Products’ Shared Visual Features Do Not Cancel in Consumer Decisions

Consumers’ product purchase decisions typically involve comparing competing products’ visual features and functional attributes. Companies strive for “product differentiation” (Liu et al., 2013, “Product Family Design Through Ontology-Based Faceted Component Analysis, Selection, and Optimization,” ASME J. Mech. Des., 135(8), p. 081007; Thevenot and Simpson, 2009, “A Product Dissection-Based Methodology to Benchmark Product Family Design Alternatives,” ASME J. Mech. Des., 131(4), p. 041002; Kota et al., 2000, “A Metric for Evaluating Design Commonality in Product Families,” ASME J. Mech. Des., 122(4), pp. 403–410; Orfi et al. 2011, “Harnessing Product Complexity: Step 1—Establishing Product Complexity Dimensions and Indicators,” Eng. Econ., 56(1), pp. 59–79; and Shooter et al. 2005, “Toward a Multi-Agent Information Management Infrastructure for Product Family Planning and Mass Customisation,” Int. J. Mass Customisation, 1(1), pp. 134–155), which makes consumers’ product comparisons fruitful but also sometimes challenging. Psychologists who study decision-making have created models of choice such as the cancellation-and-focus (C&F) model. C&F explains and predicts how people decide between choice alternatives with both shared and unique attributes: The shared attributes are “canceled” (ignored) while the unique ones have greater weight in decisions. However, this behavior has only been tested with text descriptions of choice alternatives. To be useful to designers, C&F must be tested with product visuals. This study tests C&F under six conditions defined by: The representation mode (text-only, image-only, and image-with-text) and presentation (sequentially or side-by-side) of choice alternatives. For the products tested, C&F holds for only limited situations. Survey and eye-tracking data suggest different cognitive responses to shared text attributes versus shared image features: In text-only, an attribute’s repetition cancels its importance in decisions, while in images, repetition of a feature reinforces its importance. Generally, product differences prove to attract more attention than commonalities, demonstrating product differentiation’s importance in forming consumer preferences. [DOI: 10.1115/1.4030162]

1 Introduction

People routinely make comparisons in daily life for activities such as preference judgments and purchase decisions. Psychologists have discovered that the mind has various strategies for minimizing the mental burden of comparisons between alternatives, and that these strategies can be captured in models that predict decision outcomes and preferences. Tversky [1] proposed a feature-matching model describing how choice alternatives were compared in similarity judgments. Based on this model, Houston and Sherman [2] proposed the C&F model that specifically investigates comparisons between two alternatives for preference judgments.

At its core, the C&F model suggests that the mind ignores commonalities between choice alternatives so that it can focus on important differences, thus reducing mental burden. This paper tests to see if the mind uses C&F to minimize mental burden when processing product choices involving text versus images. The mental processing of product images with commonalities and differences is important to designers. When designing into a crowded product category or designing a product line, designers must carefully decide what to share across products and what to differentiate. The C&F model suggests that differentiation is the more important design task, because consumers ignore shared attributes in product comparisons. Yet a good designer knows that the commonalities of form communicate important meaning and are not discounted by consumers. Therefore, exploring the C&F model in the context of product design is important, because if shared features are indeed ignored by consumers, this gives direction that designers should focus on differentiation. However, if people instead focus on both shared and unique features in product design comparisons, as we anticipated, this emphasizes the importance of carefully designing both commonalities and differences of products. A strategic designer can exploit commonalities to position their product(s) more favorably. “Product feature” in this paper refers to visual characteristics of a product’s appearance, while attribute refers to characteristics described using text, see Fig. 1 for examples.

The C&F model explains how preference judgments are made when the given alternatives contain both unique and shared attributes and predicts preference trends in particular situations. This is further explained in Secs. 2 and 3. An example of a choice with shared and unique attributes is presented in Fig. 1. Based on the evaluation strategies specified by the C&F model, the alternatives provided for comparison are purposefully formed into unique-good (UG) and unique-bad (UB) pairs, as explained in Sec. 2, to control preference trends. The effectiveness of the C&F model has been tested only when alternatives are described by text attributes alone [2–8], as shown in Fig. 1(a). Figure 1(a) is very similar to the original C&F experiment [2], but that experiment used a 12-point scale instead of 8-point. This is not adequate for design purposes—we must also test visual features. For example, various car models made by BMW share the same kidney-like grille design but have unique designs for headlights and side mirrors.
Likewise, across brands with competing products, some features are shared (such as a high-gloss tablet screen) and some are differentiated or unique (such as tablet aspect ratio). In such cases, preference for different product designs should theoretically follow the evaluation strategies summarized in the C&F model, as described in Sec. 2.1. We add to the original testing scenario (Fig. 1(a)) a number of scenarios more applicable to product design, such as including images, shown in Fig. 1(b). Car X and Y in Fig. 1(b) share the same grille and side mirror designs, but have unique designs for headlights and wheels. As illustrated in Fig. 2, it is hypothesized that the C&F model will predict preferences for alternatives represented by product images, and common feature designs between the alternatives will attract less attention than the unique ones, similar as that for the alternatives described by text attributes.

We test the effectiveness of the C&F model in six conditions using the research hypotheses listed in Sec. 3. The six conditions vary by description/depiction of alternatives (by image-only, text-only, or image-with-text) and presentation (sequentially or side-by-side). Two products are tested: cars and bicycles. The experiment uses structured UG and UB pairs of choice alternatives, as used in the original C&F model tests.

The study employs eye-tracking technology to help test the core of the C&F model. As introduced in Sec. 2, eye-tracking data facilitate investigations of the visual evaluation process by providing information such as what people look at and for how long. Therefore, eye-tracking data can directly indicate consumers’ evaluation patterns for unique and shared attributes/features during product evaluations and help validate if the unique attributes/features attract more attention than the shared ones.

This research differs from existing work in the study of choice alternatives that have mixed “good” and “bad” attributes levels, or levels along a spectrum—many such studies exist in design, psychology, and marketing literature (e.g., see Refs. [9–13]). The purposefully structured UG and UB pairs in C&F work lead to the identification of the effects of shared and unique attributes on consumer decisions; they also make comparisons of choice alternatives a difficult task, which explicitly invokes cognitive short-cuts (cancel and focus) that may otherwise lay dormant. Details about experiment stimuli and experiment design are provided in Sec. 4. Experiment results are presented in Sec. 5. Discussion is provided in Sec. 6. Section 7 concludes the study.

2 Background

2.1 C&F Model. The C&F model investigates the approach that people use to make a preference decision between a pair of choice alternatives that have both shared and unique attributes [2]: The foundation of the model is that, within a choice pair, the shared attributes are canceled (or ignored) by the evaluator, and the unique attributes attract the evaluator’s focus. Additionally, the model proposes that each of the two alternatives is given a special role in the decision. One alternative is the Referent and the other is the Subject. In the original experiment [2], the alternative that was shown to the participants first was considered the
Referent. The alternative shown second was named the Subject and proved more influential on the decision. Researchers [2–5] tested this by presenting UG and UB choice pairs, as shown in Table 1.

Table 1 includes highlighting to show the UG and UB alternatives used by Houston and Sherman [2]; again, the highlighting was not included in the experiment and is used here to illustrate the following: UG attributes are highlighted in light gray and UB attributes are highlighted in dark gray.

Table 1  Sample UB and UG pairs used in Ref. [2]. The highlighting was not included in the experiment and is used here to illustrate the following: UG attributes are highlighted in light gray and UB attributes are highlighted in dark gray.

<table>
<thead>
<tr>
<th>Automobile X</th>
<th>Automobile Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>UG pair</td>
<td>UB pair</td>
</tr>
<tr>
<td>Does not need repairs often</td>
<td>Does not need repairs often</td>
</tr>
<tr>
<td>Good financing available</td>
<td>Good financing available</td>
</tr>
<tr>
<td>Good ratings from a consumer guide</td>
<td>Good ratings from a consumer guide</td>
</tr>
<tr>
<td>Good acceleration</td>
<td>Good acceleration</td>
</tr>
<tr>
<td>A friend recommended this model</td>
<td>A friend recommended this model</td>
</tr>
<tr>
<td>Hard to find service outlets</td>
<td>Hard to find service outlets</td>
</tr>
<tr>
<td>Poor warranty</td>
<td>Poor warranty</td>
</tr>
<tr>
<td>Poor mileage</td>
<td>Poor mileage</td>
</tr>
<tr>
<td>High priced</td>
<td>High priced</td>
</tr>
<tr>
<td>Stereo included</td>
<td>Stereo included</td>
</tr>
<tr>
<td>Prestigious model</td>
<td>Prestigious model</td>
</tr>
<tr>
<td>Air conditioning included</td>
<td>Air conditioning included</td>
</tr>
<tr>
<td>Hard to find service outlets</td>
<td>Hard to find service outlets</td>
</tr>
<tr>
<td>Poor warranty</td>
<td>Poor warranty</td>
</tr>
<tr>
<td>Poor mileage</td>
<td>Poor mileage</td>
</tr>
<tr>
<td>High priced</td>
<td>High priced</td>
</tr>
<tr>
<td>Does not need repairs often</td>
<td>Does not need repairs often</td>
</tr>
<tr>
<td>Stereo included</td>
<td>Stereo included</td>
</tr>
<tr>
<td>Prestigious model</td>
<td>Prestigious model</td>
</tr>
<tr>
<td>Air conditioning included</td>
<td>Air conditioning included</td>
</tr>
<tr>
<td>Hard to find service outlets</td>
<td>Hard to find service outlets</td>
</tr>
<tr>
<td>Poor warranty</td>
<td>Poor warranty</td>
</tr>
<tr>
<td>Has had a lot of factory recalls</td>
<td>Has had a lot of factory recalls</td>
</tr>
<tr>
<td>Repair parts are hard to get</td>
<td>Repair parts are hard to get</td>
</tr>
</tbody>
</table>

Houston et al. [3] proposed, tested, and confirmed the above predictions of different preferences for UG and UB pairs in four experiments that manipulated the Subject and the Referent. In follow-on work, Houston and Sherman [2] tested the preference predictions in two conditions that presented alternatives side-by-side (Subject “Y” assigned as alternative on the right) and sequentially (Subject Y assigned as alternative shown last). The researchers validated their preference predictions only in the sequential condition. We reason that importance of the Subject Y holds only in the sequential condition because the participant reviews Y closer-in-time to the preference decision, whereas the side-by-side condition has no such timing difference between review of X and Y. In Secs. 5.3 and 6, we present eye-tracking evidence to support this original speculation. Houston and Sherman [2] also collected three particularly useful postpreference
evaluations: (1) Overall satisfaction with the preference decisions and (2, 3) “good-ness” ratings for both the accepted and rejected alternative (how good the participant thought the alternative was). For both evaluations, participants rated the UB pair higher than the UB pair, because in the UG pair the participants focused on UG attributes, which left good impressions for their preference evaluations.

Sütterlin et al. [5] replicated the findings of Houston et al. with sequentially presented, text-only alternatives and additionally used eye-tracking technology to show that, within the second alternative Y, the unique attributes attracted more gaze attention than the shared attributes. Dhar and Sherman [14] tested the C&F model with the addition of a no-choice alternative. They found that the no-choice alternative had a larger choice rate within UB sets than within UG sets.

Su et al. [15] manipulated the shared attributes provided in a pair and pointed out that the shared attributes can influence preferences depending on their relevance to the unique attributes and the quantity they indicated (e.g., “10 pieces chicken wings” versus “1 piece chicken wings”). Eye-tracking technology determined that shared attributes that were (a) relevant to the unique attributes and (b) indicated a large quantity (e.g., “10 pieces chicken wings”) attracted more gaze attention than irrelevant, small-quantity shared attributes.

2.2 Eye-Tracking Research. Eye movement data, recorded using eye-tracking technology, explicitly demonstrate how people visually evaluate objects and provide quantitative evidence of people’s cognitive processes [16]. The analysis of the data studies fixations, “eye movements that stabilize the retina over a stationary object of interest” [17]. Figure 3 illustrates two common metrics, termed gaze data: fixation time (temporal length of the fixation) and fixation count (number of fixations). Associated with the area of interest (AOI), the target area in a research stimulus, the two fixation metrics indicate the gaze attention attracted by the particular area.

Researchers have previously used eye-tracking technology in product design studies. Reid et al. [18] investigated design representation mode’s effects on consumers’ subjective, objective, and inference judgments of products. Eye-tracking was used as an investigation tool in addition to a survey instrument. They also looked at visual evaluation strategies related to making preference decisions and observed that some people preferred the alternative on which they spent more fixation time, while some other people did the opposite. Du and MacDonald [19] tested for correlations between gaze data for product features and feature importance to preference decisions and found significant correlations between the two. They also compared gaze data for noticeable feature size changes and those for unnoticeable ones, where significant differences were detected. Their work demonstrates eye-tracking’s potential use in predicting feature importance as well as saliency of feature size change. A study by She [20] incorporated eye-tracking to test effects of sustainable-triggering features for toasters. It was found that those features succeeded to trigger certain sustainability-related behaviors, such as spending more gaze attention on text attributes regarding the product’s sustainability, provided along with the product image. All of these studies take advantage of eye-tracking to help with product design in various ways.

Use of eye-tracking technology in areas like decision-making and information processing [17,21] is also related here. Researchers have used eye-tracking to study information acquisition behaviors [13,22,23] because it provides detailed information on what, when, and how the information is examined. Shimoo et al. [24] analyzed the gaze data during preference evaluations and proposed a “gaze cascade effect” closely related to the final preference decisions. Russo and Rosen [25] took advantage of gaze data to investigate the evaluation processes during multi-alternative choices. Russo and Dosher [26] combined the gaze data and verbal protocols to compare the use of holistic and dimensional evaluation strategies when multi-attribute binary choices were presented, and then provided suggestions for the development of decision rules. These uses of eye-tracking demonstrate its usefulness in studying preference formation.

3 Research Hypotheses

As described in Sec. 2, the C&F model can be tested with three approaches: (I) analyzing differences in preference decisions and postpreference evaluations using survey questions, (II) analyzing visual evaluation strategies using gaze data, and (III) a combination of (I) and (II).

This study uses approach (I), survey data, to test three hypotheses referenced from the original C&F testing work [2], summarized in Hypotheses 1A–1C and Eqs. (1)–(3). Note that we test all hypotheses in this study in conditions with text-only, images-only, and image-with-text.

HYPOTHESIS 1A. Choice ratings (V) lean to “strongly prefer product V” for the UG pair (G), more so than for the UB pair (B):

\[ \overline{V_G} - \overline{V_B} > 0 \]  \hspace{1cm} (1)

HYPOTHESIS 1B. Satisfaction (S) with the preference decision is higher for the UB pair than the UG pair:

\[ \overline{S_U} - \overline{S_B} > 0 \]  \hspace{1cm} (2)

HYPOTHESIS 1C. “Good-ness” rating (Γ) for both the accepted (A) and rejected (R) alternatives is higher in the UB pair than in the UG pair:

\[ \Gamma_{BA} - \Gamma_{BR} > 0, \quad \Gamma_{GR} - \Gamma_{BR} > 0 \] \hspace{1cm} (3)

Explanation of Hypothesis 1A: According to the C&F model and as explained in Sec. 2.1, the Subject (the alternative that has the larger decision weight and is thus more influential in decisions) should be preferred in the UG pair because its UG attributes/features “weigh more” than those of the Referent (the other alternative in the pair). The Referent should be preferred in the UB pair.
because the Subject’s UB attributes/features weigh more than those of the Referent. Following Ref. [2], Hypothesis 1A considers product Y as the Subject.

Similar to Ref. [2], the preference decision for each pair of product alternatives is indicated on an eight-level choice scale ranging from “strongly prefer product X” to “strongly prefer product Y”, as shown in Fig. 1. We term the value indicated on this scale choice rating (V). If V is on the right half of the scale (V \geq 5), it indicates that product Y is preferred, which we term accepted (A). If (V \leq 4), then product Y is not preferred, termed rejected (R). Product X is oppositely accepted/rejected.

Explanation of Hypothesis 1B: According to the C&F model, the unique attributes guide the preference decision. The UG pair makes people feel that they are making a decision with two good alternatives (even though these alternatives include bad attributes that are shared). Therefore, when compared to UB pair decision, people should be more satisfied with the decision made for the UG pair.

Explanation of Hypothesis 1C: It follows that people should rate both alternatives in the UG pair as better than those in the UB pair. This is tested using a scale that ranges from “very bad” to “very good,” a rating we term “good-ness” (Γ).

The study uses approach (II), gaze data, to test if unique attributes/features attract more gaze attention than shared ones, as the C&F model asserts that people focus on differences between alternatives and ignore information that is the same. This is addressed by Hypothesis 2, which is referenced from Ref. [5]:

**Hypothesis 2.** A unique (U) attribute/feature has longer fixation time (T) and higher fixation count (Q) than a shared (H) one:

\[ T_U - T_H > 0, \quad Q_U - Q_H > 0 \]  \hspace{1cm} (4)

Approach (III), survey and gaze data combined, uses gaze data to identify the alternative that has the larger total fixation time and assigns this as the Subject (gaze) regardless of presentation order/position. The other alternative is considered the Referent (gaze).

Testing of Hypothesis 3 combines this new approach of determining Subject/Referent with preference data from the survey to test if the C&F model holds:

**Hypothesis 3.** Transformed choice ratings (V') lead to “strongly prefer Subject (gaze)” for the UG pair, more so than for the UB pair:

\[ V'_U - V'_H > 0 \]  \hspace{1cm} (5)

### 4 Methodology

Testing of the C&F model was realized by part I of a five-part computer-based experiment, as described in Sec. 4.2. Results from other parts of the same experiment are reported in Ref. [19], which includes descriptions of the other parts of the survey. Section 4.1 details the preparation of experimental stimuli. Section 4.3 summarizes the experiment participant population. Section 4.4 describes data preparations prior to statistical analysis.

#### 4.1 Stimuli

Cars and electric bicycles were selected as test products. The original C&F experiment [2] tested cars, so we include cars here to facilitate a direct comparison. We include the electric bicycle because it is a novel product to U.S. consumers with low familiarity. This allows for explorations of product familiarity’s effects on the C&F model. People may have existing mature ways to evaluate a car, but not an electric bicycle. Each test product has three representation modes: image-only, text-only, and image-with-text, as shown in Fig. 4. Test stimuli were formed into UG and UB pairs. Each pair had two question versions in which the presentation order of the two stimuli in the pair was switched to eliminate potential bias in the survey. Sample UG and UB pairs in the image-with-text mode are provided in Fig. 5.

Image stimuli containing only images of the test products were generated in Adobe Photoshop by merging different feature designs into base images. The base image for cars was the 2012 Chevy Cruze [27] and for electric bicycles was the Shanyang electric bicycle [28]. We selected base images that were as neutral as possible to avoid bias. We chose neutral forms (not a sports car, for example) and muted colors. Cars had varied headlights, grille, side mirrors, and wheels; and electric bicycles have varied handlebars, seat, footrest, and cargo box. These visual features are the “varied features” mentioned in the rest of the paper. Figure 4 shows example visual features and text attributes.

It was first necessary to create “good” and “bad” attributes/features to form UG and UB choice pairs. As beauty is in the eye of the beholder, creating “good” and “bad” visual features required careful effort. To form the UB and UB image pairs, we first used our design expertise to select and modify features from web images so that some features were ugly and/or mismatched with the overall product styling (bad) and some were harmonious with the product styling (good), though not necessarily beautiful. We performed a pilot study to test our efforts in which design variants of each varied feature were verified as good and bad based on their desirability ratings [3]. The pilot study used printed cards that showed design variants of each feature merged into the base product image. The experiment ultimately used 29 out of 37 total variants, each tested on a separate card. The cards were grouped by varied feature (for example, all design variants for the feature “headlight” were grouped together).

Thirteen participants sorted design variants in these groups from most to least preferred and rated their desirability on an eight-level scale that ranged from “not desirable at all” to “very

<table>
<thead>
<tr>
<th>Image-only</th>
<th>Text-only</th>
<th>Image-with-text</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Image-only" /></td>
<td><img src="text.png" alt="Text-only" /></td>
<td><img src="image-with-text.png" alt="Image-with-text" /></td>
</tr>
<tr>
<td>Car model: Sedan, five seats Doesn’t need repairs often Has had a lot of factory recalls Stereo included Hard to find service outlets</td>
<td>Car model: Sedan, five seats Doesn’t need repairs often Has had a lot of factory recalls Stereo included Hard to find service outlets</td>
<td>Car model: Sedan, five seats Doesn’t need repairs often Has had a lot of factory recalls Stereo included Hard to find service outlets</td>
</tr>
</tbody>
</table>

*Fig. 4 Three representation modes of the car stimuli: image-only, text-only, and image-with-text*
Procedure. Text attributes for cars were referenced from Ref. [2] and those for electric bicycles were referenced from product descriptions on Amazon.com. To verify them as good or bad, the attributes were provided to the participants in the pilot study in a similar manner to the visual features. According to the t-test results, all of the good attributes used in the experiment had significantly larger desirability ratings than the bad attributes. Two UG and two UB pairs of text stimuli were generated for each product category. The stimuli in a UG pair shared two bad attributes while each having two UG attributes. The stimuli in a UB pair shared two good attributes while each having two UB attributes. Each stimulus in a pair had an attribute that described the model of the product (e.g., “car model: Sedan, five seats”). This attribute kept constant across text stimuli within a product category in order to be consistent with the constant base image used for the image stimuli.

Stimulus for the image-with-text representation mode was created using a combination of an image stimulus and a text stimulus as introduced above. Also for this mode, two UG and two UB pairs were generated for each product category.

4.2 Experiment Design. The experiment had six conditions: Image & Sequential (ISeq), Text & Sequential (TSeq), Image-with-Text & Side-by-Side (ITSBS), Text & Side-by-Side (TSBS), and Image-with-Text & Side-by-Side (ITSBS). The experiment was developed and deployed using ATTENTION TOOL software from iMotions Company [29] and shown on a Tobii T120 eye-tracking monitor screen, which tracked eye movements of participants while they were taking part in the experiment. A calibration process, provided by the ATTENTION TOOL software, was conducted for each participant before the experiment started.

The experiment started with instructions, which were followed by a practice question set and then test sets. In the test set, stimuli and corresponding survey questions were successively presented on separate screens, as demonstrated in Fig. 1, for the sequential condition. Separating the stimuli and the survey questions on separate screens ensured that the collected gaze data for the stimuli were clean and unclouded, for example, by repeatedly gazing back at a question at the top of the screen. The survey questions in the test set were the same as those in the practice question set, so participants knew the questions beforehand and kept them in mind while they were viewing the test stimuli. There was no time limit for each screen. Participants were presented with stimuli of the cars first, and then stimuli of the electric bicycles. For each product category, a participant saw a UB pair and a UG pair in a randomly determined order in two separate test sets. The pairs shown to the participant were randomly chosen from those prepared. After evaluating each pair, participants were instructed to compare product Y in the pair to product X and indicate their preferences on the eight-level choice rating scale, introduced in Sec. 3. (The paper refers to the products as “X” and “Y,” rather than “A” and “B” as in the experiment, to avoid confusion, as A and B are used in the equations here with other meanings.) Then, participants had to complete three postpreference evaluations using eight-level scales: (1) rate their satisfaction with the preference decisions from “very unsatisfied” to “very satisfied,” as demonstrated in Fig. 1; and (2, 3) rate good-ness of product X and Y from “very bad” to “very good.” Preference indications and postpreference ratings were performed on the same screen, right after the participant saw both stimuli in a pair; this is also illustrated in Fig. 1.

4.3 Participants. The experiment had two separate data-collection rounds with different participants. In the first round, participants were randomly assigned to one of the six conditions. To enlarge the sample size, a second round was conducted, which only tested the three sequential conditions (ISeq, TSeq, and

---

1Amazon.com
ITSeq), because authors of Ref. [2] found that the C&F model is effective only in the sequential condition, and our own conclusions from the first round of data-collection confirmed these findings. In the second round, participants were randomly assigned to the three sequential conditions. Excluding participants whose responses were unrecorded either because of computer issues or their failure in the eye-tracking calibration process, the experiment had 72 participants (37 males and 35 females) in the first round and 36 participants (18 males and 18 females) in the second round. The participants were recruited from the Iowa State University and compensated with $5 cash or minor extra course credit, deemed equivalent compensations by the Institutional Review Board for human subject studies. The course credit was minor compensation: For one course with total maximum score of 1850, only five extra points were given to the experiment participants; for the other course, the extra points only increased the letter grade on one assignment by a half-step. Only 14% of the 108 participants were students who took the course credit. The rest of the participants were either not in the associated courses or were staff members. All participants who came to the experiment passed an online screening survey used to avoid participants who did not meet basic criteria of participating in an eye-tracking experiment as suggested by Pernice and Nielsen [30].

4.4 Data Preparation. ATTEST TOOL software managed both the survey and gaze data. We manually created AOIs for each product stimulus, as demonstrated in Fig. 7, so that the software can identify the gaze data (fixation time and count) associated with each attribute/feature. Then, the survey and gaze data were exported from the software separately for further analysis. The software had difficulty detecting the fixations of eight participants. Any stimuli that had no fixations at all were excluded in the gaze data analysis, indicated in Table 3. The first round of data-collection included some incorrect text attributes for a UB pair of electric bicycle stimuli. Therefore, the survey and gaze data associated with that pair were excluded from the analysis. The software did not record one participant’s answers to six survey questions; we include this in the analysis as missing data.

5 Analysis and Results
The testing of the C&F model was conducted separately for each experimental condition (ISeq, TSeq, ITSeq, ISBS, TSBS, and ITSBS). Section 5.1 details the results for Hypotheses 1A–1C, which are based on the survey data. Section 5.2 details the results for Hypothesis 2, which are based on the gaze data. Section 5.3 details the results for Hypothesis 3, which are based on both the survey and gaze data.

5.1 Analysis and Results: Survey Data. Hypothesis 1A: For each participant (i), an individual-level average UG choice rating ($V_{Gi}$) and UB choice rating ($V_{Bi}$) were calculated by averaging...
choice ratings the participant gave to the UG pair of cars ($V_{UC}$) and that of electric bicycles ($V_{UE}$), and by averaging the ratings for the UB pair of cars ($V_{BC}$) and that of electric bicycles ($V_{BE}$), respectively, as indicated in Eq. (6). Pairwise $t$-tests tested if the difference between UG choice rating ($V_G$) and UB choice rating ($V_B$) was greater than 0. Equation (7) shows the calculations of $V_G$ and $V_B$, where $N$ represents the number of participants in a condition. Table 2 provides the results.

$$V_G = \frac{(V_{UC} + V_{UE})}{2}, \quad V_B = \frac{(V_{BC} + V_{BE})}{2} \tag{6}$$

$$S_G = \frac{N}{i=1}S_{iU}, \quad S_B = \frac{N}{i=1}S_{iU} \tag{7}$$

**Hypothesis 1B:** For each participant (i), an individual-level average UG satisfaction rating ($S_{GC}$) and UB satisfaction rating ($S_{GB}$) were calculated by averaging satisfaction ratings the participant gave to the UG pair of cars ($S_{GC}$) and that of electric bicycles ($S_{GC}$), and by averaging the ratings for the UB pair of cars ($S_{GB}$) and that of electric bicycles ($S_{GB}$), respectively, as indicated in Eq. (8). Pairwise $t$-tests tested if the difference between UG satisfaction rating ($S_G$) and UB satisfaction rating ($S_B$) was greater than 0. Equation (9) shows the calculation of $S_G$ and $S_B$. Table 2 provides the results.

$$S_G = \frac{N}{i=1}S_{iG}, \quad S_B = \frac{N}{i=1}S_{iB} \tag{8}$$

$$S_G = \frac{N}{i=1}S_{iG}, \quad S_B = \frac{N}{i=1}S_{iB} \tag{9}$$

**Hypothesis 1C:** For each participant (i), an individual-level average UG good-ness rating for the accepted alternative ($T_{GAC}$) and UB good-ness rating for the accepted alternative ($T_{BAC}$) were calculated by averaging good-ness ratings the participant gave to the accepted alternative in the UG pair of cars ($T_{GAC}$) and that of electric bicycles ($T_{GAC}$), and by averaging the ratings for the accepted alternative in the UB pair of cars ($T_{BAC}$) and that of electric bicycles ($T_{BAC}$), respectively, as indicated in Eq. (10). Pairwise $t$-tests were conducted to test if the difference between UG good-ness rating for the accepted alternative ($T_G$) and UB good-ness rating for the accepted alternative ($T_B$) was greater than 0. $T_G$ and $T_B$ were calculated as shown in Eq. (11). Table 2 provides the results. The same analysis was performed on the good-ness rating for the rejected alternatives; refer to Table 2.

$$T_{GAC} = \frac{N}{i=1}T_{iGAC}, \quad T_{BAC} = \frac{N}{i=1}T_{iBAC} \tag{10}$$

Table 3. Different abilities of the unique and shared attributes/features to attract gaze attention

<table>
<thead>
<tr>
<th>N</th>
<th>ISeq</th>
<th>TSeq</th>
<th>ITSeq</th>
<th>ISBS</th>
<th>TSBS</th>
<th>ITSBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>192 (11)</td>
<td>192 (5)</td>
<td>192 (1)</td>
<td>48 (1)</td>
<td>48 (1)</td>
<td>48 (1)</td>
</tr>
<tr>
<td>24</td>
<td>2.1–1.85</td>
<td>6.72–6.06**</td>
<td>3.34–1.53**</td>
<td>10.66–6.4**</td>
<td>2.31–1.8**</td>
<td>4.29–3.79*</td>
</tr>
</tbody>
</table>

$^* p < 0.1, \ast p < 0.05, \ast\ast p < 0.01$
prefer Referent (gaze)” to “strongly prefer Subject (gaze)” using Eq. (14). For example, consider a participant who strongly preferred product X in a pair, and product X was identified as the Subject (gaze) by the fact that the participant spent more time looking at product X, the transformed choice rating is “8,” indicating that the participant strongly preferred the Subject (gaze) alternative.

\[
V' = \begin{cases} 
V, & \text{if Subject (gaze) is product Y} \\
9 - V, & \text{if Subject (gaze) is product X} 
\end{cases}
\] (14)

The same analysis used for Hypothesis 1A in Sec. 5.1 was performed on V'. Here, pairwise t-tests tested if the difference between UG transformed choice rating (\(V'_U\)) and UB transformed choice rating (\(V'_B\)) was greater than 0. Table 4 provides the results.

### 6 Discussion

Table 5 summarizes the results of hypothesis testing. In general, results were mixed. The core of the C&F model holds only for the original sequential text condition. Yet, each condition finds some portion of the C&F model that significantly predicts trends in the choices made. This suggests a strong model; larger sample sizes, different numbers of attributes/features, and further stimuli production may have led to stronger results.

Hypothesis 1A is accepted in the TSBS condition, replicating the test results documented in Refs. [2] and [3]. It indicates that when the alternatives are represented by text-only and are shown sequentially, the second alternative is more likely to be preferred in the UG pair than in the UB pair, and the second alternative is confirmed as the Subject. As Houston and Sherman [2] also found, Hypothesis 1A is not accepted in the TSBS condition. In the additional conditions we added to the existing literature: ISeq, ITSeq, ISBS, and ITBSBS, Hypothesis 1A is also not accepted.

We further explored why Hypothesis 1A was not accepted in these remaining conditions. Hypothesis 3 was tested to explore our speculation that the C&F model inappropriately considers product Y as the Subject in side-by-side conditions. It seems unlikely that the right-hand product is given more weight in choice simply because it appears at right. We use gaze data to provide a new definition for “Subject” in Hypothesis 3: Subject (gaze) is the product that had the longer total fixation time. When tested, Hypothesis 3 is accepted in the TSBS condition: When the alternatives are represented by text-only and are shown side-by-side, the transformed choice rating leans to strongly preferring the alternative with the longer gaze time for the UG pair, more so than for the UB pair. This shows that the C&F model’s claim regarding the preference decision possibly holds in the TSBS condition, if gaze is accepted as a substitute for order effects. The opposite of Hypothesis 3 is found to be significant at 0.05 level in the ISeq condition. This may be due to influences of the shared features as discussed below.

When visual features are included for product alternatives, the model does not hold. In fact, the opposite trend is observed in the ISBS condition, but it does not reach statistical significance. In the ISeq condition, the preference decisions follow the trend predicted by the C&F model, but they do not reach statistical significance. ISeq also does not support Hypothesis 2 (as discussed below), suggesting underlying challenges with the presentation of image-with-text stimuli sequentially—namely, a difficulty holding the information about four attributes and four features in one’s head for comparison purposes. We believe the C&F model does not hold in all conditions that include images for the reasons discussed below.

Ratings on satisfaction and good-ness (Hypotheses 1B and 1C) in the TSeq and TSBS conditions do not fully replicate what Houston and Sherman [2] found; results from the two conditions for Hypothesis 1B and those from the TSeq condition for Hypothesis 1C (for the accepted alternative) do trend in the hypothesized direction, but do not reach statistical significance. Hypotheses 1B and 1C are both accepted in the ITBSBS condition, meaning that participants are more satisfied and feel that the alternatives are better in the UG versus UB pairs. Hypothesis 1C is accepted in the ISeq condition, but Hypothesis 1B is not. As these image-with-text conditions had double the information when compared to text- and image-only conditions, it is telling that the standout attributes/features contributed significantly to participants’ overall impressions of the decision: They could not weigh all information equally.

Hypotheses 1B and 1C are not supported in the ISeq or the ISBS condition. These two conditions even fail to show any trends that are predicted in Hypotheses 1B and 1C. Therefore, the opposite of Hypothesis 1B and 1C was tested. As shown in Table 5, the opposite of Hypothesis 1B was found to be significant at the 0.05 level and 0.1 level for the ISeq and ISBS conditions, respectively; the opposite of Hypothesis 1C (for the accepted alternative) was found to be significant at the 0.05 level in the ISeq condition. This is an interesting finding when paired with the results of Hypotheses 1A and 3, which were also rejected in these conditions.

Overall, the findings suggest that there are important differences in how people process text versus image product information, and

### Table 4 Differences in preferences between the UG and UB pairs

<table>
<thead>
<tr>
<th>Conditions</th>
<th>ISeq</th>
<th>TSeq</th>
<th>ITSeq</th>
<th>ISBS</th>
<th>TSBS</th>
<th>ITBSBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>21</td>
<td>23</td>
<td>24</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

* p < 0.05

### Table 5 A summary of the hypothesis testing results

<table>
<thead>
<tr>
<th>H</th>
<th>Hypothesized trend</th>
<th>ISeq</th>
<th>TSeq</th>
<th>ITSeq</th>
<th>ISBS</th>
<th>TSBS</th>
<th>ITBSBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A: Choice rating</td>
<td>(V'_U &gt; V'_B)</td>
<td>Op.*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1B: Satisfaction rating</td>
<td>(S'_U &gt; S'_B)</td>
<td>Op.*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1C: Good-ness rating</td>
<td>(G'_U &gt; G'_B)</td>
<td>Op.*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2: Fixation time and count</td>
<td>(T'_U &gt; T'_B)</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3: Transformed choice rating</td>
<td>(V'_U &gt; V'_B)</td>
<td>Op.*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Op. indicates that the opposite of the proposed hypothesis was found to be significant. Hypothesis is accepted at * 0.1 level, * 0.05 level, or ** 0.01 level.

Journal of Mechanical Design

JULY 2015, Vol. 137 / 071408-9
that these differences lead to the ineffectiveness of the C&F model for image-based comparisons. Our findings suggest that “cancellation” does not exist for shared image information, but rather “reinforcement.” It may be that shared or repeated features reinforce impressions rather than being canceled. Su et al. [15] found that shared text attributes did not cancel and can affect consumption decisions when (a) the shared attributes are relevant to the unique attributes and (b) indicated a large quantity. One explanation is that visual features do not cancel because they are always “relevant” to each other—they are all part of the whole to make up the image, and they all play significant roles in consumer decisions.

Hypothesis 2 is accepted in the TSeq and the three side-by-side conditions, with strong evidence that unique attributes attract more gaze attention (time and count) than shared attributes, consistent with the C&F model. For the ISeq condition, Hypothesis 2 is accepted for time but not count. Hypothesis 2 is not accepted in the ITSBS condition. This may be due to the fact that there is a large amount of information in different forms and on different screens, so the processing mode may change for this information-rich decision.

The C&F model-related hypotheses rely on an assumption that decisions are based on comparisons of attributes/features. So, for the visual product designs shown as images, the model assumes that people would deconstruct the whole design into separate features, evaluate them separately, and compare different feature designs. When this assumption does not hold, such as when the preference decision is based on holistic evaluations of the alternatives, the effectiveness of the C&F model could be compromised [6].

The car and electric bicycle had similar results when analyzed separately, except for a few cases. The car showed stronger effects (e.g., larger satisfaction difference between the UG and UB pairs) compared to the electric bicycle for Hypothesis 1B in the ITSBS condition, and for Hypothesis 1C (for the accepted alternative) in the ISeq condition; both products’ results trended consistently with the hypotheses. These differences between the two products suggest that people’s unfamiliarity with the electric bicycle’s good versus bad attributes/features may shrink the distinction between UG and UB pairs, especially in conditions that contain a large amount of information, as in the ITSBS and ISeq conditions. The two products behaved oppositely for Hypothesis 1C (for the rejected alternative) in the ITSBS condition and for Hypothesis 2 in the TSeq condition. In both of these cases, the results from the car trended consistently with the hypotheses while the bicycle did not. These results suggest that the unfamiliarity about the electric bicycle could raise shared attributes/features’ importance in decisions, but as the results were not seen in all conditions or a meaningful subset of conditions, the implications are unclear. In Hypothesis 2 test of the ISBS condition, the electric bicycle enhanced the unique features’ advantage of attracting gaze attention over the shared features, compared to the car. This indicates that people may have fewer existing mature evaluation strategies for visual design of the electric bicycle compared to the car, prompting them to rely on comparing differences while they are determining preferences for electric bicycles shown side-by-side. In summary, product familiarity could have some minor effects on post-preference evaluations of the UG and UB pairs and on the visual evaluations of the unique and shared attributes/features in a few conditions; but familiarity does not affect the core of the C&F model. The consistent results across a familiar product also suggest that bias due to choosing a particular make/model of the car did not have significant influence on the outcome of the hypotheses. Although we did not encounter the effects of this bias in the experiment, possible brand and form bias could cause, for example, cognitive dissonance between the described versus expected attributes and influence results.

7 Conclusion

This study uses both the survey and gaze data to test the C&F model in six conditions, four of which have not been tested before. While partially replicating previous findings regarding the C&F model [2], the study finds the inability of the model to predict preference or post-preference evaluation trends in UG and UB pairs when the choice alternatives include images. Importantly, trends that are opposite to the hypotheses on satisfaction and good-ness ratings are found in the two image-only conditions. It indicates that the shared feature designs between alternatives may reinforce good or bad impressions that are consistent with the valence of these designs, even though they attract less gaze attention than the unique ones. In addition, different hypothesis testing results obtained for the image-only conditions and the text-only conditions suggest that people process image versus text information differently. Using the gaze data, the study confirms in five out of the six conditions that differences between choice alternatives attract more gaze attention than commonalities.

A wider range of experimental conditions could enforce findings, particularly additional product categories, features, attributes, and representation forms for images (for example, sketches), which would be beneficial to the strength of our findings. There are some possible sources of error for Hypotheses 1A–1C and 3. The number of shared/unique attributes/features that are varied in the experiment is two of each and is small; only two UG pairs and two UB pairs of products are provided to each participant; both of the products (cars and electric bicycles) have relatively high costs. All of these can either allow or motivate the participants to carefully examine and consider the two alternatives in a pair instead of “canceling” the shared attributes/features.

The differences between text stimuli in a pair come from different product attributes (e.g., service versus mileage), as in the original C&F experiment. Each alternative has some attributes missing; the participant will know about the mileage for one alternative but not the other. This increases the cognitive load of the decision-making process, as the alternatives are more difficult to compare, and potentially magnifies the C&F model’s effectiveness. These artificially constructed pairs of product attributes can deviate from choice alternatives the consumers encounter in real-world and potentially limit implications of the results obtained from the text-only conditions. The participants may have different interpretations of or responses to missing attributes, affecting experiment results. A possible extension of the research is to study text attribute pairs without missing attributes across choice alternatives. This would test the strength of the original C&F hypotheses under reduced cognitive load. However, between image stimuli in the new work we have contributed here, the differences come from the same product feature (e.g., headlight 1 versus headlight 2). Directly emulating the original C&F experiment for the image stimuli with missing visual features would require a very creative approach, for example, we cannot think of a way to have the headlamps of a car missing in a stimulus without ruining experimental results. Additionally, our experiment used visual base images, and it should be less biased than the original text-only C&F experiment, which is very unclear as to the “car” and leaves its model and design up to the imagination of each individual respondent, allowing for much greater margins of variance. Both of these reductions in cognitive load, the use of a base image and the lack of missing features, could counteract the need for decision strategies such as C&F, thus providing a partial explanation of the results.

The study can be extended in different directions: (1) It can further detect how the shared features function in the UG and UB pairs to affect the satisfaction and good-ness ratings with image-only product stimuli. Factors that may influence the shared features’ effects can be tested. These features include number of shared features in a choice alternative, number of alternatives, and product category. (2) The study can also be furthered by testing different effects of the visual features and text attributes as discussed in Sec. 6. Visual features, as they may be more easily recognized, compared, and remembered relative to text attributes, may be weighed more rationally by consumers in decisions that can be influenced by the ordering effects of choice attributes. To test these, an experiment with two conditions (image-only versus text-only) can be designed. Participants’ memories of the features/attributes provided in each condition can be inspected.
through tests of recall or comprehension tests, in order to see if the features are more easily recognized and remembered. To verify whether or not features are weighted without ordering effects, the participants would be given two chances to indicate preferences for each pair of alternatives; these alternatives would be presented sequentially in the first chance and side-by-side in the second. The participants’ choice switch rates in the two conditions could be compared. A more consistent weighting strategy should lead to a smaller switch rate.

Implications for Design and Design Research. Our findings suggest that, possible, the cognitive processing of product images results in easier recognition, comparison, and recall as compared to text. Thus, people may be able to weigh visual product features more “rationally” in decisions and find themselves less influenced by the stimuli ordering effects that the C&F model relies on. If this is the case, this provides further evidence that design researchers should present experimental product information as visual features whenever possible, rather than trying to describe these features with text.

This study highlights the importance of shared features in design, an already intuitively important concept in fields such as industrial design. Whether designing features to be shared with product predecessors, shared with products in the same product line, or shared with competing products, designers must study what reinforcements they may create through shared features. Designers should consider and test attitudes and preference for potentially shared features in addition to considering production costs and ease of mass-customization. Otherwise, they risk damaging consumers’ overall impressions of a newly designed product or an entire brand portfolio with the presence of inappropriately shared features. Additionally, product differentiation remains an important target, as unique features are confirmed to attract extra gaze attention.

Acknowledgment

We would like to thank the iMotions Company for supporting our experiment by providing the Attention Tool software and associated technical assistance. We would also like to thank Dr. Frederick Lorenz of the Iowa State University for his advice on designing the experiment and analyzing the experimental results.

Nomenclature

\( A \) = accepted alternative
\( B \) = UB pair
\( C \) = car
\( E \) = electric bicycle
\( G \) = UG pair
\( H \) = shared attribute/feature
\( I \) = index of experiment participant
\( K \) = number of attributes/features in a condition
\( L \) = index of attribute/feature
\( N \) = number of experiment participants in a condition
\( Q \) = fixation count
\( R \) = rejected alternative
\( S \) = satisfaction rating
\( T \) = fixation time
\( U \) = unique attribute/feature
\( V \) = choice rating
\( W \) = transformed choice rating
\( \Gamma \) = “good-ness” rating

References


Downloaded From: http://mechanicaldesign.asmedigitalcollection.asme.org/ on 05/19/2015 Terms of Use: http://asme.org/terms