optimized outcomes over random outcomes (hit rate). In particular, analysis of hit rate provides support for the benefit of interactive over non-interactive models, given real initialization data. Furthermore, alignment of individualized outcomes with perceptions of comfort is rated relatively highly by participants (medians of 6 and 5 out of 7 for the Initialized and Baseline conditions respectively), with the Initialized condition tending to be rated the highest of the three conditions.

The results of Initialized condition also addresses the second question: how does prior data affect the efficiency and outcome quality of interactive optimization for individual users? When the initialization data reflects real human decisions, including this information helps improve upon the individualized models, which already perform relatively well. However, when the data is simulated and does not match how humans make real decisions (e.g., prioritizing a specific design attribute like handle length), including aggregate data can be detrimental and participants are unable to reach an area of the design space that reflects their perceptions. This is supported by the hit rate analysis, which finds a negative impact of initializing with simulated data on selecting the optimized outcome Therefore, interactive models

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

Table 4: Examples of responses to "Briefly describe any perceptions you had about the impact of your decisions/feedback on the model during the study"

Category	Responses
Impressions that decisions impacted the optimization process positively	I did notice the modifications I was making being implemented at a certain point, and it felt like the middle part of the study was trying to get me to a more narrow range of designs whereas the beginning and end showed me a very big range of designs. [P1]
	Not sure. I appreciated that I was able to see physical changes in the models that reflected my adjustment suggestions [P2]
	The options slowly changed to cater to my needs, until it offered a model that was outside of what I considered. [P3]
	Other times, the model was trying to figure out a quality (like handle length) and gave very different models until it understood my preference. [P4]
Impressions that decisions did not impact process	I didn't feel like my decisions were affecting the model at all. Each pair seemed just as random as the previous. [P5]
	I was confused because I couldn't really see any of my modifications in the reloaded cup designs. [P6]
Dissatisfaction with the optimization process	It seemed the model would go through periods of very slight changes, then suddenly give a wildly different and [uncomfortable] design after I gave feedback on what changes I want, either opposite to or overly fitting to that feedback. [P7]
Difficulty expressing preference due to visual similarity	At times I felt like the two models were the same and it was hard to decide which was better, at which point I looked at the cup angle which I felt like was less noticeable. [P1]
	Sometimes the cups looked very similar and I would get stuck in a loop of deciding between very similar cups. [P4]
	The cups started blurring together also all the designs looked so similar. [P6]

can fail if initialized with data that does not represent human perceptions well. Prior work implies through simulated experiments that changing the initial guess based on similar users is only valuable when the optimally-preferred designs are clustered [34]. Our empirical results support this based on the difference in results from initialization with simulated vs. real data. In general, the example considered here and bias in the participant pool likely induces more subtle individual differences in perception, whereas different examples may elicit larger differences that may lessen the benefits of initializing with aggregate data.

Design-relevant computational approaches such as semantic shape editing rely on the creation of large aggregate mappings of semantics to geometries [25]. Here, we find that such an approach can work well, but personalizing the semantic mappings can capture subtle differences across individuals' perceptions. Prior work within the design field has successfully moved towards both crowd-based and adaptive, personalized methods in the context of inspirational stimuli [62, 63]. In this study, the interactive models are able to lead to satisfactory outcomes with relatively few queries. Results show that incorporating real group-level data into an interactive model (Initialized condition) can lead to the best alignment with participants' perceptions, as shown by both decisions and self-reported measures on average. Table 5 shows the ranking of the "best" outcome from each model (ties allowed), obtained from the comparative ratings. Notably, there were still participants who ranked outcomes from other conditions (Baseline and Aggregate) the highest. In fact, 25% (8 of 31) of the participants rank the result of the Baseline condition as the most aligned with their perception of comfortable. While not the majority, this subset could be important to considered for application

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

of personalization if their perceptions diverge, as they may not be satisfied with outcomes that result from models with a bias towards aggregate data. Highly individualized models (i.e., the Baseline condition) may be more useful when there are highly diverging views of the dimension (e.g., for particularly abstract concepts) than in the attribute considered here. This individualization must be balanced with the cost of collecting data from a broader group, which is the greatest for the Initialized condition and the lowest for the Baseline condition. Furthermore, it is notable that even the Initialized condition does not align perfectly with true perceptions, demonstrated by participants' tendency to modify the outcome designs when given the opportunity to do so at the end. This type of behavior supports the need for interactive methods to computationally express abstract semantic attributes, but also indicates that improved feedback methods are necessary to better align interactive preference learning to perceptions.

Table 5: Frequency of rankings for "best" outcomes

	Model		
Rank	*Baseline*	*Initialized*	*Aggregate*
1	8	19	14
2	9	11	7
3	14	1	10

5.2 Model behavior and interaction with models impacts their use for embedding into human-centric applications

Analyzing the interaction process addresses the final question: what behavioral patterns and interaction strategies (e.g., design space exploration and decision-making) emerge during interactive optimization, and how do they shape satisfaction with the final designs? More of the design space is explored during the Baseline condition compared to the Initialized condition, but based on the hit rates and ratings, less exploration does not have a negative impact if the starting point is relatively aligned with human perceptions. In fact, the presentation of vastly different and seemingly "random" designs can be detrimental, based on behavior (i.e., the tendency to stay rather than switch) and comments about the interactive optimization process. It appears that the benefit of the Initialized condition is its ability to reach a general consensus of a comfortable mug and then allow adaptation to more subtle user preferences. This adaptation is likely enabled by guidance towards parameters that are more important from prior data, which is unavailable when the model starts from scratch in the Baseline condition. In a study comparing human and optimizer-led design, Chan et al. find that though performance-related outcomes can be improved, people lose agency and ownership when they are being guided by an optimizer [64]. Some of this can be mitigated by allowing people to provide co-active feedback, like in the form of design modifications here. People may have been more satisfied with outcomes because they had the option for active guidance rather than only passive evaluation. People may also feel frustrated if they do not understand Bayesian optimization, which trades off exploration and exploitation [64]. Thus, it may appear the optimizer is giving worse examples when it is simply trying to gain more information. Prior work also notes that human steering can impact optimization if information about how the optimization process works is provided [65]. Open-ended comments from the survey in our study indicate that some participants felt like they could see the impact of their decisions and feedback throughout the optimization process, while others felt frustrated if they felt like their decisions did not make a difference or if the differences were not visually perceivable. It is possible that if participants were able to understand the impact of their actions on the optimization process, or receive training [66], they would reach better outcomes. Providing more transparency around what the optimizer is doing in the moment could mitigate some of the negative aspects observed in this study. Considering these factors is critical to improve computational representations of hard-to-quantify qualities relevant for design.

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

5.3 Preference-based interactive optimization for capturing semantic attributes can complement text-to-image and text-to-3D models

The results of the study, although applied to only a single semantic attribute, reinforce that it is possible to reflect subjective qualities into visual outputs using aggregate mappings to a certain extent. Large-scale mappings of semantic and visual information [19] enable the rapid generation of outputs via text input (e.g., "a comfortable mug with a large handle") while, in the context of design, efforts have been taken to guide generative models to better align with subjective evaluations [67, 68, 69]. These approaches can facilitate design space exploration without the need to manually visualize and search through alternatives. However, is may be difficult to reflect subtle individual differences in perceptions of subjective attributes into visual artifacts using current text-to-image or 3D generation models, as current interactions generally need the users to adapt to the model (e.g., wording their prompts correctly and with enough detail) rather than the model adapting to the individual users. Results from this study indicate that interactivity, compared to outputs from a pre-trained model (i.e., the Aggregate condition), may help improve perception alignment for subjective qualities. This outcome is perhaps because of the ability of interactive models to support human decision making agency (e.g., through guidance or feedback) when computationally generating design outputs. In other words, though non-interactive aggregate models can represent a subjective attribute to some degree, interactivity can lead to improved perceptions. The approach taken in the study conducted here could be used in conjunction with more flexible generative models to generate outputs that are aligned with a specific, personalized semantic attribute. Furthermore, a blend of approaches may allow designers to build upon real data, leveraging empirical evidence from user preferences while simultaneously maintaining the agency to influence design outcomes. Such future advances can eventually lead to tools that enable designers to efficiently explore a vast design space while not having to sacrifice their individual styles.

5.4 Considerations for scaling to higher dimensions

In more complex design problems, capturing the interactions between a large number design parameters is particularly critical for translating abstract semantic goals into geometry. Individuals' perceptions may also vary to a greater degree when there are more design parameters (or there is a more abstract subjective goal), making interactive adaptation to the individual particularly important as complexity increases. At the same time, increasing the dimensionality impacts optimization performance (the method used here has not been tested beyond six dimensions [50]), making it difficult to implement interactive preference learning practically in these cases. With a more complex example, it may be necessary to increase the number of evaluations per individual or initialize with greater amounts of aggregated data to ensure that the optimization reaches the relevant areas of the design space. When scalability of the number of dimensions becomes particularly challenging, it may be necessary to pursue dimensionality reduction techniques and learn more concise design representations that capture the relevant features (e.g., design manifolds [70]). This approach can likely be achieved offline and used to initialize the preference learning process, for example, for specific product categories.

5.5 Limitations and future work

Some limitations and further work should be considered. First of all, the use of a GP allows for the estimation of a surrogate function which can represent a subjective scale in relation to design features, but this function is more representative of positive values (more comfortable) than negative ones (less comfortable). It should also be noted that the approach taken here may have to be modified if the adjectives describing the semantic pair are very different (e.g., traditional to elegant) compared to a quantity (less or more comfortable). Furthermore, because the adaptive querying involves sampling from the Gaussian Process and then evaluating over a line, there is a degree of randomness in what participants were presented with as a new design (i.e., exploration), which may impact decisions, particularly given the blackbox nature of the optimizer. Particularly with the method applied here, the granularity at which preference is expressed cannot be changed. Preferences are expressed over combinations of all parameters (global). This is helpful to address interaction effects, but it prevents the flexibility of specifying parameter level (local) preference. Mechanisms to enable changes in the granularity of expression could be to selectively allow manually changes (e.g., through direct specification or manipulation [47]). A challenge here, however, is encoding these local preferences properly without negatively impacting representations of global preference.

Some limitations also relate to the design representation. Most notably, using virtual (e.g., in virtual reality) prod-

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

uct representations over 2D product representations (i.e., images) leads to judgments that are more aligned with judgments for real products [71], but it is known that evaluations can differ between digital and physical models [72]. In early-stage design, creating even a moderate number of physical representations to assess preference comprehensively can be impractical. Therefore, the approach here can be used to narrow the design space down for more expensive design representations to be evaluated downstream. Additionally, in this study, the design was parameterized into five features that were assumed to be relatively important for the subjective attribute being considered. The surrogate function for each participant could be used to optimize a mug of a different shape for "comfort" as long as it can be parameterized by the features represented here (e.g. handle length, handle width). However, this is not inclusive to many other features people may have considered that contribute to further inter-individual differences. In our case, the adaptive querying was conducted using design features that mapped directly to low-level parameters. Participants were allowed to provide a degree of abstract feedback, but an interesting area of future investigation would be the incorporation of higher-level conceptual feedback without a one-to-one mapping to the low-level parameters.

6 CONCLUSION

In this work, interactive and non-interactive optimization models are utilized and evaluated for their ability to capture and reflect differences in human perceptions of a subjective attribute. We provide insight into the ability to capture an aggregate-level representation and show how subtle individual differences can be captured by interactive models, resulting in different outcomes and satisfaction with these outcomes. The results show that a non-interactive aggregate-level model can represent human perceptions, but that reflecting individuals' decisions through an interactive process leads to even better alignment. Thus, while people may share some commonalities in their perceptions, individualization can be useful to generate and adapt designs at a semantic level. These systems can blend the strength of computational models with designer expertise and intuition, allowing designers to guide the translation of abstract semantic concepts into design features in real-time, and efficiently customize generic parametric designs to fit their own perceptions of a quality (e.g., more sporty, more luxurious). Such approaches can be complementary to large-scale semantic-to-visual mappings, particularly when perceptions diverge from the "group" or for design contexts not encompassed in the existing training data for foundation models.

ACKNOWLEDGEMENTS

This work has been supported by the Regents of University of California and NSF (2145432-CAREER). The findings represent the views of the authors and not necessarily those of the sponsors. A preliminary version of this study was presented for ASME IDETC 2023 [73].

REFERENCES

[1] Krippendorff, Klaus and Butter, Reinhart. "Product semantics-exploring the symbolic qualities of form." *Innovation* Vol. 3 (1984): pp. 4–9.

[2] Demirbilek, Oya and Sener, Bahar. "Product design, semantics and emotional response." *Ergonomics* Vol. 46 No. 13-14 (2003): pp. 1346–1360.

[3] Benaissa, Brahim and Kobayashi, Masakazu. "The consumers' response to product design: a narrative review." *Ergonomics* Vol. 66 No. 6 (2023): pp. 791–820.

[4] Krippendorff, Klaus and Butter, Reinhart. "Semantics: Meanings and contexts of artifacts." *Product experience*. Elsevier (2008): pp. 353–376.

[5] Lin, David Chuan-En and Martelaro, Nikolas. "Learning Personal Style from Few Examples." *Designing Interactive Systems Conference 2021*: p. 1566–1578. 2021. Association for Computing Machinery, New York, NY, USA. DOI 10.1145/3461778.3462115.

[6] Burnap, Alexander, Hartley, Jeffrey, Pan, Yanxin, Gonzalez, Richard and Papalambros, Panos Y. "Balancing design freedom and brand recognition in the evolution of automotive brand styling." *Design science* Vol. 2 (2016): p. e9.

[7] Liberman-Pincu, Ela, Korn, Oliver, Grund, Jonas, van Grondelle, Elmer D. and Oron-Gilad, Tal. "Designing

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

Socially Assistive Robots Exploring Israeli and German Designers' Perceptions." *J. Hum.-Robot Interact.* Vol. 13 No. 2 (2024). DOI 10.1145/3657646. URL https://doi.org/10.1145/3657646.

[8] Petiot, Jean-François and Yannou, Bernard. "Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics." *International Journal of Industrial Ergonomics* Vol. 33 No. 6 (2004): pp. 507–525.

[9] Orsborn, Seth, Cagan, Jonathan and Boatwright, Peter. "Quantifying aesthetic form preference in a utility function." *Journal of Mechanical Design* Vol. 131 No. 6 (2009): p. 061001.

[10] Kelly, Jarod C, Maheut, Pierre, Petiot, Jean-François and Papalambros, Panos Y. "Incorporating user shape preference in engineering design optimisation." *Journal of Engineering Design* Vol. 22 No. 9 (2011): pp. 627–650.

[11] Ren, Yi, Burnap, Alex and Papalambros, Panos. "Quantification of perceptual design attributes using a crowd." *DS 75-6: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 6: Design Information and Knowledge, Seoul, Korea, 19-22.08. 2013*. 2013.

[12] Sylcott, Brian and Cagan, Jonathan. "Modeling aggregate choice for form and function through metaconjoint analysis." *Journal of Mechanical Design* Vol. 136 No. 12 (2014): p. 124501.

[13] Goucher-Lambert, Kosa and Cagan, Jonathan. "The impact of sustainability on consumer preference judgments of product attributes." *Journal of Mechanical Design* Vol. 137 No. 8 (2015): p. 081401.

[14] Valencia-Romero, Ambrosio and Lugo, José E. "Part-worth utilities of Gestalt principles for product esthetics: a case study of a bottle silhouette." *Journal of Mechanical Design* Vol. 138 No. 8 (2016): p. 081102.

[15] Sylcott, Brian, Michalek, Jeremy J and Cagan, Jonathan. "Towards understanding the role of interaction effects in visual conjoint analysis." *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 55881: p. V03AT03A012. 2013. American Society of Mechanical Engineers.

[16] Gustafsson, Anders, Ekdahl, Fredrik and Bergman, Bo. "Conjoint analysis: a useful tool in the design process." *Total Quality Management* Vol. 10 No. 3 (1999): pp. 327–343.

[17] Liu, Vivian, Vermeulen, Jo, Fitzmaurice, George and Matejka, Justin. "3DALL-E: Integrating text-to-image AI in 3D design workflows." *Proceedings of the 2023 ACM designing interactive systems conference*: pp. 1955–1977. 2023.

[18] Brisco, Ross, Hay, Laura and Dhami, Sam. "Exploring the role of text-to-image AI in concept generation." *Proceedings of the Design Society* Vol. 3 (2023): pp. 1835–1844.

[19] Radford, Alec, Kim, Jong Wook, Hallacy, Chris, Ramesh, Aditya, Goh, Gabriel, Agarwal, Sandhini, Sastry, Girish, Askell, Amanda, Mishkin, Pamela, Clark, Jack et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*: pp. 8748–8763. 2021. PMLR.

[20] Rombach, Robin, Blattmann, Andreas, Lorenz, Dominik, Esser, Patrick and Ommer, Björn. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*: pp. 10684–10695. 2022.

[21] Liao, Jiayi, Chen, Xu, Fu, Qiang, Du, Lun, He, Xiangnan, Wang, Xiang, Han, Shi and Zhang, Dongmei. "Text-to-image generation for abstract concepts." *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 4: pp. 3360–3368. 2024.

[22] Kang, Namwoo, Ren, Yi, Feinberg, Fred and Papalambros, Panos. "Form+ function: Optimizing aesthetic product design via adaptive, geometrized preference elicitation." *arXiv preprint arXiv:1912.05047* (2019).

[23] Tucker, Maegan, Novoseller, Ellen, Kann, Claudia, Sui, Yanan, Yue, Yisong, Burdick, Joel W and Ames, Aaron D. "Preference-based learning for exoskeleton gait optimization." *2020 IEEE international conference on robotics and automation (ICRA)*: pp. 2351–2357. 2020. IEEE.

[24] Lun, Zhaoliang, Kalogerakis, Evangelos, Wang, Rui and Sheffer, Alla. "Functionality Preserving Shape Style Transfer." *ACM Trans. Graph.* Vol. 35 No. 6 (2016). DOI 10.1145/2980179.2980237.

[25] Yumer, Mehmet Ersin, Chaudhuri, Siddhartha, Hodgins, Jessica K and Kara, Levent Burak. "Semantic shape editing using deformation handles." *ACM Transactions on Graphics (TOG)* Vol. 34 No. 4 (2015): pp. 1–12.

[26] Nagamachi, Mitsuo. "Kansei engineering: a new ergonomic consumer-oriented technology for product development." *International Journal of industrial ergonomics* Vol. 15 No. 1 (1995): pp. 3–11.

[27] Reid, Tahira N, Gonzalez, Richard D and Papalambros, Panos Y. "Quantification of perceived environmental friendliness for vehicle silhouette design." *Journal of Mechanical Design* Vol. 132 No. 10 (2010): p. 101010.

[28] Zintgraf, Luisa M., Roijers, Diederik M., Linders, Sjoerd, Jonker, Catholijn M. and Nowé, Ann. "Ordered Preference Elicitation Strategies for Supporting Multi-Objective Decision Making." *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*: p. 1477–1485. 2018. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.

[29] Petiot, Jean-François and Grognet, Stephane. "Product design: a vectors field-based approach for preference modelling." *Journal of Engineering Design* Vol. 17 No. 03 (2006): pp. 217–233.

[30] Petiot, Jean-François and Dagher, Antoine. "Preference-oriented form design: application to cars' headlights." *International Journal on Interactive Design and Manufacturing (IJIDeM)* Vol. 5 (2011): pp. 17–27.

[31] Nandy, Ananya and Goucher-Lambert, Kosa. "Do human and computational evaluations of similarity align? An empirical study of product function." *Journal of Mechanical Design* Vol. 144 No. 4 (2022): p. 041404.

[32] Jiang, Wei, Zhao, Wu, Du, Lin, Zhang, Kai and Yu, Miao. "Product perceptual similarity evaluation: From attributive error to human knowledge hierarchy." *Journal of Computing and Information Science in Engineering* Vol. 23 No. 2 (2023): p. 021002.

[33] Carroll, J Douglas and Chang, Jih-Jie. "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition." *Psychometrika* Vol. 35 No. 3 (1970): pp. 283–319.

[34] Ren, Yi and Papalambros, Panos Y. "On design preference elicitation with crowd implicit feedback." *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 45028: pp. 541–551. 2012. American Society of Mechanical Engineers.

[35] Wan, Jie and Krishnamurty, Sundar. "Learning-based preference modeling in engineering design decision-making." *Journal of Mechanical Design* Vol. 123 No. 2 (2001): pp. 191–198.

[36] Thurston, Deborah L. "Real and misconceived limitations to decision based design with utility analysis." *Journal of Mechanical Design* Vol. 123 No. 2 (2001): pp. 176–182.

[37] Valencia-Romero, Ambrosio and Lugo, José E. "An immersive virtual discrete choice experiment for elicitation of product aesthetics using Gestalt principles." *Design Science* Vol. 3 (2017): p. e11.

[38] Reid, Tahira N, Frischknecht, Bart D and Papalambros, Panos Y. "Perceptual attributes in product design: Fuel economy and silhouette-based perceived environmental friendliness tradeoffs in automotive vehicle design." *Journal of Mechanical Design* Vol. 134 No. 4 (2012): p. 041006.

[39] Burnap, Alexander, Hartley, Jeffrey, Pan, Yanxin, Gonzalez, Richard and Papalambros, Panos Y. "Balancing design freedom and brand recognition in the evolution of automotive brand styling." *Design Science* Vol. 2 (2016): p. e9. DOI 10.1017/dsj.2016.9.

[40] Poirson, E and Petiot, J-F. "Interactive genetic algorithm to collect user perceptions. Application to the design of stemmed glasses." *Nature-Inspired Methods for Metaheuristics Optimization: Algorithms and Applications in Science and Engineering* (2020): pp. 35–51.

[41] Tseng, Ian, Cagan, Jonathan and Kotovsky, Kenneth. "Concurrent optimization of computationally learned stylistic form and functional goals." *Journal of Mechanical Design* Vol. 134 No. 11 (2012): p. 111006.

[42] Burnap, Alex, Pan, Yanxin, Liu, Ye, Ren, Yi, Lee, Honglak, Gonzalez, Richard and Papalambros, Panos Y. "Improving design preference prediction accuracy using feature learning." *Journal of Mechanical Design* Vol. 138 No. 7 (2016): p. 071404.

[43] Tuarob, Suppawong and Tucker, Conrad S. "Quantifying product favorability and extracting notable product features using large scale social media data." *Journal of Computing and Information Science in Engineering* Vol. 15 No. 3 (2015): p. 031003.

[44] Jin, Jian, Liu, Ying, Ji, Ping and Kwong, Che Kit. "Review on recent advances in information mining from big consumer opinion data for product design." *Journal of Computing and Information Science in Engineering* Vol. 19 No. 1 (2019): p. 010801.

[45] Bıyık, Erdem, Huynh, Nicolas, Kochenderfer, Mykel J and Sadigh, Dorsa. "Active preference-based gaussian process regression for reward learning." *Robotics: Science and Systems 2020*. 2020.

[46] Brochu, Eric, Cora, Vlad M and De Freitas, Nando. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." *arXiv preprint arXiv:1012.2599* (2010).

[47] Koyama, Yuki, Sato, Issei, Sakamoto, Daisuke and Igarashi, Takeo. "Sequential line search for efficient visual design optimization by crowds." *ACM Transactions on Graphics (TOG)* Vol. 36 No. 4 (2017): pp. 1–11.

[48] Koyama, Yuki, Sato, Issei and Goto, Masataka. "Sequential gallery for interactive visual design optimization."

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

ACM Transactions on Graphics (TOG) Vol. 39 No. 4 (2020): pp. 88–1.

[49] Koyama, Yuki and Goto, Masataka. "BO as Assistant: Using Bayesian Optimization for Asynchronously Generating Design Suggestions." *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*: pp. 1–14. 2022.

[50] Tucker, Maegan, Cheng, Myra, Novoseller, Ellen, Cheng, Richard, Yue, Yisong, Burdick, Joel W and Ames, Aaron D. "Human preference-based learning for high-dimensional optimization of exoskeleton walking gaits." *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*: pp. 3423–3430. 2020. IEEE.

[51] Nielsen, Jens Brehm Bagger, Nielsen, Jakob and Larsen, Jan. "Perception-based personalization of hearing aids using Gaussian processes and active learning." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* Vol. 23 No. 1 (2014): pp. 162–173.

[52] Tao, Siyu, Van Beek, Anton, Apley, Daniel W and Chen, Wei. "Multi-model Bayesian optimization for simulation-based design." *Journal of Mechanical Design* Vol. 143 No. 11 (2021).

[53] Iyer, Akshay, Yerramilli, Suraj, Rondinelli, James M, Apley, Daniel W and Chen, Wei. "Descriptor aided Bayesian optimization for many-level qualitative variables with materials design applications." *Journal of Mechanical Design* Vol. 145 No. 3 (2023): p. 031701.

[54] Campbell, Matthew and Hoyle, Christopher. "Constraining the Feasible Design Space in Bayesian Optimization With User Feedback." *Journal of Mechanical Design* Vol. 146 (2024): pp. 041703–1.

[55] Williams, Christopher KI and Rasmussen, Carl Edward. *Gaussian processes for machine learning*. Vol. 2. MIT Press, Cambridge, MA (2006).

[56] Ren, Yi and Papalambros, Panos Y. "A design preference elicitation query as an optimization process." *Journal of Mechanical Design* Vol. 133 No. 11 (2011).

[57] Lepird, John R, Owen, Michael P and Kochenderfer, Mykel J. "Bayesian preference elicitation for multiobjective engineering design optimization." *Journal of Aerospace Information Systems* Vol. 12 No. 10 (2015): pp. 634–645.

[58] Desmedt, Nicolas, Iliopoulou, Vicky, Lopez, Carlos and De Grave, Kurt. "Active preference learning in product design decisions." *Procedia CIRP* Vol. 100 (2021): pp. 277–282.

[59] Chu, Wei and Ghahramani, Zoubin. "Preference Learning with Gaussian Processes." *Proceedings of the 22nd International Conference on Machine Learning*: p. 137–144. 2005. Association for Computing Machinery, New York, NY, USA. DOI 10.1145/1102351.1102369.

[60] Balandat, Maximilian, Karrer, Brian, Jiang, Daniel R., Daulton, Samuel, Letham, Benjamin, Wilson, Andrew Gordon and Bakshy, Eytan. "BOTORCH: A Framework for Efficient Monte-Carlo Bayesian Optimization." *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 2020. Curran Associates Inc., Red Hook, NY, USA.

[61] Afshari, Hamid, Peng, Qingjin and Gu, Peihua. "Design Optimization for Sustainable Products Under Users' Preference Changes." *Journal of Computing and Information Science in Engineering* Vol. 16 No. 4 (2016): p. 041001.

[62] Goucher-Lambert, Kosa and Cagan, Jonathan. "Crowdsourcing inspiration: Using crowd generated inspirational stimuli to support designer ideation." *Design Studies* Vol. 61 (2019): pp. 1–29.

[63] Goucher-Lambert, Kosa, Gyory, Joshua T, Kotovsky, Kenneth and Cagan, Jonathan. "Adaptive inspirational design stimuli: using design output to computationally search for stimuli that impact concept generation." *Journal of Mechanical Design* Vol. 142 No. 9 (2020): p. 091401.

[64] Chan, Liwei, Liao, Yi-Chi, Mo, George B, Dudley, John J, Cheng, Chun-Lien, Kristensson, Per Ola and Oulasvirta, Antti. "Investigating positive and negative qualities of human-in-the-loop optimization for designing interaction techniques." *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*: pp. 1–14. 2022.

[65] Colella, Fabio, Daee, Pedram, Jokinen, Jussi, Oulasvirta, Antti and Kaski, Samuel. "Human strategic steering improves performance of interactive optimization." *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*: pp. 293–297. 2020.

[66] Simpson, Timothy W and Zhang, Xiaolong Luke. "The Importance of Training for Interactive Trade Space Exploration: A Study of Novice and Expert Users." *Journal of Computing and Information Science in Engineering* Vol. 11 (2011): pp. 031009–1.

[67] Yuan, Chenxi, Marion, Tucker and Moghaddam, Mohsen. "DDE-GAN: Integrating a Data-Driven Design Eval-

Journal of Mechanical Design. Received November 26, 2024;
Accepted manuscript posted September 4, 2025. doi:10.1115/1.4069687

uator Into Generative Adversarial Networks for Desirable and Diverse Concept Generation." *Journal of Mechanical Design* Vol. 145 No. 4 (2023). DOI 10.1115/1.4056500. 041407.

[68] Yuan, Chenxi, Marion, Tucker and Moghaddam, Mohsen. "Leveraging End-User Data for Enhanced Design Concept Evaluation: A Multimodal Deep Regression Model." *Journal of Mechanical Design* Vol. 144 No. 2 (2021). DOI 10.1115/1.4052366. 021403.

[69] Jiang, Zhoumingju, Wen, Hui, Han, Fred, Tang, Yunlong and Xiong, Yi. "Data-driven generative design for mass customization: A case study." *Advanced Engineering Informatics* Vol. 54 (2022): p. 101786. DOI 10.1016/j.aei.2022.101786.

[70] Chen, Wei, Fuge, Mark and Chazan, Jonah. "Design manifolds capture the intrinsic complexity and dimension of design spaces." *Journal of Mechanical Design* Vol. 139 No. 5 (2017): p. 051102.

[71] Tovares, Noah, Boatwright, Peter and Cagan, Jonathan. "Experiential conjoint analysis: an experience-based method for eliciting, capturing, and modeling consumer preference." *Journal of Mechanical Design* Vol. 136 No. 10 (2014): p. 101404.

[72] Häggman, Anders, Tsai, Geoff, Elsen, Catherine, Honda, Tomonori and Yang, Maria C. "Connections between the design tool, design attributes, and user preferences in early stage design." *Journal of Mechanical Design* Vol. 137 No. 7 (2015): p. 071408.

[73] Nandy, Ananya and Goucher-Lambert, Kosa. "Adaptive Optimization of Subjective Design Attributes: Characterizing Individual and Aggregate Perceptions." *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 87349: p. V006T06A001. 2023. American Society of Mechanical Engineers.