2016

Investigating effects of product visual designs on consumer judgments with the aid of eye-tracking

Ping Du
Iowa State University

Investigating effects of product visual designs on consumer judgments with the aid of eye-tracking

by

Ping Du

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Mechanical Engineering

Program of Study Committee:
Erin MacDonald, Co-Major Professor
Judy Vance, Co-Major Professor
Frederick Lorenz
Eric Cooper
Rafael Radkowski

Iowa State University

Ames, Iowa

2016

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to thank my co-major professor, Dr. Erin MacDonald. Thank you for all the knowledge you have taught me and for being patient with me while I was learning. I appreciate you sometimes pushed me to seek answers myself. This allowed me to improve and discover my potentials, although the processes might be difficult. Thank you for all your support and assistance. You helped me receive assistantships and financial support—crucial to me. You supported me as a visiting student researcher at Stanford University. You helped me find jobs and supported me when I met difficulties. I appreciate them all.

I would like to thank my other co-major professor, Dr. Judy Vance. Thank you for willing to be my co-major professor and for helping me with my research, as well as a variety of logistical things. Thank you for allowing me to join your research team. You and other team members were very kind to me, which made me feel welcome.

It is my honor to have Dr. Frederick Lorenz, Dr. Eric Cooper, and Dr. Rafael Radkowski as my committee members. Dr. Lorenz, thank you for teaching me knowledge on statistics and how to apply it to solve research problems. Dr. Cooper, thank you for your advice on my research projects. It was very valuable. Dr. Radkowski, thank you for sharing your research and your inspiring questions about my research projects.

I am so lucky to be a member of the IRIS Design Lab, where I met great colleagues like Le Chen, Jinjuan She, and Tahira Reid. I enjoyed discussing research and life with you. I appreciate all the help I received from my lab members. Thank you iMotions Inc. for generously providing the iMotions software, which allowed me to incorporate eye-tracking into experiments easily and straightforward.

I would like to thank my parents. Thank your encouragement and support over these years. These helped me during difficult times. Thank you for all your love!

ABSTRACT

Product visual designs convey a variety of information about these products to consumers. These designs play an important role in affecting consumer judgments, which further determine purchase decisions. Understanding how consumers decode visual designs to form judgments as well as how to use visual designs to affect consumer judgments are important. Insights in these will help designers make better design decisions and also present new possibilities of product design.

This dissertation employs eye-tracking technology to assist in understanding consumers' decoding processes. First, eye-tracking is used to examine how consumers evaluate visual designs to determine preferences and product differences. Then, eye-tracking is utilized to help investigate influences of (1) pairing products that have both commonalities and differences and (2) visual cues on consumer judgments separately.

Product features, defined as visible product attributes, are important constitutions of product visuals. Study 1 uses eye-tracking to address two topics about product features: (1) feature importance in preference decisions and (2) whether or not consumers can detect a feature's size change. Results from eye-tracking how subjects evaluated product images to determine preferences showed a feature's gaze data (e.g., how long the subjects fixated on the feature) significantly correlated with the feature's importance rating provided by the subjects. Results from eye-tracking how subjects detected differences between product images showed noticeable and unnoticeable feature size changes had significantly different corresponding gaze data. Statistical models of gaze data can predict importance and size change saliency of a feature.

Purchase decisions often require comparing products that have both commonalities and

differences. Study 2 investigates how this configuration of choice alternatives influences consumer judgments by testing a model of choice from psychology, the cancellation-and-focus (C&F) model, in the product design domain. The C&F model specifies when facing two choice alternatives that have both shared and unique attributes, people tend to ignore the shared attributes and focus on the unique ones, which can affect both preferences and certain postpreference judgments. The model had only been tested with text-only stimuli, where text-described attributes represented products. Study 2 tested the model with image-only, text-only, and image-with-text stimuli separately. It tested each stimuli type with two conditions: (1) presenting stimuli sequentially and (2) side-by-side. The C&F model held only in limited situations for the tested products. Generally, the unique attribute/feature had more gaze attention than the shared one, indicating the importance of product differences in consumer preferences. While a shared attribute was canceled in decisions, a shared feature reinforced impressions.

Consumers extract cues from visual designs and mentally associate them with unobservable product attributes to aid judgments. Study 3 investigates the possibility to rapidly build mental associations to influence consumer judgments. The study also compares the effectiveness of cuing holistically, through body shapes, and cuing by features. Subjects participated in an association-building task, where a visual cue was associated with either a positive or a negative judgment of environmental friendliness. Results from a latter testing task demonstrated that mental associations between body shape cues and environmental friendliness formed. Body shape cues affected products' environmental friendliness ratings in the desired direction, but feature cues did not. Gaze data showed the subjects adjusted their distributions of attention to a product after the association-building task, indicating the ability of cues to promote a more efficient decision-making behavior.

CHAPTER 1

INTRODUCTION

1.1 Motivations

Nowadays, consumers have a variety of product choices that can fulfill their needs, which makes it difficult for a product to be conspicuous and purchased in the crowded market. The initial designer-oriented market is changing into a consumer-oriented market [1], suggesting the necessity of considering consumers in product design [2]. Hsu et al. [3] identified that designers and consumers have significantly different perceptions of product designs. This indicates the need of an in-depth understanding of how consumers evaluate and judge products. This understanding can provide guidance to designers and help them design products perceived in desired ways by consumers.

Products often act as a major medium of the communication between designers and consumers. Designers encode their design intentions into both observable and unobservable product characteristics [4], while consumers decode information through interactions with products. During decoding, the visual design of a product—focus of this dissertation—plays an important role because it can affect many initial judgments of the product [5]. A product's visual design carries different kinds of information about the product. This design can leave aesthetic impressions, indicate functionality, usability, and social significance of a product, and help categorize a product [4, 6]. After decoding the visual design, consumers form judgments about products. As Reid et al. [7] summarized, consumers can form three types of judgments: "opinion," "inference," and "objective evaluation." Opinion refers to a judgment without specifiable correctness, such as preference and satisfaction; inference refers to a judgment that

has specifiable correctness and is made with missing or incomplete information, such as inferred safety of a car when the safety information of the car is not directly provided; objective evaluation refers to a judgment that has specifiable correctness and is made when needed information is provided, such as a judgment of the length of a car when the car is provided [7]. Outside factors, such as situational and individual factors [4, 6], can affect these consumer judgments. Examples for situational factors include social settings, marketing programs, and product predecessors; examples for individual factors include age, gender, and personality of consumers [4, 6].

Previous research has provided insights in the effects of visual designs on consumer judgments [5, 8-11]. For instances, MacDonald et al. [8] identified that people tended to judge wine contained in vertically stretched bottles as “sweet and fruity” or “dry and crisp” instead of “nutty and oaky”. Reid et al. [9] found that people perceived cars with smooth body shapes as more environmentally friendly than those with boxy body shapes. However, previous research has not directly studied how consumers use or decode visual designs to form judgments.

This dissertation attempts to fill this gap by using eye-tracking technology. Eye-tracking technology can monitor visual evaluation processes of people and inform researchers how people look at a given stimulus in the form of both quantitative data and qualitative visualizations. It enables researchers to directly examine the decoding processes of consumers and investigate how consumers form product judgments. In this way, it can help designers identify origins of visual design effects and take appropriate actions.

1.2 Contributions and Summary

This dissertation is comprised of three studies. Study 1 uses eye-tracking to directly examine how consumers decode products' visual designs to form opinions and objective evaluations. It identifies relationships between eye-tracking data and two basic aspects of product features (defined as visible product attributes)—importance and size—respectively. This suggests new solutions to address design concerns. Study 2 identifies how commonalities and differences between a pair of product choice alternatives affect consumers' opinions and evaluation processes. It uses eye-tracking to examine visual attention for commonalities and differences. Results of Study 2 provide guidance for designers to approach product commonalities and differences. Study 3 investigates a topic related to proactively affecting consumer judgments using visual designs. It proves the possibility to rapidly build mental associations between a product's visual cues and its environmental friendliness and then affect consumer inferences and decision-making behaviors. It uses eye-tracking to examine consumers' decision-making behaviors during and after building mental associations. Figure 1 [4, 6] illustrates the main theme of these three studies. Chapters 2, 3, and 4 of this dissertation detail these three studies, respectively. Insights obtained from these studies provide useful feedback for designers and contribute to product improvements.

Specifically, to identify the aforementioned relationships between eye-tracking data and aspects of product features, a computer-based experiment was conducted for Study 1. This study tested two case products: cars and electric bicycles. Subjects completed the experiment while a Tobii T120 eye-tracker tracked their eye movements. This experiment showed the subjects pairs of product images. The experiment obtained (1) eye-tracking information on product features

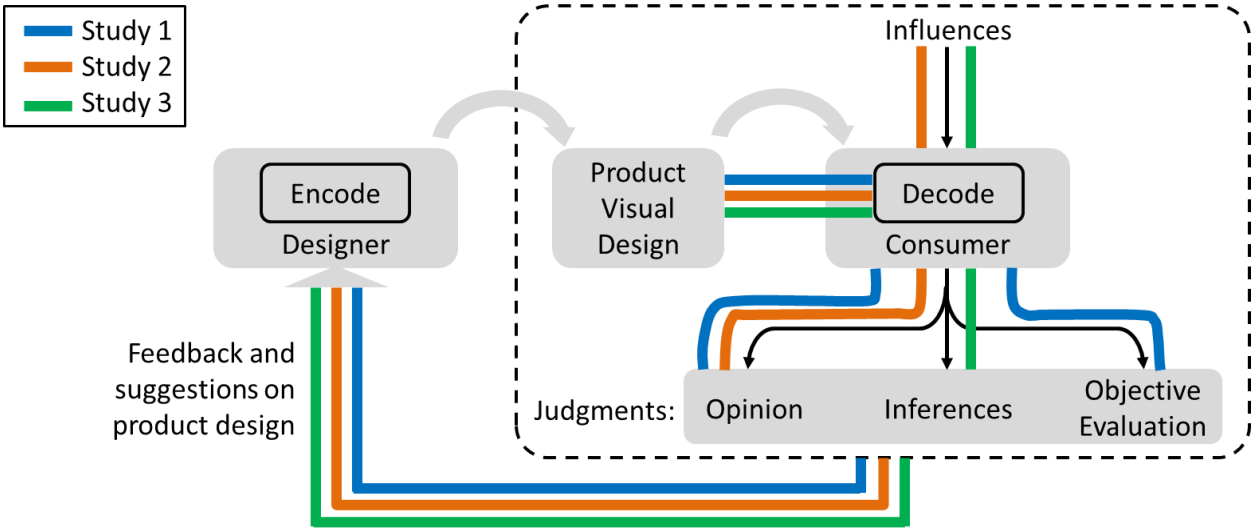


Fig. 1 This dissertation is comprised of three studies

while the subjects evaluated the images to determine preferences, such as how long and how frequently the subjects looked at the features, (2) eye-tracking information on features with size variants while the subjects evaluated the product images to determine their differences, (3) survey answers regarding whether or not the subjects noticed the feature size changes, and (4) self-reported feature importance ratings in forming preference decisions. Relationships between (1) and (4) and those between (2) and (3) were identified separately.

Having both commonalities and differences is common for products available in the market. Designers strive to create differences between products to better satisfy different needs and tastes of consumers. They also retain some commonalities between products to allow consumers to easily categorize these products, to help with brand communication, and to optimize the use of product lines. A model of choice from psychology, the cancellation-and-focus (C&F) model [12], specifies that when people are determining preferences for a pair of products that have both shared and unique attributes, they tend to ignore the shared attributes and only focus on the unique attributes. The C&F model only tests text-only stimuli, where products are described by text attributes. Study 2 translates the C&F model from the realm of psychology

to the realm of product design to identify effects of visual product commonalities and differences on consumer judgments. It tested the C&F model using six conditions—four had not been tested previously. These six conditions considered three representation modes (image-only, text-only, and image-with-text) and two presentations (sequential and side-by-side) of product choice alternatives. Study 2 shared experiment with Study 1. In the experiment, subjects saw purposefully configured pairs of products while the eye tracker tracked their eye movements. They were asked to indicate preferences and complete three postpreference evaluations for each pair, through which effects of shared and unique features/attributes on the subjects' opinions were identified. Eye-tracking data showed how the subjects evaluated the shared and unique features/attributes.

Designers use product visual cues to deliver messages about products to consumers. This communication relies on how consumers mentally associate the cues with unobservable product attributes. Study 3 attempted to rapidly build mental associations between a product's visual cues and its environmental friendliness through an association-building task, which referenced the feedback training method from psychology [13]. It also compared the effectiveness of cuing holistically through body shapes and cuing by features. An eye-tracking computer-based experiment was conducted for Study 3, using electric bicycles and electric heaters as case products. The association-building task inside the experiment showed subjects product images varied in the cues they had, asked the subjects to rate each image's environmental friendliness, and provided them with feedback information on the image's predetermined environmental friendliness rating. A latter testing task appraised whether or not the desired associations were built. Eye-tracking data examined subjects' visual attention for cued and uncued areas in the

testing task and compared them with those in the association-building task to detect if the subjects changed their decision-making behaviors.

References

[1] Nagamachi, M., 1995, "Kansei Engineering: A New Ergonomic Consumer-Oriented Technology for Product Development," International Journal of Industrial Ergonomics, 15(1), pp. 3-11.

[2] Petiot, J. F., and Grognet, S., 2006, "Product Design: A Vectors Field-based Approach for Preference Modelling," Journal of Engineering Design, 17(3), pp. 217-233.

[3] Hsu, S. H., Chuang, M. C., and Chang, C. C., 2000, "A Semantic Differential Study of Designers' and Users' Product Form Perception," International Journal of Industrial Ergonomics, 25(4), pp. 375-391.

[4] Crilly, N., Moultrie, J., and Clarkson, P. J., 2004, "Seeing Things: Consumer Response to the Visual Domain in Product Design," Design Studies, 25(6), pp. 547-577.

[5] Tseng, I., Cagan, J., Kotovsky, K., and Wood, M., 2013, "Form Function Fidelity," Journal of Mechanical Design, 135(1), p. 011006.

[6] Bloch, P. H., 1995, "Seeking the Ideal Form: Product Design and Consumer Response," Journal of Marketing, 59(3), pp. 16-29.

[7] Reid, T., MacDonald, E., and Du, P., 2013, "Impact of Product Design Representation on Customer Judgment," Journal of Mechanical Design, 135(9), p. 091008.

[8] MacDonald, E. F., Lubensky, A., Sohns, B., and Papalambros, P. Y., 2009, "Product Semantics and Wine Portfolio Optimisation," International Journal of Product Development, 7(1/2), pp. 73-98.

[9] Reid, T. N., Gonzalez, R. D., and Papalambros, P. Y., 2010, "Quantification of Perceived Environmental Friendliness for Vehicle Silhouette Design," Journal of Mechanical Design, 132(10), p. 101010.

[10] Swamy, S., Orsborn, S., Michalek, J., and Cagan, J., 2007, "Measurement of Headlight Form Preference Using Choice Based Conjoint Analysis," Proceedings of the ASME 2007 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Las Vegas, Nevada, pp. 197-206.

[11] Orbay, G., Fu, L., and Kara, L. B., 2013, "Shape Spirit: Deciphering Form Characteristics and Emotional Associations Through Geometric Abstraction," Proceedings of the ASME 2013 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, Portland, Oregon.

[12] Houston, D. A., and Sherman, S. J., 1995, "Cancellation and Focus: The Role of Shared and Unique Features in the Choice Process," Journal of Experimental Social Psychology, 31(4), pp. 357-378.

[13] Ashby, F. G., Maddox, W. T., and Bohil, C. J., 2002, "Observational versus Feedback Training in Rule-based and Information-integration Category Learning," Memory & Cognition, 30(5), pp. 666-677.

CHAPTER 2

EYE-TRACKING DATA PREDICT IMPORTANCE OF PRODUCT FEATURES AND SALIENCY OF SIZE CHANGE

A paper published in *Journal of Mechanical Design*

Ping Du
Department of Mechanical Engineering, Iowa State University, Ames, IA 50011

Erin MacDonald
Assistant Professor, Department of Mechanical Engineering, Stanford University, Stanford, CA, 94305

Abstract

Features, or visible product attributes, are indispensable product components that influence customer evaluations of functionality, usability, symbolic impressions and other qualities. Two basic components of features are visual appearance and size. This work tests whether or not eye-tracking data can (1) predict the relative importances between features, with respect to their visual design, in overall customer preference and (2) identify how much a feature must change in size in order to be noticeable by the viewer. The results demonstrate that feature importance is significantly correlated with a variety of gaze data. Results also show that there are significant differences in fixation time and count for noticeable versus unnoticeable size changes. Statistical models of gaze data can predict feature importance and saliency of size change.

2.1 Introduction

Product visuals are an important determinant of customer preference in almost all product categories. The preference for the overall visual design can be thought as based, wholly or in part, on the preferences for the visual design of individual product *features*, defined as visible

product attributes or characteristics. We focus on two challenges in the visual design of products: (1) determining feature importance, which refers to the importance of a feature to a customer forming a preference for the whole product; and (2) how noticeable size changes of a feature are to the customer.

Both of these challenges, importance and size, are directly linked to the profitability of a design. Designers cannot spend equal amounts of time perfecting all visual features. They must focus on those that are most important to the customer. Likewise, production budgets for intricate molds, labor-intensive manufacturing processes, and expensive materials must be weighted toward investing in product features that are most likely to increase sales. Size concerns present budgetary constraints as well; for example, a company may have the opportunity to save 10% on production costs by reducing the size of a product feature by 2%, but may have worries that customers will notice this change and perceive it as a loss of quality or luxury. Visual appearance and size of features can both be constrained by product function and other product objectives, such as weight.

As compared with a survey approach for gathering such information, eye-tracking offers more information with less exposure to stimuli. The data offered by eye-tracking hardware/software systems include gaze fixation location, or where a subject is looking on a computer monitor, fixation duration, and fixation timing and ordering—referred to here as gaze data. Gaze data have been used to indicate attribute importance [1], but the relationship between gaze data and importance ratings has not been directly proven. This paper lays a foundation for future use of the gaze data to facilitate product design.

The paper proceeds as follows: Section 2.2 provides background information for eye-tracking research, attribute importance, and feature size; Sec. 2.3 contains research propositions

and associated hypotheses for this paper; and Sec. 2.4 specifies the methodology. Results regarding feature importance are presented in Sec. 2.5, and results regarding size changes are provided in Sec. 2.6. Sections 2.7 and 2.8 discuss the results and present conclusions.

2.2 Background

2.2.1 Eye-Tracking Research. Gaze data provide quantitative information on the visual acquisition of information. Eye-tracking devices and corresponding software collect, refine, and analyze gaze data. According to the "eye-mind" hypothesis, what people look at is an indication of what they are mentally processing [2, 3]. Gaze data provide insights into human cognitive processes to facilitate the investigation of the origins of decisions or behaviors [4] and have been used in research areas, including psychology [5-7], marketing [8-10], human-computer interaction [4, 11, 12] and industrial engineering [13].

Eye-tracking has become one of the major process-tracing methods for information acquisition research [14]. Another major process-tracing method is computerized process tracing, which is usually conducted through the MOUSELAB software. The MOUSELAB software [15, 16] displays information on the computer with covered boxes; people acquire the information by moving the mouse cursor over a box; in the end, the software provides details about which boxes have been visited, the time spent on each box, and so forth. This kind of output is similar to that from the eye-tracking process, but as Lohse and Johnson [14] identified, eye-tracking technology can monitor the process of how the information is acquired more completely and naturally.

There are a number of eye-tracking studies related to the work presented here. Pieters and Warlop [9] used eye-tracking technology to study visual attention during brand choice. They found that, on average, subjects had longer fixation times on the brand they eventually chose

compared with other alternatives; neither time pressure nor task motivation altered this relationship. Gofman et al. [17] found that the first gaze location on food packages was correlated with both the total amount of time spent on the packages and the purchase decisions.

Koivunen et al. [18] analyzed gaze path data to study how people perceived product designs with different given tasks: memorizing the product, evaluating its aesthetics, usability, and durability. They also tested how the products were evaluated when no instructions were given. They observed that gaze paths and fixation times varied for the different tasks. Reid et al. [19] used both the gaze data and survey data to elucidate how customer judgments were affected by different representations of product design.

2.2.1.1 Eye-Tracking Equipment and Data. Eye-tracking equipment can be used while investigating 3D surroundings, but it is most typically used in conjunction with a computer monitor or screen. The screen presents different visual stimuli to a subject as he or she proceeds through an experimental session. This study used a Tobii T120 commercial eye tracker, shown in Fig. 1. The 17-in. thin-film transistor monitor displayed the experiment with a 1024 × 768 resolution. The Tobii hardware was used in conjunction with IMOTIONS' ATTENTION TOOL software [20] on a control computer, which managed the gaze data for further analysis. The software can also be configured to record survey data, for example, as in Qualtrics [21].

Fig. 1 The Tobii T120 eye tracker (left) and the associated control computer (right)

Prior to data analysis, gaze data are parsed into areas of interest (AOIs), or areas of a given stimulus related to the research hypothesis [11]. Fixation time and fixation count are commonly analyzed types of gaze data. Fixations are "eye movements that stabilize the retina over a stationary object of interest" [22]; fixation time refers to the duration of one fixation, and count refers to the number of fixations. Data on percentage-fixation time and first-located time are also used in this paper. The percentage-fixation time is the fixation time spent on an AOI divided by the total fixation time spent on the stimulus. Compared with the fixation time, the absolute measure of the gaze attention, the percentage-fixation time is a relative measurement. It takes into account of the fact that different stimuli attract different total fixation times, and different people have varied evaluation speeds. The first-located time for an AOI is a measurement of the time between initial exposure to a stimulus and first fixation on that AOI. Information about additional eye-tracking measurements can be found in Refs. [11, 12].

2.2.2 Attribute Importance. Addressing feature importance, specifically the importance of the visual design of features, can be thought of as studying a particular type of product attribute importance. Relative attribute importance identifies product attributes that are most likely to change customer preference through variation in attribute configuration. Bettman et al. define customer decision-making rules, such as compensatory and lexicographic decision rules [23] for which attribute importance can either directly or indirectly determine product choice. They model customer preference decisions (choices) by assigning different importance to different product attributes. In this model, differences in attribute importance cause each attribute of a choice option to have a weighted subjective value. These values are added together to get the total utility for the option. In this model, the customer's final choice decision largely depends on

the attributes that are most important to the customer, due to the larger weights on these attributes.

Attribute importance has been assessed in different ways. The most direct way of estimating attribute importance is to ask subjects why they choose a product option. By collecting the attributes indicated in an interview, the relative importance of an attribute can be estimated by the number of times the attribute is mentioned [24]. Attribute importance can also be estimated by establishing relationships between attributes and preference decisions or other evaluations [24]. Banks [25] applied linear discriminant functions to relate preference ratings on attributes to preference of overall products; functions' coefficients were then "converted to units of the standard deviation of the corresponding variable" to indicate the relative importance of the attributes.

In conjoint or discrete choice analysis, attribute importance can be interpreted from the estimated part-worths of attribute/levels (configurations). Orsborn et al. [26] apply this specifically to visual product features. They estimated customer preferences for quantified aesthetic forms using a logit model, and mentioned that attribute importance was indicated by the magnitude of the estimated part-worths. MacDonald et al. [27] studied importance of product attributes, but not visual features. They refined the definition of attribute importance in product design in order to perform statistical tests on this metric in a discrete choice study. Importance was defined as the percentage of customer choice that is determined by a specific attribute, in a hypothetical market where a full factorial combination of products is available. Jaccard et al. [28] conducted an information search task and gained insight on how customers searched for information about different attribute dimensions while making automobile purchasing decisions. Subjects evaluated a choice with available product profiles. Each profile had nine attribute

dimensions, each with associated information available to the subjects. The authors calculated two indices of importance for each subject: the order and number of pieces of information collected by subjects. In a mobile phone purchasing case study, Reisen et al. [29] used eye-tracking to test the relationship between attribute importance rankings and the frequency of evaluating related text. They found that the two variables were highly correlated, but there was potential bias in the nonrandom ordering of the related text. Warell and Nåbo [30] proposed a "design format modeling" method to capture and describe the visual form of products. With their method, important design features for a collection of products can be identified by comparing the weighted occurrence frequencies for the features. In a case study of home electronics by Bang and Olufsen, they found that features like "metal finishes," "black surfaces," and "geometrical forms" appeared more frequently compared with the others.

There are other studies that address importance of components of alternatives, that do not study products per se. Jaccard and King [31] estimated attribute importance by comparing two conditional probabilities, defined as the absolute difference between the probability of an intention, such as the intention to vote for a candidate, with a presence of an attribute and without. Schkade and Johnson [32] used the MOUSELAB system to investigate how people evaluated two-payoff gambles in two response modes, pricing, and choice, separately. They used the duration of time spent on an attribute as a measure of attention and indirectly demonstrated that the amount of attention that an attribute attracts may be an indication of its salience or importance.

2.2.3 Feature Size. Designers determine feature size using customer preference, technical requirements, and other sources of input. For example, a large grille on a car promotes engine cooling and better performance, but may look ugly to customers. Designers want a size

change to be noticed when it has positive effects on customer preference, for example, “30% more free” in a detergent bottle. But designers work to hide, conceal, or diminish size changes that could have negative effects on customer preference, for example, a decrease in car trunk size versus last year’s model. Noticeable versus unnoticeable difference is referred to as “saliency,” and this term is used in a binary sense (salient or not).

The relationship between attribute size and customer preference has been studied in the marketing literature. Michalek et al. [33] observed that large number size on a dial-readout scale was preferred as it indicated easy readability. Coelho do Vale et al. [34] discussed that package sizes of tempting products, small versus large, could affect customer choices through the activation of self-regulation. With self-regulation activated, customers were more likely to approach small packages, believing the packages would help regulate consumption. Chandon and Ordabayeva [35] found that compared with supersizing a product in three dimensions (height, width, and length), supersizing a product in only one dimension largely increased its choice share. This relationship was not affected by the fact that volume increase was clearly marked. A product downsized in three dimensions would have larger choice share than that downsized in one dimension. These results were due to the visual bias that the same amount of volume change through three dimensions was considered smaller than that through one dimension. Yang and Raghubir [36] have found that elongated containers for frequently purchased goods were considered to have a larger volume which could lead to decreased purchase quantity. Krider et al. [37] studied the perception of container shape and showed that a rectangular cream cheese container was considered larger than a round one even though they actually had the same volume, leading people to buy a lower quantity if packaged as a rectangle.

Krider et al. also discovered that customers initially relied on a single dimension, which was most salient, to compare areas.

2.3 Research Propositions and Associated Hypotheses

People use different viewing strategies for evaluating stimuli sequentially (Seq) and side-by-side (SBS), so both are tested here. In the experiment, all stimuli are evaluated in pairs. Within a SBS given pair, "product A" refers to the left-side stimulus and "product B" to the right-side stimulus. Within a given Seq pair, "product A" refers to the stimulus shown first and "product B" to the stimulus shown subsquently, on the next screen.

PROPOSITION 1. *Feature importance is correlated with gaze data in preference choices between two products.* This proposition is inferred from and supported by the literature presented in Secs. 2.2.1 and 2.2.2; see Refs. [1, 9, 17, 24, 28, 32]. The proposition is tested by the following hypotheses. This first set of hypotheses is accompanied by explanations in plain English to assist in understanding. For explanation of terms mentioned in the hypotheses below, refer to Sec. 2.2.1.1.

- *Hypothesis 1a: There is a positive correlation between feature importance and the feature's fixation time.* It is hypothesized that subjects spend a longer time looking at more important features during the choice task, and that the longer they look, the more important the feature.
- *Hypothesis 1b: There is a positive correlation between feature importance and the feature's percentage-fixation time.* It is hypothesized that subjects spend a larger percentage of a product stimulus' total fixation time looking at important features.

- *Hypothesis 1c: There is a positive correlation between feature importance and the feature's fixation count.* It is hypothesized that subjects look more frequently at important features than other features.
- *Hypothesis 1d: There is a negative correlation between feature importance and the feature's first-located time.* It is hypothesized that subjects look at important features first.

PROPOSITION 2. *Saliency of size change can be predicted by gaze data*. Sütterlin et al. [1] used gaze data to examine how customers evaluated pairs of options, which were described by text information and were shown sequentially. Some information provided in a pair was the same between options while the other information in the two options was different. In this way, the two options in a pair had both shared and unique information. They observed that the shared information between the two options was evaluated normally when it appeared in the first option but almost ignored when it appeared again in the second option. Based on their findings, we expect that this phenomenon will appear when the size change of a feature for a pair is unnoticeable. Gaze data, therefore, could be used to detect the saliency of size changes by testing whether features are ignored or not.

Two more measurements, Δ fixation time and Δ fixation count, are defined to test the proposition. Δ fixation time/count represents the difference in these quantities for features appearing on product B versus product A. The proposition is tested by a number of hypotheses summarized in Table 1. Note that the blank cells in Table 1 are also tested in the analysis, for completeness. For further explanation of terms and calculations mentioned in the hypotheses below, refer to Sec. 2.6.1.

As compared with its unnoticeable size-change counterpart:

- *Hypothesis 2a: A noticeable size change of a feature in product B has a longer fixation time (Seq).*
- *Hypothesis 2b: A noticeable size change of a feature in product B has a higher fixation count (Seq).*
- *Hypothesis 3a: A noticeable-size-change feature pair (for example, two car grilles of noticeable unequal size) has a longer total fixation time (SBS).*
- *Hypothesis 3b: A noticeable-size-change feature pair has a higher total fixation count (SBS).*
- *Hypothesis 4a: Δ fixation time of a noticeable-size-change feature is different (Seq).*
- *Hypothesis 4b: Δ fixation count of a noticeable-size-change feature is different (Seq).*

Table 1 An illustration of the test for proposition 2 about size changes.

		Condition	
Associated AOI	**Fixation metric**	Seq	SBS
Size-changed feature in Product B	Time	H2a: Noticeable, longer	
	Count	H2b: Noticeable, higher	
Feature pair with size change	Time		H3a: Noticeable, longer
	Count		H3b: Noticeable, higher
Feature pair with size change	Δ time	H4a: Noticeable, longer	
	Δ count	H4b: Noticeable, higher	

2.4 Method and Procedure

To test the hypotheses, a computer-based experiment was designed for 72 subjects and implemented using a Tobii eye-tracker and IMOTIONS' ATTENTION TOOL software, introduced in Sec. 2.2.1.1. Two product categories, cars, and electric bicycles, were used, described in Sec. 2.4.1. Table 2 provides an overview of the experiment design, described in detail in Sec. 2.4.2. Subjects for the experiment are introduced in Sec. 2.4.3. Data preparations, conducted before the results analysis, are introduced in Sec. 2.4.4.

Table 2 An overview of the experiment design. "*H*" refers to the associated hypotheses tested

Sect.	Survey Questions	Stimuli	Condition	*H*	Gaze Data	Survey Data Used?
I	1) Indicate preferences 2) Rate satisfaction of (1) 3) Rate Product A 4) Rate Product B	Feature Design Variants	Seq & SBS	1a	Fixation Time	No
				1b	%-Fixation Time	
				1c	Fixation Count	
				1d	First-located Time	
II	Indicate preferences	Feature Size and Design Variants	Seq & SBS	1a	Fixation Time	No
				1b	%-Fixation Time	
				1c	Fixation Count	
				1d	First-located Time	
III	Identify and write down the features that are different between two products	Feature Size and Design Variants	Seq	2a	Fixation Time	Size-changed features that subjects mentioned labeled as "noticeable-size-change"; otherwise as "unnoticeable-"
				2b	Fixation Count	
				4a	Δ fixation Time	
				4b	Δ fixation Count	
			SBS	3a	Fixation Time	
				3b	Fixation Count	
IV	Rate importance for different features	-	-	1a to 1d	-	Ratings used to test correlations
V	Demographic questions	-	-	-	-	Yes

2.4.1 Stimuli. Sample stimuli used in the experiment are shown in Fig. 2. The 2012 Chevy Cruze from the Chevrolet website [38] and a Shanyang electric bicycle model from the Global-tradekey website [39] were used as the base digital photographs of the stimuli. Only one base photograph for each of the products was used; different perspectives of the products were not shown. This car was selected because it provided a basic sedan model that was familiar to customers. This electric bicycle was selected because it was transformed from a bike model, which made for a product that was familiar in some ways, but unfamiliar in others. The car and the electric bicycle are both vehicles and are both durable goods, but they differed in their familiarity to customers in the U.S. This allows for explorations as to how product familiarity influences the relationships between feature importance and gaze data. For example, while brand

may complicate car evaluations, it is unlikely to affect electric bicycle evaluations as brand is not yet strongly identified with physical features for this new category of bikes.

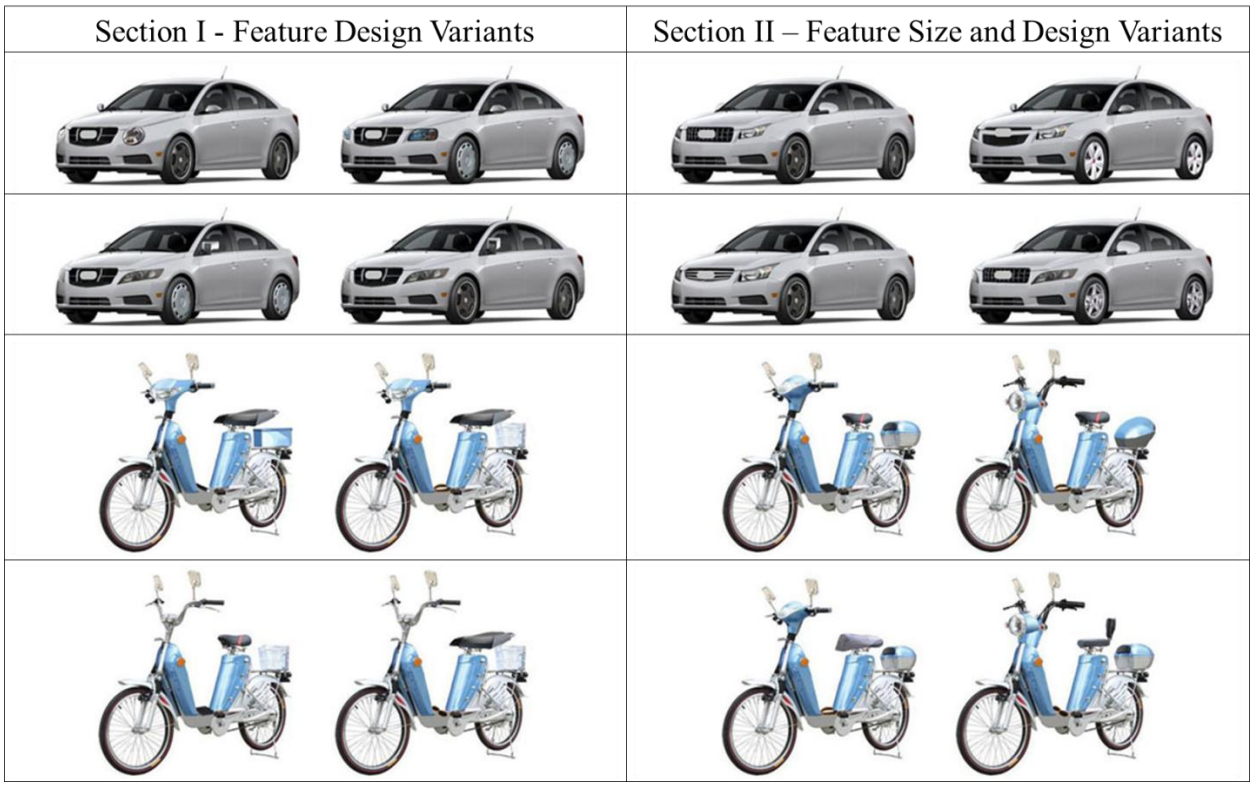


Fig. 2 Sample pairs used in the experiment (section II size variants are headlight (15%), side mirror (20%), seat (15%), and cargo box (10%), from top to bottom).

Sets of stimuli were created from the base photographs in Adobe Photoshop. The features that are varied are called "varied features." The design permutations of these varied features are referred to as "design variants" or "feature design variants," and the size permutations are referred to as "size variants" or "feature size variants." The car stimuli included four design variants each for four varied features: headlights, grille, side mirrors, and wheels. The electric bicycle stimuli included four design variants each for four varied features: handlebars, footrest, seat, and cargo box. The design variants were taken directly or modified from existing cars and bicycles, as shown in Fig. 3. To form product stimuli, the variants were "pasted" onto the base photographs and carefully blended. A pilot study was conducted to evaluate response to the

stimuli. Subjects found that the stimuli looked natural and did not distract them from evaluating the product designs—the variant components were not noticeably different from the rest of the stimuli.

Car					Electric Bicycle				
Design Variants	Headlight	Grill	Wheel	Side mirror	Design Variants	Handlebar	Seat	Cargo box	Footrest
1					1				
2					2				
3					3				
4					4				

Fig. 3 Design pool for varied features

For size variants, the headlights, grille, and side mirrors of the car and the handlebars, seat, and cargo box of the electric bicycle were proportionally resized from the base photographs. Each of these features had three size variants. The levels of size variants for the features of the car were 15%, 20%, and 25%, and the electric bicycle were 10%, 15%, 20%. Each feature size variant appeared in one stimulus, which was paired with a stimulus that had the base size of the feature. Table 3 shows the three pairs of stimuli formed for the size variants of the headlights. Numbers “1” through “4” in the table stand for different feature design variants (see Fig. 3), and the percentages represent size variants (enlargements).

2.4.2 Experiment Design. The experiment was composed of instruction screens, name tag screens that indicated the name(s) of the upcoming stimulus(or stimuli) like “car A”, product

Table 3 Pairs of car stimuli that involve size variants of headlights, and design variants for other features (numbers "1" through "4" represent different design variants, percentages represent enlargements)

		Headlight	Grill	Wheel	Side Mirror
Pair 1	Car A	1	4	2	1
	Car B	1 at 115%	1	1	4
Pair 2	Car A	1	3	3	2
	Car B	1 at 120%	2	4	3
Pair 3	Car A	1	1	2	3
	Car B	1 at 125%	2	1	1

stimulus screens that were used to collect gaze data, and survey question screens used to collect information, typically presented after product stimulus screens. A "screen" refers to the information presented on the computer screen, sometimes called a "page" or "slide." There was no time limit for each screen. In this paper, the gaze data from the product stimulus screens are analyzed (not the gaze data from survey question screens). The experiment flow and an example set of screens are illustrated in Figs. 4 and 5, respectively.

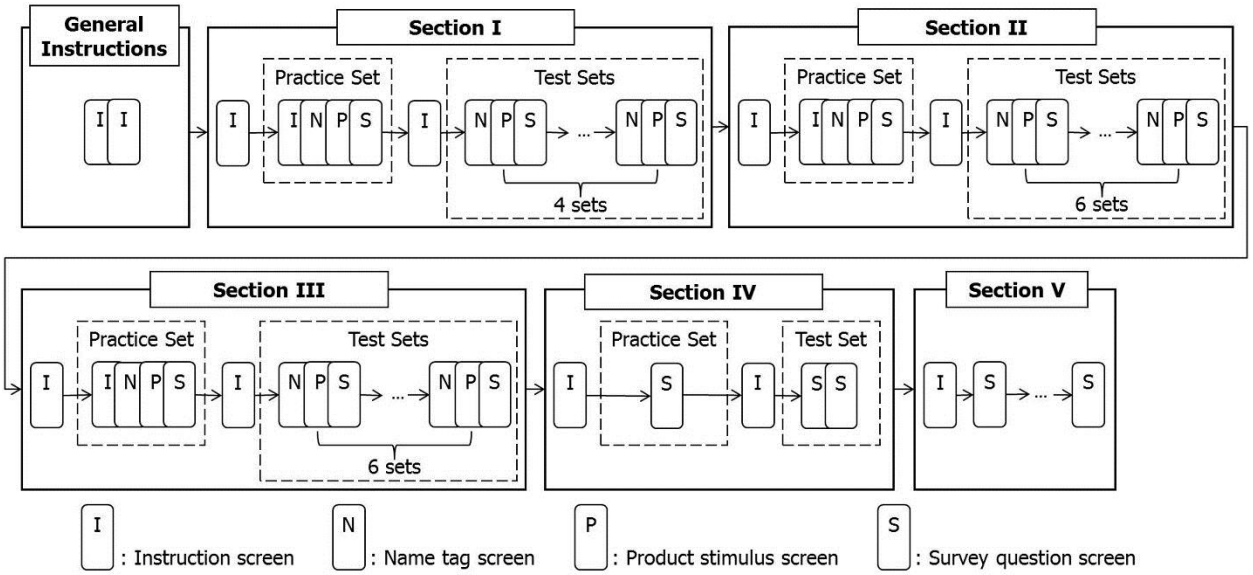


Fig. 4 An illustration of the experiment flow (demonstrated by the SBS condition)

The experiment had five sections, as summarized in Table 2, which took about twenty minutes to complete. Each section began with instructions and a practice question. Sample

Fig. 5 Screens from experiment section II (images from SBS condition, with enlarged text)

survey questions asked in the experiment are provided in Table 4. The experiment sections served different goals. Sections I and II were associated with the importance ratings collected in section IV to test proposition 1/hypotheses: 1a-1d. Section III tested proposition 2/hypotheses: 2a-4b. Product stimuli for section I of the experiment had only feature design variants (no size variants). Stimuli used in section II and III had both size and design variants.

Table 4 Sample survey questions asked in the experiment about cars

Section I	1. Please compare car B to car A and indicate your preference using the following scale. (sliding scale from 1=“strongly prefer car A” to 8=“strongly prefer car B”) 2. Please evaluate your decision according to the following requirements. (1) Please think about the car you prefer in this pair and rate your satisfaction with the decision using the following scale. (sliding scale from 1=“very unsatisfied” to 8=“very satisfied”) (2) Please rate for car A using the following scale. (sliding scale from 1=“very bad” to 8=“very good”) (3) Please rate for car B using the following scale. (sliding scale from 1=“very bad” to 8=“very good”)
Section II	Given these two options of cars, which one do you prefer? Please use the following scale to rate your preference for the two cars. (sliding scale from 1=“strongly prefer car A” to 8=“strongly prefer car B”)
Section III	Please identify the differences between these two cars (just list the names of the parts which are different).
Section IV	We have divided a basic car model into nine components: hood/windshield, grill, headlight, bumper/lower grill, wheel, side door, side mirror, side window, and tail, as shown in the image below. How important are these different components' design in forming your preference for the car? Please rate the importance for the design of these components respectively using the following scales. (sliding scale from 1=“not important at all” to 7=“very important”)

Six pairs of cars and six pairs of electric bicycles were prepared for section I (for a total of 24 stimuli). In this section, two randomly-determined pairs of cars and then two randomly determined pairs of electric bicycles were shown to each subject for preference evaluations (eight stimuli per subject). As a result, each of the 12 prepared pairs was seen by four subjects in an experimental condition. After evaluating each pair of stimuli, subjects completed four questions in one screen: (1) indicate preferences using an eight-level scale, which ranged from "strongly prefer product A" to "strongly prefer product B"; (2) rate their satisfaction with the preference decision they just made using an eight-level scale, ranging from "very unsatisfied" to "very satisfied"; and (3)/(4) rate products A and B using an eight-level scale, ranging from "very bad" to "very good". These scales are adapted from Houston and Sherman [40]. Using eight-level scales in these questions forces preference for either product A or B; data that will be analyzed for cancelation/focus behavior [40] in related work.

Nine pairs of cars and nine pairs of electric bicycles were prepared for section II (36 stimuli). In this section, three randomly determined pairs of cars and then three randomly determined pairs of electric bicycles were shown to each subject for preference evaluations (12 stimuli per subject). The stimuli were chosen with the requirement that in each product category the subject saw three different levels of size variants (enlargements) of different features. As an example, three pairs of cars presented to a subject could be a pair in which the headlights of product B were enlarged 15%; a pair in which the grille of product B was enlarged 25%; and a pair in which the side mirror of product B was enlarged 20%. Each of the 18 pairs of stimuli prepared for this section was seen by 12 subjects in an experimental condition. After each pair of stimuli was presented and evaluated, subjects were asked to indicate their preferences using an eight-level scale as in section I.

Section III repeated the stimuli of section II for each subject. After each pair was evaluated, the subject was asked to identify and write down the features that were different between the two stimuli in a pair—the experiment gave no direct indication of the presence of size variants. This written task allowed for comparison in size-noticing between different types of tasks: implicit and explicit size evaluations, which will be addressed in future work.

Section IV collected importance ratings for all stimuli features. Figure 6 shows the car survey question screen. The rated features included both varied features and unvaried ones, as shown in Table 5, predetermined by the authors. The rating screen showed their names and outlined regions. The section IV rating scales are a typical seven-level Likert scale that made it possible for the subjects to indicate a neutral response. This is different from the preference rating scales in section I and II (data not analyzed here), which sought to force a preference, as previously discussed in Sec. 2.4.2. The experiment ended with section V, which collected demographic information.

2.4.3 Subjects. As stated at the beginning of Sec. 4, the experiment was designed for 72 subjects. However, due to initial computer issues which resulted in unrecorded data for 11 subjects, a total of 83 adults from Iowa State University participated in the experiment, compensated with $5 or extra credit. An online screening survey was conducted to make sure subjects met eye-tracking study requirements. They had normal to corrected vision; did not wear bifocals, trifocals, layered lenses, or regression lenses; did not have difficulty reading a computer screen unassisted; and did not have cataracts, eye implants, glaucoma or permanently dilated pupils [19, 41]. Table 6 provides counts of subject gender and ages.

Subjects were randomly assigned to either the Seq or SBS condition. To begin the experiment, subjects were provided with the informed consent document. They were then

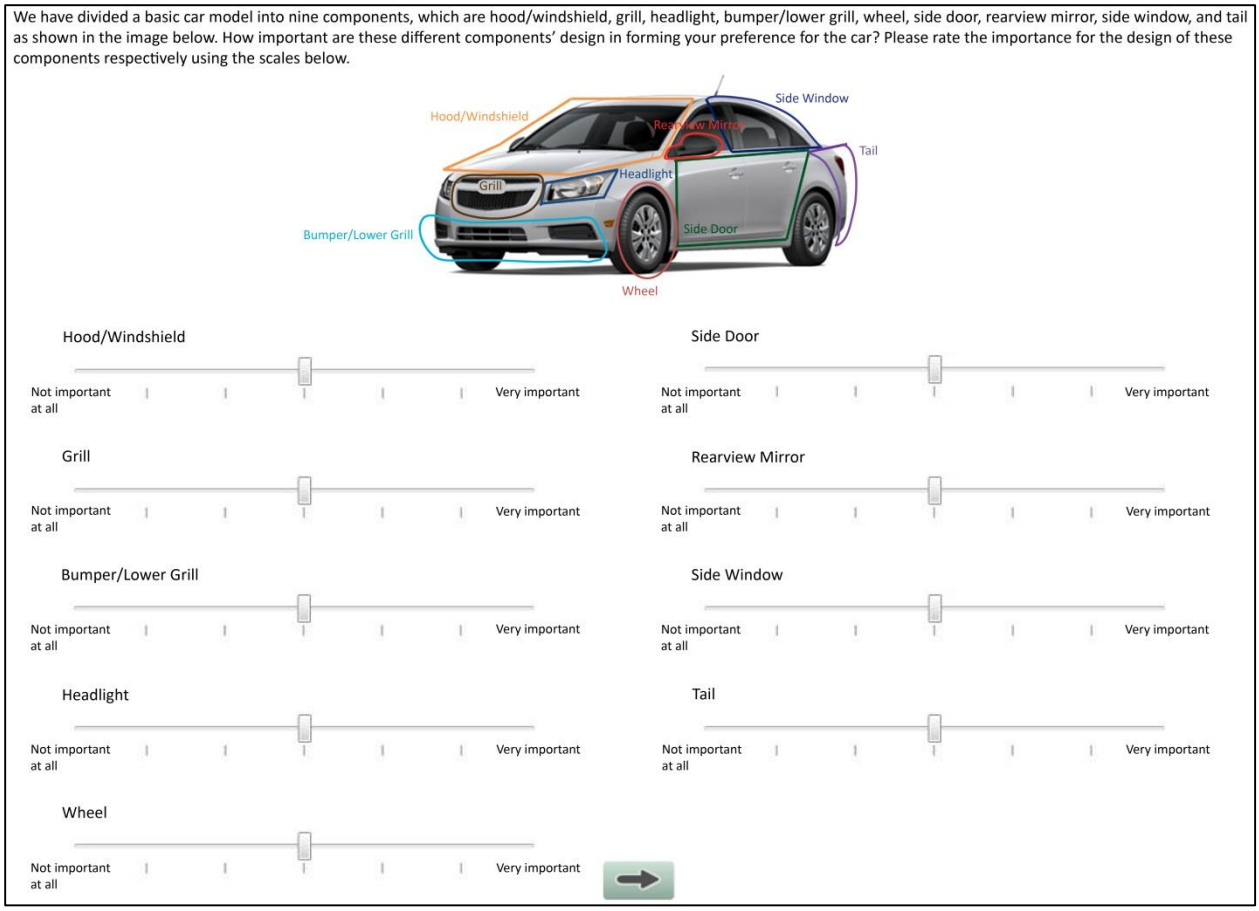


Fig. 6 Car feature-importance-rating survey question screen (rearview mirror referred to as side mirror in this paper)

Table 5 A collection of features rated in the survey

Car								
Varied features				Unvaried features				
Grille	Headlight	Side mirror	Wheel	Bumper/lower grille	Door	Hood/ windshield	Tail	Window

Electric bicycle									
Varied features				Unvaried features					
Cargo box	Footrest	Handlebar	Seat	Front frame	Kick stand	Pedal	Rear frame	Rearview mirror	Tire

Table 6 Counts of subject gender and ages; N= number of subjects

Gender	Male	Female	**Age**	18-24	25-34	35-44	45-54	55-64	65-74
N	37	35	N	27	22	15	4	3	1

instructed to sit in front of the eye tracker, and adjusted themselves such that their eyes were in the optimal position according to the ATTENTION TOOL's Eye Finder. The subjects were

instructed to maintain a consistent posture while completing the experiment. They then performed a calibration exercise, after which the experiment began automatically.

2.4.4 Data Preparation. The gaze data, together with the survey responses, were collected and managed using the ATTENTION TOOL software. The authors used this software to manually define an area of interest (AOI, see Sec. 2.2.1.1), for each feature of each stimulus, shown in Fig. 7. The car wheels, the car headlights, and the electric bicycle's rearview mirrors each required the creation of two AOIs for which the gaze data were combined. Gaze data, organized by AOIs, were exported to R, a free statics software platform, for postprocessing, and then analyzed using the statistical software package JMP.

While the ATTENTION TOOL software worked well to identify fixations overall, it failed for two subjects, which had very few fixations identified for almost all product stimuli. An additional eight subjects had no fixations for only a few stimuli. These missing fixations were identified in postprocessing and treated as missing data in the analysis.

Fig. 7 An example of the AOIs generated for a car

2.5 Proposition 1 Results

The relationships between importance ratings and gaze data from sections I and II were tested separately. The reason for this is that the relevant subject group from section I had only 12 subjects and showed only feature design variants with no size variants, while section II included 36 subjects and showed both design and size variants. The gaze data from section II were used to

fit linear regressions to predict the importance of product features. Data from the Seq condition and the SBS condition were analyzed separately, and data from the car and the electric bicycle were analyzed both separately and jointly. The collected survey data from sections I and II are not analyzed. The survey questions served to make the subjects evaluate the stimuli and their related features, but knowing the results (like what product stimuli are preferred and how satisfying the decisions are) is not the interest of this study.

2.5.1 Experiment Section I Results. Subject-level or individual-level averages were calculated for fixation time, percentage-fixation time, fixation count, and first-located time—throughout Sec. 2.5, these four types of data will be collectively referred to as *gaze data*. As one subject saw two pairs of stimuli (four stimuli) for each product category, to calculate the subject-level averages of the gaze data, four measurements were averaged for each feature. Example calculations for fixation time are shown in Table 7 as $\bar{T}_{1,1}$ through $\bar{T}_{n,19}$, where $I_{1,1}$ through $I_{n,19}$ represent importance ratings directly collected from experiment section IV.

Table 7 Data used for the correlation analysis, demonstrated with fixation time calculations

	Features				
	Headlight		…	Kick stand	
Subject	Fixation time	Importance rating	…	Fixation time	Importance rating
1	$\bar{T}_{1,1}$	$I_{1,1}$	…	$\bar{T}_{1,19}$	$I_{1.19}$
2	$\bar{T}_{2,1}$	$I_{2,1}$	…	$\bar{T}_{2,19}$	$I_{2.19}$
…	…	…	…	…	…
n	$\bar{T}_{n,1}$	$I_{n,1}$	…	$\bar{T}_{n,19}$	$I_{n.19}$
Averages across subjects (used for the correlation test)	$\overline{T_1}$	$\overline{I_1}$	…	$\overline{T_{19}}$	$\overline{I_{19}}$

Next, average gaze data for each feature across all the subjects in an experimental condition were calculated, shown in the last row of Table 7. As a few stimuli for ten subjects were excluded in the analysis due to missing fixations as explained in Sec. 2.4.4, this process ensured each

subject's gaze data contributing equally to the average gaze data. Average importance ratings across the subjects for each feature were also calculated, shown in Table 7 as $\bar{I}_1$ through $\bar{I}_{19}$.

The average gaze data and average importance ratings were used to conduct Pearson correlation tests, shown in Table 8. Conclusions indicated from tests are consistent: there are significant positive correlations between feature importance rating and fixation time, percentage-time and count; and there is a significant negative correlation between feature importance rating and the first-located time on the feature. Hypotheses 1a-1d are strongly supported by these results.

Table 8 Correlations between feature's average importance and associated average gaze data for sections I and II ($^{+}$ p<0.1, * p<0.05, ** p<0.01, *** p<0.0001, one-tailed test)

Fixation Metric		Time (ms)	% Time	Count	First-located (ms)
ISeq	Car	0.54^{+}	0.59*	0.56^{+}	-0.51^{+}
	Electric bicycle	0.58*	0.65*	0.60*	-0.51^{+}
	Car and electric bicycle combined	0.51*	0.63**	0.54**	-0.54**
ISBS	Car	0.86**	0.89**	0.81**	-0.70*
	Electric bicycle	0.63*	0.58*	0.66*	-0.58*
	Car and electric bicycle combined	0.70**	0.69**	0.70**	-0.63**
IISeq	Car	0.65*	0.68*	0.68*	-0.72*
	Electric bicycle	0.59*	0.58*	0.57*	-0.51^{+}
	Car and electric bicycle combined	0.61**	0.61**	0.61**	-0.58**
IISBS	Car	0.55^{+}	0.56^{+}	0.50^{+}	-0.50^{+}
	Electric bicycle	0.79**	0.81**	0.78**	-0.70*
	Car and electric bicycle combined	0.73**	0.74**	0.71**	-0.64**

To visualize gaze data across feature importance ratings, observations of how a feature was visually evaluated by each subject were grouped based on the subject's importance rating for the feature, i.e., all features that received a rating of 1 had their gaze data averaged together. Thus, seven averages for each type of gaze data were plotted against the corresponding feature importance ratings. Similar trends were obtained for fixation time, percentage-fixation time, and fixation count, so fixation time is used as a demonstration. Data from the car and the electric

bicycle stimuli are combined. Figure 8 shows that for both the ISeq condition and the ISBS condition, there is a clear trend showing that a longer average fixation time is spent for a higher importance rating. These trends are consistent with the significant positive correlations found for hypothesis 1a. The average first-located time was plotted with the importance rating, shown in Fig. 9. The first-located time for a feature decreases with its importance rating in ISeq; while in the ISBS condition, the pattern is less clear. This may be due to the limited data in that condition (12 subjects).

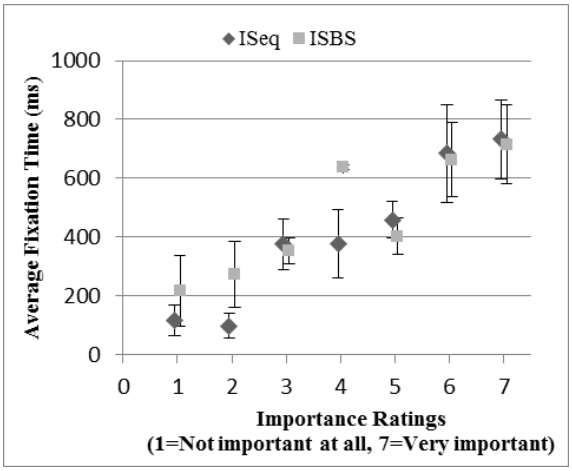


Fig. 8 Average fixation time spent on a feature increases with its importance rating (section I); error bars indicate ±1 standard errors (the two series of data are nudged along the horizontal axis to avoid overlapping of the error bars)

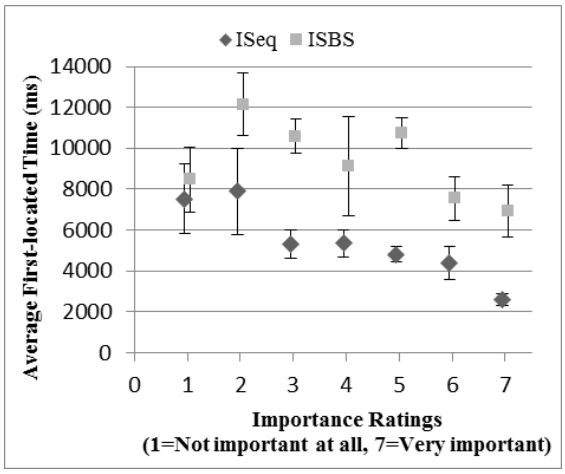


Fig. 9 Average first-located time on a feature varies with importance ratings (section I); error bars indicate ±1 standard errors

2.5.2 Experiment Section II Results. The method described in Sec. 2.5.1 was used with the data in section II as well. Average importance ratings for each feature are provided in Table 9. Results of the Pearson correlations are shown in Table 8. In all situations, there are significantly positive correlations between feature importance rating and the fixation time, percentage-time and count. There is a significantly negative correlation between feature importance rating and the first-located time. Hypotheses 1a-1d are strongly supported by these results. Graphs of trends are shown in Figs. 10 and 11. In both conditions, there is a clear trend showing that the average fixation time spent on a feature increases and the average first-located time on the feature decreases as its importance rating goes up. This indicates that features considered as more important are examined earlier and for longer. When the car and electric bicycle were examined separately, similar trends were observed, as shown in Fig. 12.

Table 9 Average importance ratings for all features

	Car									
	Varied Features				Un-varied Features					
	Grill	Headlight	Side Mirror	Wheel	Bumper/ Lower Grill	Door	Hood/ Windshield	Tail	Window	
Average Importance Rating	4.73	5.57	4.56	5.14	4.37	4.29	5.16	4.20	3.94	
Standard Error	0.16	0.13	0.16	0.16	0.16	0.15	0.17	0.18	0.15	
	Electric Bicycle									
	Varied Features				Un-varied Features					
	Cargo Box	Footrest	Handlebar	Seat	Front Frame	Kick Stand	Pedal	Rear Frame	Rearview Mirror	Tire
Average Importance Rating	5.13	3.57	5.60	6.22	4.54	3.38	4.00	4.00	4.59	4.65
Standard Error	0.17	0.17	0.12	0.11	0.15	0.18	0.18	0.16	0.17	0.19

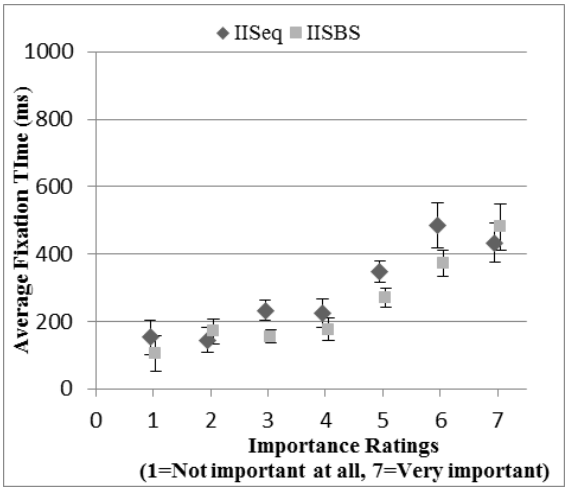


Fig. 10 Average fixation time spent on a feature increases with its importance rating (section II); error bars indicate ±1 standard errors

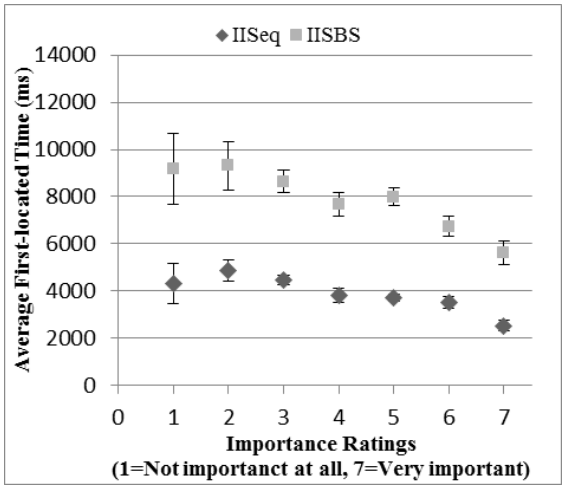


Fig. 11 Average first-located time on a feature decreases with its importance rating (section II); error bars indicate ±1 standard errors

Linear regressions were applied to predict feature importance using the gaze data. The regressions were based on the average gaze data and average importance rating for each feature, which was the same data set as that used to obtain the correlation values shown in Table 8. To avoid any potential multicollinearity problems [42] caused by involving correlated variables, only one type of gaze data was chosen as the independent variable. For each situation considered here, the type of gaze data that had the largest correlation with the importance rating, as indicated in Table 8, was chosen. The summary of the fit is shown in Table 10. In all situations, except for

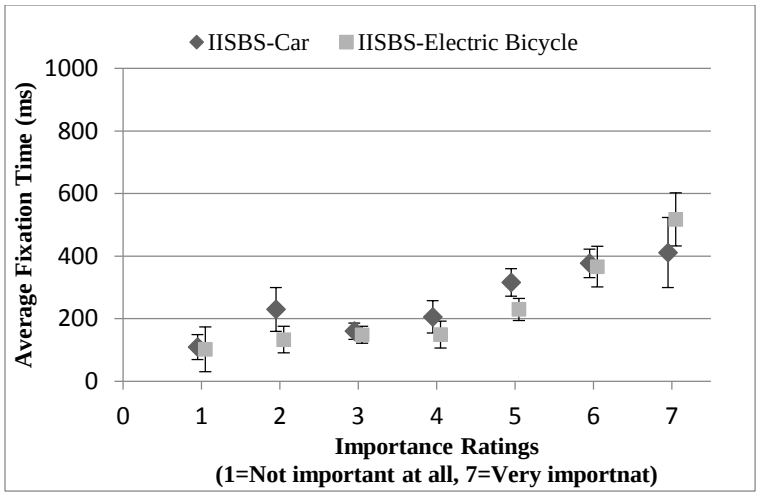


Fig. 12 Trends of average fixation time spent on a feature as its importance varies are similar for the car and the electric bicycle (section II – SBS condition); error bars indicate ±1 standard errors

Table 10 Linear regressions show gaze data predict feature importance; the intercept is the constant term ($^{+}$ p<0.1, * p<0.05, ** p<0.01, *** p<0.0001)

		Independent Variable (x)	Intercept			Coefficient for x		
			Estimate	Std. Error	t	Estimate	Std. Error	t
IISeq	Car	First-located time	5.90	0.48	12.30 ***	-3.50e-4	1.29e-4	-2.72*
	Electric bicycle	Fixation time	3.93	0.36	10.85 ***	1.83e-3	8.85e-4	2.07^{+}
	Car and electric bicycle combined	Fixation time	3.98	0.23	17.29 ***	1.80e-3	5.68e-4	3.16**
IISBS	Car	Percentage-fixation time	4.16	0.33	12.64 ***	0.21	0.12	1.79
	Electric bicycle	Percentage-fixation time	3.91	0.27	14.70 ***	0.34	0.09	3.86**
	Car and electric bicycle combined	Percentage-fixation time	3.96	0.19	20.50 ***	0.30	0.07	4.56**

the car in the SBS condition, the chosen types of gaze data are significant predictors of the feature importance.

2.6 Proposition 2 Results

Based on the write-in responses from experiment section III (indicating what size changes are noticed), data from the AOIs of features that had size variants were classified into two sets: noticeable size changes and unnoticeable size changes. These two sets were compared

using fixation time/count and Δ fixation time/count. The results from the IIISeq and IIISBS conditions were analyzed separately and tested by one-way ANOVA.

2.6.1 Results from Experiment Section III, Sequential (IIISeq). As detailed in Table 11, noticeable size changes in product B have significantly larger values of average fixation time and count than unnoticeable ones. For feature pairs, for example, the two car grilles in a stimuli pair, noticeable-size-change feature pairs have a significantly larger average fixation time and count than unnoticeable ones. Average Δ fixation time for the noticeable size changes is significantly different from the unnoticeable ones, with the former value above zero and the latter one below zero. Average Δ fixation count shows the similar results. These results strongly support hypotheses 2a, 2b and 4a, 4b.

Table 11 Proposition 2 is supported by results (* p<0.05, ** p<0.01, *** p<0.0001)

		Condition	
Associated AOI	**Fix. metric**	Seq	SBS
		Noticeable vs. Unnoticeable	Noticeable vs. Unnoticeable
Size-changed feature in Product B	Time	1080ms vs. 400ms ***	906ms vs. 483ms **
	Count	4.13 vs. 1.63 ***	3.66 vs. 2.38 **
Feature pair with size change	Time	1843ms vs. 1110ms *	1697ms vs. 919ms ***
	Count	7.22 vs. 4.33 **	7.25 vs. 4.43 **
Feature pair with size change	Δ time	319ms vs. -303ms **	115ms vs. 47ms
	Δ count	1.03 vs. -1.06 **	0.07 vs. 0.32

A logistic regression used gaze data to predict saliency (noticed versus unnoticed) of a size change. This regression is suitable such a binary dependent variable [43, 44]. It models the odds of a size change to be noticed by a subject and estimates "the effects of independent variables on these odds [44]." The following equation represents the standard regression model:

$$P(y = 1) = \frac{1}{1+e^{-(Intercept+\beta x)}} \quad (1)$$

In the model, y is a dummy variable, indicating whether a size change is noticed (1) or not (0); P(…) stands for the probability of an event; β is the coefficient for the independent variable x; intercept is the constant term in the model. The model is fit following the maximum likelihood principle. All gaze data measurements produce significant results when used as the independent variable (separately). Fixation time for a size-changed feature in product B is shown as an example; a summary of the fit is provided in Table 12. The small value of "Prob>ChiSq" demonstrates that the current model with gaze data as an independent variable is significantly better than a model with intercepts alone. The significant nonzero coefficient for the independent variable validates that fixation time for the size-changed feature in product B has a significant effect on differentiating noticeable and unnoticeable size changes. Similar results were obtained for data combined from pairs of size-changed features.

Table 12 Logistic models show fixation time (and other gaze data not shown here) predict saliency of size changes (with the unnoticeable size change as reference level) (** p<0.01, *** p<0.0001)

Seq			**SBS**		
Prob>ChiSq	<0.0001		**Prob>ChiSq**	<0.001	
Term	**Estimate**	**Standard Error**	**Term**	**Estimate**	**Standard Error**
Intercept	2.23***	0.27	Intercept	1.49***	0.22
β for "fixation time for size-changed feature in product B"	-8.97×10^{-4} **	2.41×10^{-4}	β for "fixation time for size-changed feature in product B"	-7.38×10^{-4} **	2.23×10^{-4}

2.6.2 Results from Experiment Section III, Side-by-Side (IIISBS). The analysis in Sec. 2.6.1 was performed for data from the IIISBS condition. Detailed results are demonstrated in Table 11. Noticeable size changes in product B have significantly larger values of average fixation time and count than the unnoticeable ones. When considering the feature pair, noticeable-size-change feature pairs have significantly larger values of average fixation time and

count than unnoticeable ones. Average Δ fixation time and count show no differences between the two sets of features that are compared. Hypotheses 3a and 3b are strongly supported here.

The logistic regression in Sec. 2.6.1 was applied. All available measurements, except the Δ fixation time/count for a feature pair with size change, produce significant results when used as the independent variable separately. Fixation time for a size-changed feature in product B is shown as an example, in Table 12.

2.7 Discussion

The experiment results support both research propositions: (1) feature importance is correlated with gaze data in preference choices between two products and (2) saliency of size changes can be predicted by gaze data.

Hypotheses 1a-1d hold true in all situations tested here. During the processes of making preference decisions, there are positive correlations between feature importance and three types of gaze data: the fixation time, the percentage-time, and the count. Each shows a clear trend with increasing feature importance. These findings are consistent with the conclusions of Bettman et al.: People pay more attention to the information that has a larger weight in achieving the decision goal [23]. There is a negative correlation between feature importance and the feature's first-located time. As the feature importance rises, its first-located time decreases. The results from the SBS condition in section I did not show a clear decreasing trend. As mentioned in Sec. 2.5.1, this may be due to the limited data in that condition. In section I, each condition has only 12 subjects, while in section II, each condition has 36 subjects. In section I, each subject only evaluated two pairs for a product category, while in section II each subject evaluated three pairs for a product category. Evaluating more pairs is more likely to indicate the true measurement of

how the subject evaluates a feature. The negative correlation found between feature importance and the feature's first-located time may suffer if people are evaluating a new product, an extremely unfamiliar product, or a product with new features, because in these cases people may not have clear ideas ready in mind about the feature importance and would look around the product stimulus in a less systematic manner to learn the product. But this concern is mitigated by the significant results found for the electric bicycle, which is less familiar (compared with the car) to the subjects. It may be that even for an unfamiliar product, after a one or two evaluation "burn-in," the negative correlation builds.

Hypotheses 1a-1d were analyzed both separately and together for the car and the electric bicycle stimuli. When analyzed separately, the absolute values of the correlations for the car range from 0.50 to 0.89 and those for the electric bicycle range from 0.51 to 0.81; when analyzed together, the absolute values of the correlations range from 0.51 to 0.74. Similar results are obtained in both cases (separate and together), indicating the potential robustness of the tested correlations across different levels of product familiarity. However, more products would need to be tested to confirm this robustness.

For hypotheses 1a-1d, type I error (falsely accepting the research hypothesis) could occur for a number of reasons. One potential cause for type I error is that an important product feature could attract more gaze attention for reasons other than its importance in preference decisions. For example, some features varied and other features did not. One might expect that the varied features attracted more eye attention because they were changing, and also because they were changing they "primed" subjects to exaggerate their importance ratings (as compared with unvaried features). However, the average importance of the features listed in Table 9 indicate that unvaried features remained important in decisions, see especially the car hood, although for

the electric bicycle, varied features are on average more important in decisions than unvaried ones. Furthermore, the average importance ratings in Table 9 suggest that no one feature received a rating of “6” or “7” from all subjects, lending a useful variability to the data (although the bicycle seat is understandably rated as very important to many subjects).

Other reasons for increased gaze time could include a unique design that requires further mental processing; a feature that occupies a larger area of the screen; or a feature that stands out because it is unrealistic or unharmonious with the rest of the design. Arguments against these sources of error include the inclusion of familiar and unfamiliar stimuli, multiple feature variations, more and less harmonious features, and small and large features. Another potential cause of the type I error could be the participation of engineering students, who may have importance-oriented viewing strategies as compared with normal consumers. This error is mitigated by the fact that engineering students comprise only 21% of the sample population.

Type I error could also occur if the important features are in easy-to-locate positions, which could increase the correlation between first-located time and feature importance. This source of error is mitigated by the strong support of the hypotheses from three other types of gaze data. The use of product photographs as compared with real products could increase the correlation between first-located time and feature importance because real products contain rich information that can distract people and delay the located time for some important features. The differences in gaze data between photographic and real evaluations of products are likely substantial, as is true of many types of product preference data.

Type II error (falsely rejecting the research hypotheses) does not apply in this study because the results for hypotheses 1a-1d are strongly accepted. But when researchers apply the conclusions of this study to other experiments, type II error is a possibility. One factor to

consider is the amount of time provided for stimulus evaluation; providing unlimited time can allow people to "invest" gaze in features that are not important to them. Another potential cause is placing special design efforts (like flashy designs, attractive colors, etc.) on some features, which could enable these features to attract unbalanced gaze attention.

This study proved that there is correlation between the importance of product features and associated gaze data, but they are not perfectly correlated. This suggests that including other variables in the regression would improve prediction of feature importance. Such factors could be (1) determining customers' use of particular features to extrapolate information missing from the decision, like price, comfort, brand, and safety information, and (2) recording customers' willingness-to-pay for designs.

Results from section III of the experiment show that the saliency of a size change can be predicted with gaze data. In both the IIISeq and IIISBS conditions, noticeable and unnoticeable size changes can be differentiated with gaze data. Hypotheses 2a, 2b, 4a, and 4b are strongly supported by the data of IIISeq. A noticeable size change in product B has significantly larger values of fixation time, count, Δ fixation time, and Δ count than an unnoticeable one. The noticeable size change attracts extra attention; while the unnoticeable one is ignored. Therefore, when a feature with an unnoticeable size change appears in two stimuli shown sequentially, its latter appearance (in product B) is considered as a repetition of the former one and attracts less attention. These findings are consistent with Refs. [1, 40]; when a pair of stimuli is evaluated for preference decisions, their shared information is likely to be ignored in its second appearance. Even though not originally hypothesized, results show that a noticeable-size-change feature pair has significantly longer total fixation time and higher count.

In IIISeq, the break between the two stimuli weakens the memory of the first stimulus. It is possible that only abstract representations of the first stimulus remain while its details are overwritten by the second stimulus, according to the "overwriting" explanation for "change blindness" [45]. Even in this situation, there are some size changes that trump overwriting. This is worth further investigation, as exposure to minor product variations is a common situation for customers, for example, when viewing products on websites, such as Amazon.com.

As hypotheses 3a and 3b predict, in the IIISBS condition, a noticeable-size-change feature pair has significantly larger fixation time and count than an unnoticeable one. Even though not hypothesized, in the IIISBS condition, a noticeable size change in product B has significantly larger values of fixation time and count than an unnoticeable one. This may have resulted from the fact that gaze typically moves from left to right, and product B is on the right, thus mimicking sequential behavior even though the two stimuli are shown simultaneously. The Δ fixation time and count do not support a hypothesis for the IIISBS condition. This suggests that different viewing strategies are adopted in the Seq and SBS conditions. Presenting stimuli side-by-side enables pairwise comparisons between options, so it is less likely that the saliency of a size change will be identified with Δ fixation time or count.

For hypotheses 2a - 4b (all accepted), there are a number of sources of type I error. One is the unlimited exposure time of the stimulus, which allows the subjects to carefully check for size changes—with a time limit, gaze patterns may change. Another is the use of digital photographs rather than real products. Subjects can stare at the almost identical images to identify size changes, which could enlarge the gaze attention difference between the noticeable and unnoticeable size changes as compared with reviewing real products. Real products allow for physical interactions, and thus other ways to identify size changes, such as holding small

products against each other or measuring precisely. In this experimental setting (section III), only three features have design variants while size is also changing. This could be a source of type I error: As the number of varied features with design variants increases, the gaze attention spent on noticeable and unnoticeable size changes may become more similar.

Type II error does not apply for the testing of hypotheses 2a-4b in this study as these hypotheses are all accepted. But when researchers apply the conclusions of this study to other experiments, type II error is a possibility. Providing stimuli with size changes that are extremely obvious and require little gaze effort to notice would significantly decrease the gaze differences between the noticeable and unnoticeable size changes. Another potential cause could be including clear and constant reference-of-scale, such as a ruler, close to feature size variants as it would decrease the gaze efforts needed to notice the size change.

All hypotheses are tested under two general conditions, showing stimuli sequentially and side-by-side. Based on the results, the relationships found between feature importance and the gaze data are almost the same in the two conditions. This suggests that researchers studying the importance of product attributes using eye-tracking can present two stimuli at a time (such as in choice decisions) and reliably draw conclusions about attribute importance. It is not necessary to show product stimuli individually. But the two conditions have different results when the Δ fixation time and count are used to differentiate noticeable and unnoticeable size changes: The Δ fixation time and count are useful only in the Seq condition.

2.8 Conclusion

Results from this study indicate that product feature importance is correlated with a variety of gaze data (fixation time, percentage-fixation time, fixation count, and first-located

time). The importance rating of a feature can be predicted by the gaze data using linear regression. These findings can help designers by providing a new approach to identify the importance of product features. They suggest that feature importance can be identified at the individual subject level in only three questions, without directly asking about feature importance. This could (a) significantly reduce the subject's mental burden associated with current methods, such as discrete choice analysis and complex rating schemes and (b) remove context effects caused by drawing attention to the purpose of the experiment (ascertaining feature importance), and instead let subjects evaluate products naturally. These directions will be pursued in future research. The study can also be furthered by investigating the effect of product viewing perspective, sizes, etc., and setting time constraints for viewing the stimuli on the results.

This study also demonstrates that gaze data can be used to identify whether or not someone notices a change in the size of a product feature. This can be used in a variety of ways, such as determining when manufacturing imperfections in the form of geometrical variations are noticeable. A considerable amount of time and money has been spent on manufacturing processes to ensure the quality appearance of products [46, 47]. If one can predict how likely it is that an imperfection will be noticed, optimization analysis can be performed to reduce the manufacturing costs while maintaining the targeted quality appearance of products. This study could be furthered by developing a method to determine the just-noticeable threshold for size changes, which would be immediately useful to practicing designers.

The study's conclusions have some potential sources of type I error, as noted in Sec. 2.7. One area that should be noted, in particular, is the difference in the predictive power of gaze data in the evaluation of digital photographs or renderings versus real products. Studies involving real products are considerably more complex, with difficult-to-create stimuli, expensive eye-tracking

equipment, and difficult-to-decipher gaze data in three dimensions. These challenges all suggest that for the time being, the usefulness of gaze data in understanding product evaluations is most readily applied to computer screen experiments.

Acknowledgment

The authors would like to thank the iMotions company for providing the ATTENTION TOOL software and the associated technical support; Dr. Frederick Lorenz, Dr. Heike Hofmann, and Kent Kroeger at Iowa State University for statistical advice; and colleagues in the IRIS Design Lab in Iowa State University for their valuable input.

References

[1] Sütterlin, B., Brunner, T. A., and Opwis, K., 2008, "Eye-Tracking the Cancellation and Focus Model for Preference Judgments," J. of Exp. Soc. Psychol., **44**(3), pp. 904–911.

[2] Just, M. A., and Carpenter, P. A., 1976, "Eye Fixations and Cognitive Processes," Cognit. Psychol., **8**(4), pp. 441–480.

[3] Nielsen, J., and Pernice, K., 2010, *Eyetracking Web Usability*, New Riders, Berkeley, CA.

[4] Schiessl, M., Duda, S., Thölke, A., and Fischer, R., 2003, "Eye Tracking and Its Application in Usability and Media Research," MMI-Interact. J., **1**(6), pp. 41-50.

[5] Rayner, K., 1998, "Eye Movements in Reading and Information Processing: 20 Years of Research," Psychol. Bull., **124**(3), pp. 372-422.

[6] Findlay, J. M., and Gilchrist, I. D., 1998, "Eye Guidance and Visual Search," *Eye Guidance in Reading and Scene Perception*, G. Underwood, eds., Elsevier, Oxford.

[7] Glaholt, M. G., and Reingold, E. M., 2011, "Eye Movement Monitoring as a Process Tracing Methodology in Decision Making Research," J. Neurosci., Psychol., and Econ., **4**(2), pp. 125-146.

[8] Rosbergen, E., Wedel, M., and Pieters, R., 1997, "Analyzing Visual Attention to Repeated Print Advertising Using Scanpath Theory," Research Institute SOM, University of Groningen, Groningen, Netherlands, Report No. 97B32.

[9] Pieters, R., and Warlop, L., 1999, "Visual Attention During Brand Choice: The Impact of Time Pressure and Task Motivation," Int. J. of Res. in Mark., **16**(1), pp. 1-16.

[10] Pieters, R., and Wedel, M., 2004, "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-size Effects," J. of Mark., **68**(2), pp. 36-50.

[11] Jacob, R. J. K., and Karn, K. S., 2003, "Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises," *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, J. Hyönä, R. Radach, and H. Deubel, eds., Elsevier, Amsterdam, pp. 573-605.

[12] Poole, A., and Ball, L. J., 2005, "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects," *Encyclopedia of Human Computer Interaction*, C. Ghaoui, eds., Idea Group Reference, PA.

[13] Konstantopoulos, P., 2009, "Investigating Drivers' Visual Search Strategies: Towards an Efficient Training Intervention," Ph.D. thesis, University of Nottingham, Nottingham, UK.

[14] Lohse, G. L., and Johnson, E. J., 1996, "A Comparison of Two Process Tracing Methods for Choice Tasks," Org. Behav. and Human Decis. Process., **68**(1), pp. 28-43.

[15] Payne, J. W., Bettman, J. R., and Johnson, E. J., 1988, "Adaptive Strategy Selection in Decision Making," J. of Exp. Psychol.: Learning, Memory, and Cognition, **14**(3), pp. 534-552.

[16] Dhar, R., Nowlis, S. M., and Sherman, S. J., 1999, "Comparison Effects on Preference Construction," J. of Consum. Res., **26**(3), pp. 293-306.

[17] Gofman, A., Moskowitz, H. R., Fyrbjork, J., Moskowitz, D., and Mets, T., 2009, "Extending Rule Developing Experimentation to Perception of Food Packages with Eye Tracking," Open Food Sci. J., **3**, pp. 66-78.

[18] Koivunen, K., Kukkonen, S., Lahtinen, S., Rantala, H., and Sharmin, S., 2004, "Towards Deeper Understanding of How People Perceive Design in Products," *CADE2004 Web Proceedings of Computers in Art and Design Education Conference*, M. A. Eriksen., L. Malmborg, and J. Nielsen, eds., Sweden.

[19] Reid, T., MacDonald, E., and Du, P., 2013, "Impact of Product Design Representation on Customer Judgment," ASME J. of Mech. Des., **135**(9), p. 091008.

[20] iMotions, 2013, retrieved on Dec. 20, 2013, http://imotionsglobal.com/

[21] Qualtrics, 2013, retrieved on Dec. 20, 2013, http://qualtrics.com/

[22] Duchowski, A. T., 2007, *Eye Tracking Methodology: Theory and Practice*, Springer, London.

[23] Bettman, J. R., Luce, M. F., and Payne, J. W., 1998, “Constructive Consumer Choice Processes,” J. of Consum. Res., **25**(3), pp. 187-217.

[24] Myers, J. H., and Alpert, M. I., 1968, “Determinant Buying Attitudes: Meaning and Measurement," J. of Mark., **32**(4), pp. 13-20.

[25] Banks, S., 1950, “The Relationship Between Preference and Purchase of Brands,” J. of Mark., **15**(2), pp. 145-157.

[26] Orsborn, S., Cagan, J., and Boatwright, P., 2009, “Quantifying Aesthetic Form Preference in a Utility Function,” ASME J. of Mech. Des., **131**(6), p. 061001.

[27] MacDonald, E. F., Gonzalez, R., and Papalambros, P., 2009, “The Construction of Preferences for Crux and Sentinel Product Attributes," J. of Eng. Des., **20**(6), pp. 609-626.

[28] Jaccard, J., Brinberg, D., and Ackerman, L., 1986, “Assessing Attribute Importance: A Comparison of Six Methods,” J. of Consum. Res., **12**(4), pp. 463-468.

[29] Reisen, N., Hoffrage, U., and Mast, F. W., 2008, “Identifying Decision Strategies in A Consumer Choice Situation,” Judgm. Decisi. Making, **3**(8), pp. 641-658.

[30] Warell, A., and Nåbo, M., 2002, “Handling Product Identity and Form Development Issues in Design Management Using Design Format Modeling,” Proceedings of DMI 2002, the 11th International Forum on Design Management Research and Education Strategies, Resources & Tools for Design management Leadership, Northeastern University, Boston.

[31] Jaccard, J., and King, G. W., 1977, “The Relation Between Behavioral Intentions and Beliefs: A Probabilistic Model,” Human Commun. Res., **3**(4), pp. 326-334.

[32] Schkade, D., and Johnson, E., 1989, “Cognitive Processes in Preference Reversals," Org. Behav. and Human Decis. Process., **44**(2), pp. 203-231.

[33] Michalek, J. J., Ebbes, P., Adiguzel, F., Feinberg, F. M., and Papalambros, P. Y., 2011, “Enhancing Marketing With Engineering: Optimal Heterogeneous Markets,” Int. J. of Res. Mark., **28**(1), pp. 1-12.

[34] Coelho do Vale, R., Pieters, R., and Zeelenberg, M., 2008, “Flying Under the Radar: Perverse Package Size Effects on Consumption Self-Regulation,” J. Consum. Res., **35**(3), pp. 380-390.

[35] Chandon, P., and Ordabayeva, N., 2008, “Downsize in 3D, Supersize in 1D: Effects of the Dimensionality of Package and Portion Size Changes on Size Estimations, Consumption, and Quantity Discount Expectations,” INSEAD Working Papers Collection, Paper No. 2008/46/MKT.

[36] Yang, S., and Raghubir, P., 2005, “Can Bottles Speak Volumes? The Effect of Package Shape on How Much to Buy,” J. Retailing, **81**(4), pp. 86-96.

[37] Krider, R. E., Raghubir, P., and Krishna, A., 2001, "Pizzas: π or Square? Psychophysical Biases in Area Comparisons," Mark. Sci., **20**(4), pp. 405-425.

[38] Chevrolet, 2013, retrieved on Dec. 20, 2013, http://www.chevrolet.com/flash.html

[39] Global-Tradekey, 2013, retrieved on Apr. 21, 2013, http://www.global-tradekey.com/company/G02189/C64201.htm

[40] Houston, D. A., and Sherman, S. J., 1995, "Cancellation and Focus: The Role of Shared and Unique Features in the Choice Process," J. Exp. Soc. Psychol., **31**(4), pp. 357-378.

[41] Pernice, K., and Nielsen, J., 2009, "Eyetracking Methodology: How to Conduct and Evaluate Usability Studies Using Eyetracking," Nielsen Norman Group, Fremont, CA.

[42] Paul, R. K., "Multicollinearity: Causes, Effects and Remedies," IASRI, New Delhi.

[43] Agresti, A., 1996, *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Inc., New York.

[44] O'Connell, A. A., 2006, *Logistic Regression Models for Ordinal Response Variables*, SAGE Publications, Inc, Thousand Oaks, CA.

[45] Simons, D. J., 2000, "Current Approaches to Change Blindness," Visual Cognit., **7**(1-3), pp. 1-15.

[46] Forslund, K., Kero, T., and Söderberg, R., 2009, "Appearance FMEA: A Method for Appearance Quality Evaluation of Early Design Concepts," Proceedings of the ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, San Diego, CA, pp. 217-225.

[47] Söderberg, R., Wickman, C., and Lindkvist, L., 2008, "Improving Decision Making by Simulating and Visualizing Geometrical Variation in Non-Rigid Assemblies," CIRP Ann., **57**(1), pp. 175-178.

CHAPTER 3

PRODUCTS' SHARED VISUAL FEATURES DO NOT CANCEL IN CONSUMER DECISIONS

A paper published in *Journal of Mechanical Design*

Ping Du
Department of Mechanical Engineering, Iowa State University, Ames, IA 50011

Erin MacDonald
Assistant Professor, Department of Mechanical Engineering, Stanford University, Stanford, CA, 94305

Abstract

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.

3.1 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. 3.2 and 3.3. 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. 3.2, 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

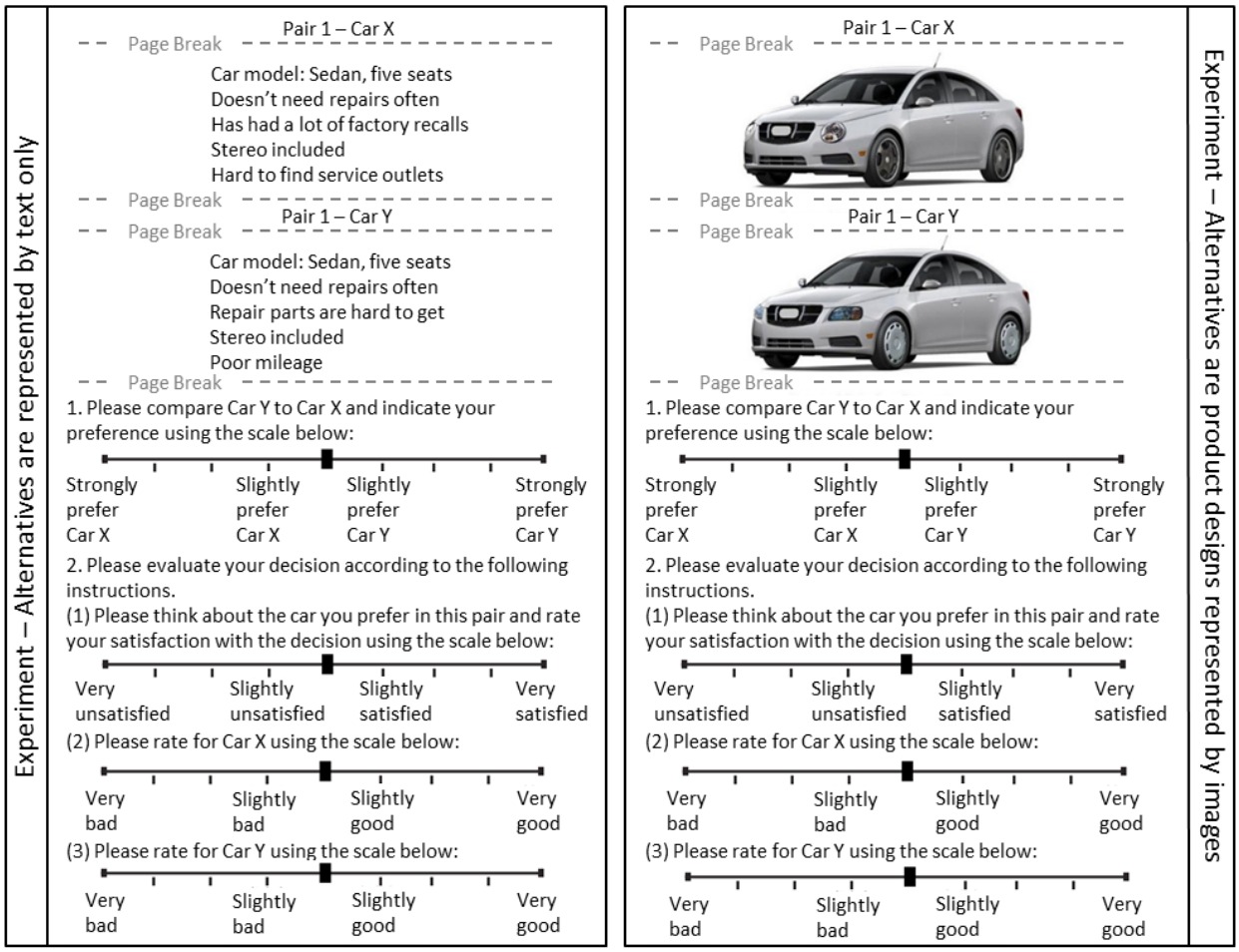


Fig. 1 The original testing scenario for the C&F model (a) and an additional scenario tested here (b) that involves product designs shown as images, which had both shared and unique feature designs.

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. 3.2.1. 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.

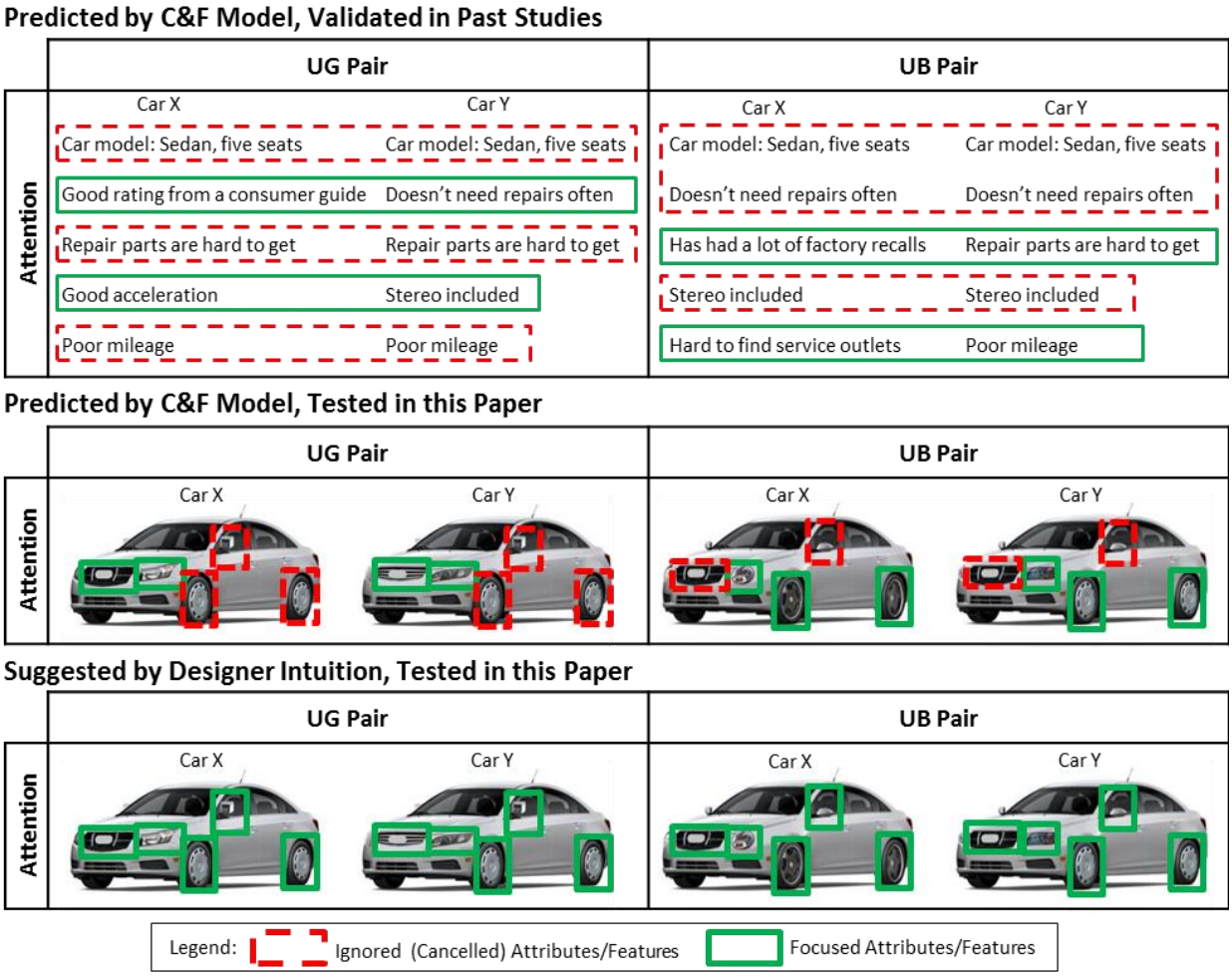


Fig. 2 The C&F model predicted preferences for text described alternatives. We tested if C&F held for product designs shown as images, as opposed to designer intuition, and also combinations of text and images

We test the effectiveness of the C&F model in six conditions using the research hypotheses listed in Sec. 3.3. 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. 3.2, 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 shot-cuts (cancel and focus) that may otherwise lay dormant. Details about experiment stimuli and experiment design are provided in Sec. 3.4. Experiment results are presented in Sec. 3.5. Discussion is provided in Sec. 3.6. Section 3.7 concludes the study.

3.2 Background

3.2.1 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 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.

	Automobile X	Automobile Y
UG pair	Doesn't need repairs often Stereo included Prestigious model Air conditioning included Hard to find service outlets Poor warranty Poor mileage High priced	Good financing available Good ratings from a consumer guide Good acceleration A friend recommended this model Hard to find service outlets Poor warranty Poor mileage High priced
UB pair	Doesn't need repairs often Stereo included Prestigious model Air conditioning included Hard to find service outlets Poor warranty Poor mileage High priced	Doesn't need repairs often Stereo included Prestigious model Air conditioning included High insurance costs Has had a lot of factory recalls Available in only a few colors Repair parts are hard to get

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. Alternatives X and Y in the UG pair shared the same bad attributes (e.g. "poor warranty" and "poor mileage"), but have UG attributes (e.g. "doesn't need repairs often" and "good financing available"). According to the C&F model, when people are comparing the two alternatives in a UG pair, effects of the shared-bad attributes are canceled leaving the effects of the UG attributes, and more decision

weight is given to the UG attributes of the Subject; therefore, people are more likely to prefer the Subject because of the prominent good attributes. Similarly for a UB pair, the effects of the shared-good attributes are canceled, leaving the effects of the UB attributes. As people place more decision weight on the Subject's UB attributes, they are less likely to prefer the Subject and instead prefer the Referent.

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. 3.5.3 and 3.6, 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 UG 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.

3.2.2 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.

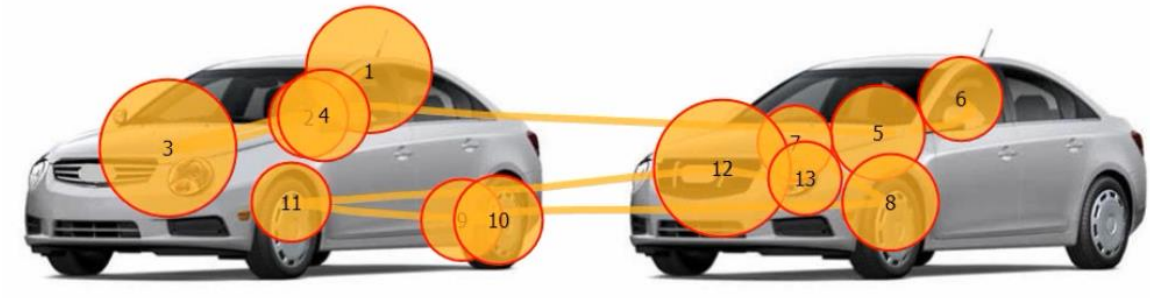


Fig. 3 Illustration of fixations. A circle represents a fixation; a larger circle indicates a longer fixation time; more circles indicates a higher fixation count

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. Shimojo 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.

3.3 Research Hypotheses

As described in Sec. 3.2, 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 Y" for the UG pair (G), more so than for the UB pair (B):*

$$\overline{V_G} - \overline{V_B} > 0 \tag{1}$$

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

$$\overline{S_G} - \overline{S_B} > 0 \tag{2}$$

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

$$\overline{\Gamma_{GA}} - \overline{\Gamma_{BA}} > 0, \overline{\Gamma_{GR}} - \overline{\Gamma_{BR}} > 0 \tag{3}$$

Explanation of Hypothesis 1A: According to the C&F model and as explained in Sec. 3.2.1, 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:*

$$\overline{T_U} - \overline{T_H} > 0, \overline{Q_U} - \overline{Q_H} > 0 \tag{4}$$

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') lean to "strongly prefer Subject (gaze)" for the UG pair, more so than for the UB pair:*

$$\overline{V'_G} - \overline{V'_B} > 0 \tag{5}$$

3.4 Methodology

Testing of the C&F model was realized by part I of a five-part computer-based experiment, as described in Sec. 3.4.2. Results from other parts of the same experiment are reported in Ref. [19], which includes descriptions of the other parts of the survey. Section 3.4.1 details the preparation of experiment stimuli. Section 3.4.3 summarizes the experiment participant population. Section 3.4.4 describes data preparations prior to statistical analysis.

3.4.1 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-only	Text-only	Image-with-text
	Car model: Sedan, five seats Doesn't need repairs often Has had a lot of factory recalls Stereo included Hard to find service outlets	 Car model: Sedan, five seats Doesn't need repairs often Has had a lot of factory recalls Stereo included Hard to find service outlets

Fig. 4 Three representation modes of the car stimuli: image-only, text-only, and image-with-text

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, grilles, side mirrors, and wheels; and electric bicycles had varied handlebars, seats, footrests, and cargo boxes. 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

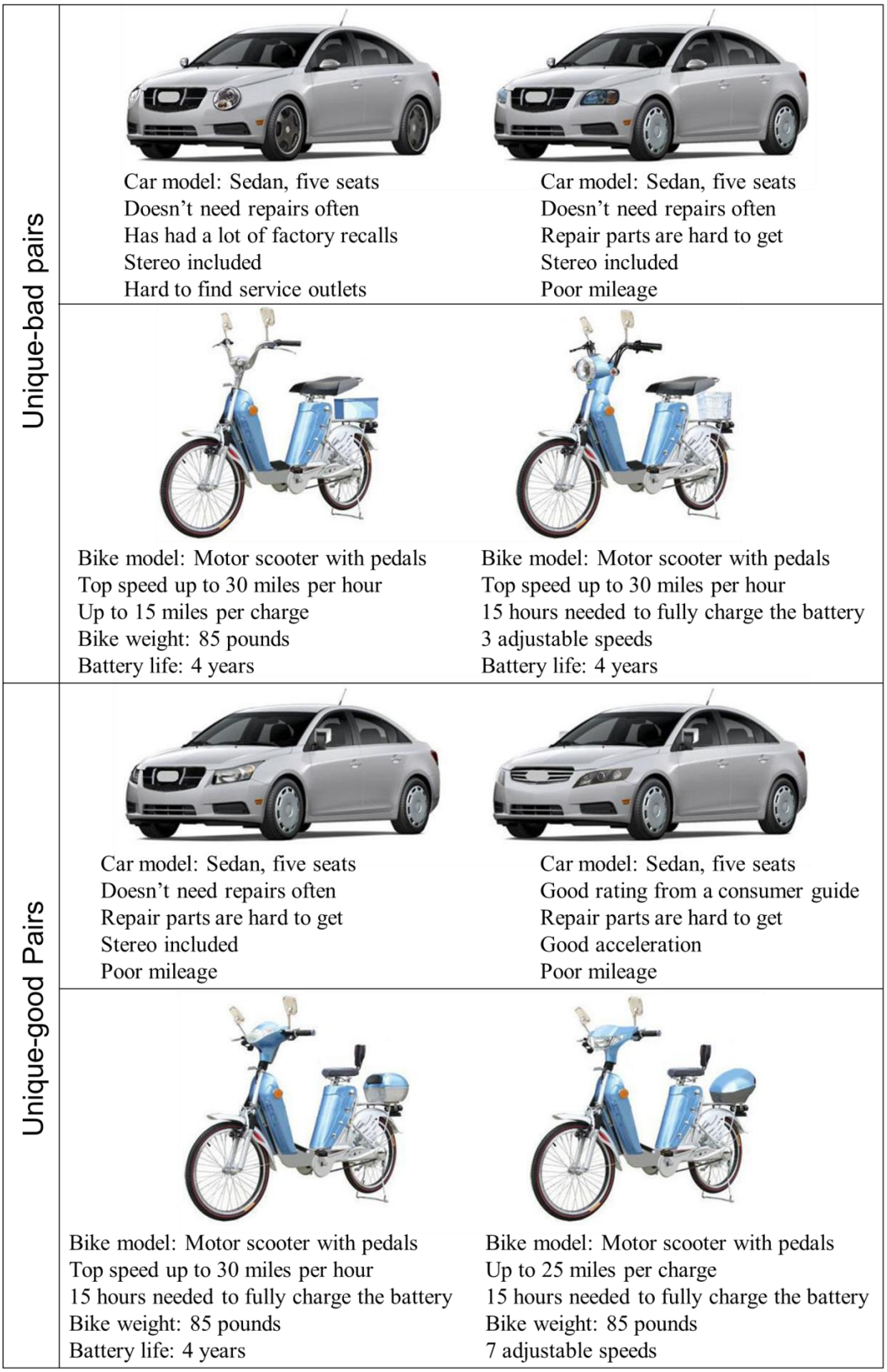


Fig. 5 Sample stimuli in the ITSBS condition

required careful effort. To form the UG 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 desirable." Using these data, the design variants were verified as good or bad using a desirability rating of 4.5 (the middle of the scale) as a split point. A design variant with an average rating greater than 4.5 was verified as good; lower than 4.5 was verified as bad. No design variants had an average rating exactly equal to 4.5. We selected the good and bad design variants with the most extreme average desirability ratings, as shown in Fig. 6, to create stimuli for the experiment. We used t-tests to validate that the selected good and bad variants had statistically significant differences in ratings, except for a single good design variant of the footrest, which followed the trend but did not achieve significance. No design variants of the grille and only one design variant of the seat were verified as bad, indicated by "—" in Fig. 6. Therefore, the experiment did not use the grille as a "unique-bad" or "shared-bad" feature, but only used it as a "unique-good" or "shared-good" feature.

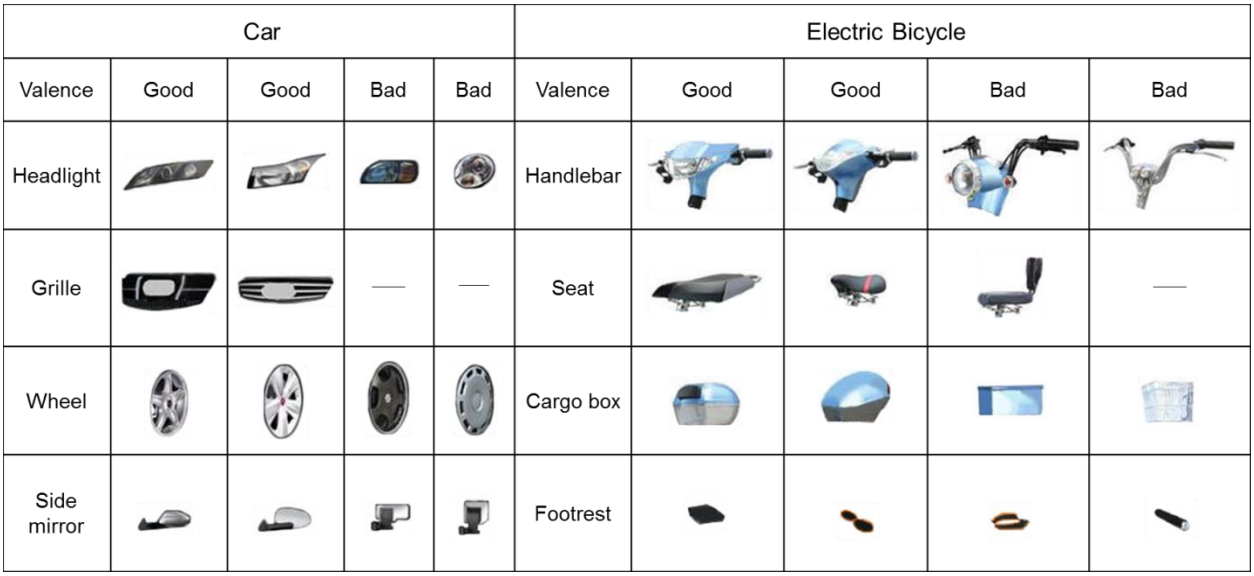


Fig. 6 Design variants of features explicitly selected to be "good" and "bad" were verified with a t-test of desirability ratings

To create the experimental stimuli, three UG and three UB pairs of image stimuli were formed for each product category. In a UG pair, the two image stimuli had shared-bad design variants for two varied features and had UG design variants for the other two varied features. Accordingly, two image stimuli in a UB pair had shared-good design variants for two varied features and had UB design variants for the other two varied features.

Creating good/bad text attributes was approached with a similar 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.

3.4.2 Experiment Design. The experiment had six conditions: Image & Sequential (ISeq), Text & Sequential (TSeq), Image-with-Text & Sequential (ITSeq), Image & Side-by-Side (ISBS), 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. 3.3. (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 indication 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.

3.4.3 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 ITSeq), because authors of Ref. [2] found 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].

3.4.4 Data Preparation. ATTENTION 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.

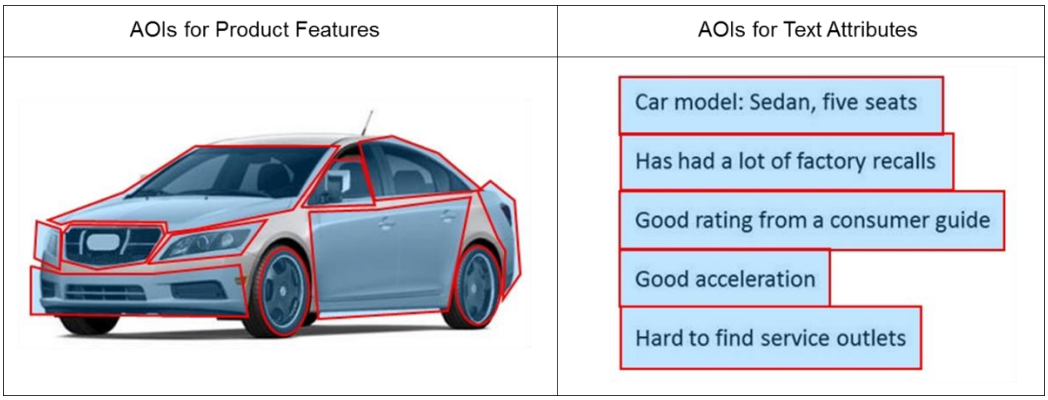


Fig. 7 Sample AOIs generated for the product attributes/features

3.5 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 3.5.1 details the results for Hypotheses 1A-1C, which are based on the survey data. Section 3.5.2 details the results for Hypothesis 2, which are based on the gaze data. Section 3.5.3 details the results for Hypothesis 3, which are based on both the survey and gaze data.

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

$$\overline{V_{Gi}} = (V_{GCi} + V_{GEi})/2,\ \overline{V_{Bi}} = (V_{BCi} + V_{BEi})/2 \tag{6}$$

$$\overline{V_G} = \sum_{i=1}^{N} \overline{V_{Gi}}\,/N,\ \overline{V_B} = \sum_{i=1}^{N} \overline{V_{Bi}}\,/N \tag{7}$$

Hypothesis 1B: For each participant (i), an individual-level average UG satisfaction rating ($\overline{S_{Gi}}$) and UB satisfaction rating ($\overline{S_{Bi}}$) were calculated by averaging satisfaction ratings the participant gave to the UG pair of cars (S_{GCi}) and that of electric bicycles (S_{GEi}), and by averaging the ratings for the UB pair of cars (S_{BCi}) and that of electric bicycles (S_{BEi}), respectively, as indicated in Eq. (8). Pairwise t-tests tested if the difference between UG

satisfaction rating ($\overline{S_G}$) and UB satisfaction rating ($\overline{S_B}$) was greater than 0. Equation (9) shows the calculation of $\overline{S_G}$ and $\overline{S_B}$. Table 2 provides the results.

$$\overline{S_{Gi}} = (S_{GCi} + S_{GEi})/2,\ \overline{S_{Bi}} = (S_{BCi} + S_{BEi})/2 \tag{8}$$

$$\overline{S_G} = \sum_{i=1}^{N} \overline{S_{Gi}}/N,\ \overline{S_B} = \sum_{i=1}^{N} \overline{S_{Bi}}/N \tag{9}$$

Table 2 There are differences in preferences and postpreference evaluations between the UG and UB pairs in some cases ($^{+}$ p<0.1, * p<0.05, ** p<0.01)

	ISeq	TSeq	ITSeq	ISBS	TSBS	ITSBS
Participants (N)	23	24	24	12	12	12
Choice Rating $\overline{V_G}$ - $\overline{V_B} > 0$	4.33 - 4.30	4.67 - 3.81 **	4.67 - 4.60	4.04 - 4.63	4.25 - 4.54	4.5 - 4.5
Satisfaction $\overline{S_G}$ - $\overline{S_B}$ >0	5.00 - 5.39	4.98 - 4.90	5.02 - 5.02	5.67 - 6.00	5.67 - 5.25	5.54 - 4.71 *
Good-ness Ac. $\overline{\Gamma_{GA}}$ - $\overline{\Gamma_{BA}}$ >0	5.20 - 5.67	5.38 - 5.13	5.65 - 5.21 **	5.71 - 6.00	5.54 - 5.08 +	5.42 - 4.67 **
Good-ness Rj. $\overline{\Gamma_{GR}}$ - $\overline{\Gamma_{BR}}$ >0	4.09 - 4.30	4.40 - 3.79 **	4.19 - 3.90 *	4.5 - 4.63	4.38 - 3.71 **	4.29 - 3.79 *

Hypothesis 1C: For each participant (i), an individual-level average UG good-ness rating for the accepted alternative ($\overline{\Gamma_{GAi}}$) and UB good-ness rating for the accepted alternative ($\overline{\Gamma_{BAi}}$) were calculated by averaging good-ness ratings the participant gave to the accepted alternative in the UG pair of cars (Γ_{GACi}) and that of electric bicycles (Γ_{GAEi}), and by averaging the ratings for the accepted alternative in the UB pair of cars (Γ_{BACi}) and that of electric bicycles (Γ_{BAEi}), 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 ($\overline{\Gamma_{GA}}$) and UB good-ness rating for the accepted alternative ($\overline{\Gamma_{BA}}$) was greater than 0. $\overline{\Gamma_{GA}}$ and $\overline{\Gamma_{BA}}$ 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.

$$\overline{\Gamma_{GAi}} = (\Gamma_{GACi} + \Gamma_{GAEi})/2, \overline{\Gamma_{BAi}} = (\Gamma_{BACi} + \Gamma_{BAEi})/2 \tag{10}$$

$$\overline{\Gamma_{GA}} = \sum_{i=1}^{N} \overline{\Gamma_{GAi}}/N,\ \overline{\Gamma_{BA}} = \sum_{i=1}^{N} \overline{\Gamma_{BAi}}/N \tag{11}$$

3.5.2 Analysis and Results: Gaze Data. *Hypothesis 2:* Individual-level average fixation times spent on a unique attribute/feature ($\overline{T_{Ui}}$) was calculated by averaging the fixation time a participant spent on all the unique attributes and/or features of the car and the electric bicycle, as shown in Eq. (12), where T_{UCil} and T_{UEil} are fixation time that participant i spent on the lth unique attribute/feature of the car and of the electric bicycle, respectively; K_{UC} and K_{UE} are the number of unique attributes/features of the car and of the electric bicycle, respectively. Similarly, an individual-level average fixation time spent on a shared attribute/feature ($\overline{T_{Hi}}$) was calculated by participant, as shown in Eq. (12). The features and text attributes that remained the same among all stimuli for the car and the electric bicycle were considered as basic features/attributes and were not included in the analysis. Pairwise t-tests tested if the difference between the unique attribute/feature's fixation time ($\overline{T_U}$) and the shared attribute/feature's fixation time ($\overline{T_H}$) was greater than 0. Equation (13) shows the calculation of $\overline{T_U}$ and $\overline{T_H}$. Table 3 provides the results, and also reports the number of excluded stimuli; see Sec. 3.4.4 for explanation.

$$\overline{T_{Ui}} = (\sum_{l=1}^{K_{UC}} T_{UCil} + \sum_{l=1}^{K_{UE}} T_{UEil})/(K_{UC} + K_{UE}),$$
$$\overline{T_{Hi}} = (\sum_{l=1}^{K_{HC}} T_{HCil} + \sum_{l=1}^{K_{HE}} T_{HEil})/(K_{HC} + K_{HE}) \tag{12}$$

$$\overline{T_U} = \sum_{i=1}^{N} \overline{T_{Ui}}/N,\ \overline{T_H} = \sum_{i=1}^{N} \overline{T_{Hi}}/N \tag{13}$$

The same analysis was performed on fixation count. Pairwise t-tests tested if average fixation count for the unique attribute/feature ($\overline{Q_U}$) was greater than that for the shared one ($\overline{Q_H}$); refer to Table 3.

Table 3 Different abilities of the unique and shared attributes/features to attract gaze attention ($^{+}$ p<0.1, * p<0.05, ** p<0.01)

	ISeq	TSeq	ITSeq	ISBS	TSBS	ITSBS
N	23	24	24	12	12	12
Stimuli (Excluded)	192 (11)	192 (5)	192 (1)	48 (1)	48(1)	48(1)
Fixation time $\overline{T_U}$ - $\overline{T_H}$ (ms)	532 - 433 +	1283 - 1157 *	498 - 484	911 - 320 **	2019 - 1196 **	391 - 297 **
Fixation count $\overline{Q_U}$ - $\overline{Q_H}$	2.1 – 1.85	6.72 – 6.06 **	2.65 – 2.65	3.34 – 1.53 **	10.66 – 6.4 **	2.31 – 1.8 **

3.5.3 Analysis and Results: Survey and Gaze Data Combined. *Hypothesis 3:* First, the Subject (gaze) and Referent (gaze) alternatives for each pair of stimuli that a participant saw, as defined in Sec. 3.3, were identified based on fixation time (the Subject having the longer fixation time). Then, the choice rating given by a participant for each stimulus was transformed to range from "strongly 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.

$$V' = \begin{cases} V, & if\ Subject(gaze) is\ Product\ Y \\ 9 - V, & if\ Subject(gaze) is\ Product\ X \end{cases} \tag{14}$$

The same analysis used for Hypothesis 1A in Sec. 3.5.1 was performed on V' here. Pairwise t-tests tested if the difference between UG transformed rating ($\overline{V'_G}$) and UB transformed choice rating ($\overline{V'_B}$) was greater than 0. Table 4 provides the results.

Table 4 Differences in preferences between the UG and UB pairs (* p<0.05)

	ISeq	TSeq	ITSeq	ISBS	TSBS	ITSBS
N	21	23	24	12	12	12
Transformed choice rating $\overline{V'_G}$ - $\overline{V'_B}$	4.10 - 4.60	4.80 - 4.80	4.67 - 4.56	4.92 - 5.21	5.13 - 4.04 *	5.13 - 4.75

3.6 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.

Table 5 A summary of the hypothesis testing results (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)

H	Hypothesized Trend	Conditions					
		ISeq	TSeq	ITSeq	ISBS	TSBS	ITSBS
1a: Choice rating	$\overline{V_G} > \overline{V_B}$		**				
1b: Satisfaction rating	$\overline{S_G} > \overline{S_B}$	Op. *			Op. +		*
1c: Good-ness rating	$\overline{\Gamma_{GA}} > \overline{\Gamma_{BA}}$	Op. *		**		+	**
	$\overline{\Gamma_{GR}} > \overline{\Gamma_{BR}}$		**	*		**	*
2: Fixation time and count	$\overline{T_U} > \overline{T_H}$	+	*		**	**	**
	$\overline{Q_U} > \overline{Q_H}$		**		**	**	**
3: Transformed choice rating	$\overline{V'_G} > \overline{V'_B}$	Op. *				*	

Hypothesis 1A is accepted in the TSeq 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 ITSBS, 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 of "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 ITSeq condition, the preference decisions follow the trend predicted by the C&F model, but they do not reach statistical significance. ITSeq 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 ITSBS 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 ITSeq 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 Hypotheses 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 Hypothesis 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 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 consumer decisions when (a) they were 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 ITSeq 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 ITSeq condition, though 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 ITSeq 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 postpreference 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 and unfamiliar 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.

3.7 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

postpreference 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 headlights 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 factors 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. 3.6. 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 alternatives. 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.

Acknowledgement

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 Iowa State University for his advice on designing the experiment and analyzing the experimental results.

Nomenclature

A = Accepted alternative

B = Unique-bad pair

C = Car

E = Electric bicycle

G = Unique-good 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

V′= Transformed choice rating

Γ = Good-ness rating

Reference

[1] Tversky, A., 1977, "Features of Similarity," Psychol. Rev., **84**(4), pp. 327-352.

[2] Houston, D. A., and Sherman, S. J., 1995, "Cancellation and Focus: The Role of Shared and Unique Features in the Choice Process," J. Exp. Soc. Psychol., **31**(4), pp. 357-378.

[3] Houston, D. A., Sherman, S. J., and Baker, S. M., 1989, "The Influence of Unique Features and Direction of Comparison on Preferences," J. Exp. Soc. Psychol., **25**(2), pp. 121-141.

[4] Houston, D. A., Sherman, S. J., and Baker, S. M., 1991, "Feature Matching, Unique Features, and the Dynamics of the Choice Process: Predecision Conflict and Postdecision Satisfaction," J. Exp. Soc. Psychol., **27**(5), pp. 411-430.

[5] Sütterlin, B., Brunner, T. A., and Opwis, K., 2008, "Eye-Tracking the Cancellation and Focus Model for Preference Judgments," J. Exp. Soc. Psychol., **44**(3), pp. 904–911.

[6] Sanbonmatsu, D. M., Kardes, F. R., and Gibson, B. D., 1991, "The Role of Attribute Knowledge and Overall Evaluations in Comparative Judgment," Organ. Behav. Hum. Decis. Processes, **48**(1), pp. 131-146.

[7] Brunner, T. A., and Opwis, K., 2008, "The WReSt Heuristic: The Role of Recall as Well as Feature-Importance in and Beyond the Cancellation and Focus Model," Soc. Cognit., **26**(1), pp. 25-43.

[8] Brunner, T. A., and Wänke, M., 2006, "The Reduced and Enhanced Impact of Shared Features on Individual Brand Evaluations," J. Consum. Psychol., **16**(2), pp. 101-111.

[9] Sylcott, B., Cagan, J., and Tabibnia, G., 2011, "Understanding of Emotions and Reasoning During Consumer Tradeoff Between Function and Aesthetics in Product Design," Proceedings of the ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Washington, DC, Aug. 28-31, pp. 165-176.

[10] Dagher, A., and Petiot, J.-F., 2007, "Study of the Correlations Between User Preferences and Design Factors: Application to Cars Front-End Design," The International Conference on Engineering Design, ICED'07, Cité des Sciences et de l'Industrie, Paris, Aug. 28-31.

[11] Orsborn, S., Cagan, J., and Boatwright, P., 2009, "Quantifying Aesthetic Form Preference in a Utility Function," ASME J. of Mech. Des., **131**(6), p. 061001.

[12] MacDonald, E. F., Gonzalez, R., and Papalambros, P., 2009, "The Construction of Preferences for Crux and Sentinel Product Attributes," J. of Eng. Des., **20**(6), pp. 609-626.

[13] Reisen, N., Hoffrage, U., and Mast, F. W., 2008, "Identifying Decision Strategies in a Consumer Choice Situation," Judgment and Decis. Making, **3**(8), pp. 641-658.

[14] Dhar, R., and Sherman, S. J., 1996, "The Effect of Common and Unique Features in Consumer Choice," J. Consum. Res., **23**(3), pp. 193-203.

[15] Su, Y., Rao, L. L., Li, X. S., Wang, Y., and Li, S., 2012, "From Quality to Quantity: The Role of Common Features in Consumer Preference," J. of Econ. Psychol., **33**(6), pp. 1043-1058.

[16] Just, M. A., and Carpenter, P. A., 1976, "Eye Fixations and Cognitive Processes," Cognit. Psychol., **8**(4), pp. 441–480.

[17] Duchowski, A. T., 2007, *Eye Tracking Methodology: Theory and Practice*, Springer, London.

[18] Reid, T., MacDonald, E., and Du, P., 2013, "Impact of Product Design Representation on Customer Judgment," ASME J. Mech. Des., **135**(9), p. 091008.

[19] Du, P., and MacDonald, E., 2014, "Eye-Tracking Data Predict Importance of Product Features and Saliency of Size Change," ASME J. Mech. Des., **136**(8), p. 081005.

[20] She, J., 2013, "Designing Features That Influence Decisions About Sustainable Products," Ph.D. thesis, Iowa State University, Ames, IA.

[21] Rayner, K., 1998, "Eye Movements in Reading and Information Processing: 20 Years of Research," Psychol. Bull., **124**(3), pp. 372-422.

[22] Lohse, G. L., and Johnson, E. J., 1996, "A Comparison of Two Process Tracing Methods for Choice Tasks," Organ. Behav. and Hum. Decis. Processes, **68**(1), pp. 28-43.

[23] Glaholt, M. G., and Reingold, E. M., 2011, "Eye Movement Monitoring as a Process Tracing Methodology in Decision Making Research," J. Neurosci., Psychol. Econ., **4**(2), pp. 125-146.

[24] Shimojo, S., Simion, C., Shimojo, E., and Scheier, C., 2003, "Gaze Bias Both Reflects and Influences Preference," Nat. Neurosci., **6**(12), pp. 1317-1322.

[25] Russo, J. E., and Rosen, L. D., 1975, “Eye Fixation Analysis of Multialternative Choice,” Mem. Cognit., **3**(3), pp. 267-276.

[26] Russo, J. E., and Dosher, B. A., 1983, “Strategies for Multiattribute Binary Choice,” J. Exp. Psychol.: Learn. Mem. Cognit., **9**(4), pp. 676-696.

[27] “Chevrolet,” Accessed Dec. 4, 2014, http://www.chevrolet.com/flash.html

[28] “Global-Tradekey,” Accessed Apr. 21, 2013, http://www.global-tradekey.com/company/G02189/C64201.htm

[29] “iMotions,” Accessed Dec. 4, 2014, http://imotionsglobal.com/

[30] Pernice, K., and Nielsen, J., 2009, “How to Conduct Eyetracking Studies," Nielsen Norman Group, Fremont, CA.

CHAPTER 4

PRODUCT BODY SHAPES, NOT FEATURES, PROVIDE FAST AND FRUGAL CUES FOR ENVIRONMENTAL FRIENDLINESS

A manuscript accepted to *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*

Ping Du
Department of Mechanical Engineering, Iowa State University, Ames, IA 50011

Erin MacDonald
Assistant Professor, Department of Mechanical Engineering, Stanford University, Stanford, CA, 94305

Abstract

Mental associations between a product's visual design and its unobservable characteristics aid consumer judgments. It is hypothesized these associations, or cues, allow people to decrease the mental load required to make a decision. This paper investigates the rapid-building of mental associations between visual cues and unobservable attributes. It questions if it is more effective to cue holistically, through body-shape, or by individual features. Subjects participated in an association-building task and were then surveyed for retention of positive and negative cues for environmental friendliness ratings. Results demonstrate retention of body shapes cues but not feature cues. Additionally, eye-tracking data demonstrate that people redistribute their attention to a product after the association-building task, increasing the percentage of attention in the cued visual areas-of-interest. This supports the hypothesis that cues work to distribute mental load more efficiently; subjects' evaluations became more targeted when judging environmental friendliness.

4.1 Introduction

Consumers use a product's visual design as cues that they mentally associate with unobservable attributes. For example, previous research has shown consumers judge a car with a body shape that has more smooth lines as more environmentally friendly [1], and they judge paper towels with quilted lines as more absorbent [2]—refer to Section 4.2 for other related research. Designers can make use of cues to deliver desired messages about a product to consumers. By providing visual cues to consumers, designers can also help consumers make the right inferences about products, and decrease the mental load required to make decisions [3] once consumers learn the cues. This paper uses the word "cue" in two ways: (1) as a noun, and (2) as a verb, meaning "providing a cue."

When considering a product's visual form, designers can: use their experience with form design to speak through cues that are instinctively known to them; test for existing cues with experiments as in [1, 2]; or choose to build new cues (perhaps for new features) through advertising, marketing, visibility, and word-of-mouth. Past research indicates that product visuals do not always behave as one might predict, for example shared visual features between products are not ignored but rather magnified in judgements [4]. Thus, designers should approach the design of new visual cues carefully. Here we investigate the nuances of building visual cues into products, testing: (1) if body or feature cues are more effective, in the case of building cues quickly to assess environmental friendliness; (2) if both positive and negative cues can be built, and (3) what happens when cues contradict each other.

According to the theory of fast-and-frugal decision-making, proposed by Gigerenzer [3], people use mental shortcuts to both ease the cognitive load of decisions and make them faster. We postulate that cues are one such shortcut to reduce mental burden in decisions, a proposition

that we investigate using eye-tracking data. The paper borrows its title from this theory: "fast" because the cues are built and tested with a 20 minute experiment, and "frugal" because we demonstrate that these quickly-built cues change decision-making behavior to be more efficient.

Designers have spent significant effort on minimizing products' environmental impacts. Unfortunately it is common that these efforts remain hidden inside the product and unlinked to the visual design, instead relying on marketing messages to communicate their environmental friendliness. This may introduce the issue of “trust” in these claims [5, 6] and therefore have a negative effect on consumer judgments. Visual cues can not only help with communicating a product’s environmental friendliness, but may also lend cognitive support to the marketing claims and help with the issue of trust, though this is not investigated here. This research uses an association-building task to let subjects learn visual cues and their associations with environmental friendliness rating. Half of the tested visual cues are tied to a positive judgment of environmental friendliness (termed as *positive cues*), the other half with a negative judgment (termed as *negative cues*).

Researchers have found that both the body and the individual feature of a product can affect consumer judgments, see Figure 1 and Section 4.2, which demonstrates the possibility to associate *body-cues* and/or *feature-cues* with certain product attributes. This paper defines the *body* of a product as the main part of the product that encloses or holds product features. Body shapes usually affect products’ silhouettes the most comparing with other parts of the products. The paper defines a product *feature* as a visible and distinct (or individual) product attribute. Features determine a product’s sub-structures. This paper considers the body-cue and the feature-cue as two types of product visual cues which are associated with unobservable attributes, and studies both of them in terms of building mental associations.

Situations considered in the study include: (1) both the body-cue and the feature-cue appear in a product; (2) only the body-cue appears in the product; and (3) only the feature-cue appears in the product and (a) only positive cue appears; (b) only a negative cue appears; and (c) both a positive and negative cue appear. Considering these situations allows a comparison between the effectiveness of body-cues and feature-cues. The study includes two case products: electric bicycles and electric space heaters. It uses eye-tracking technology to study how consumers use visual cues when making product judgments. The paper proceeds as follows: Section 4.3 lists research propositions and hypotheses; Section 4.4 provides details on the experiment; Section 4.5 presents analysis approaches and experiment results; Section 4.6 discusses the results; and Section 4.7 provides conclusions.

4.2 Background

4.2.1 Effects of Cues on Consumer Judgments. The review here draws from literature on various forms of non-visual and visual cues to demonstrate that cues affect consumer judgments. MacDonald et al. [2] identified through manipulated discrete choice surveys that quilting of paper towels served as a cue for subjects to judge the paper towels' absorbency. Berkowitz [7] studied consumer preferences for buying ears of corn with untrimmed ends versus those with squared-off ends. He found that experimental subjects used the untrimmed end of the corn as a visual cue to differ the two types of corns in terms of their less-obvious attributes like freshness, taste/flavor, and overall quality, resulting in a preference for corns with untrimmed ends. Chandler et al. [8] found that a heavier copy of a book received a higher importance rating than a lighter copy of the same book when subjects had read the book before. According to Wänke et al. [9], consumers use brand name as a cue to judge the relevance of brand extensions

to the original brand. They created a scenario where a sports-car manufacturer introduced a compact car as a brand extension. The name of the compact car reflected either continuation or discontinuation of the names of the brand's sports cars in different experimental conditions. They found that subjects rated the compact car as less sport-car-typical when its name reflected discontinuation than when it reflected continuation. Dawar and Parker [10] found that consumers used brand names, price, visual design, and retailer reputation as cues that signal product quality. Kirmani [11] identified that perceived advertising costs served as a cue for judging brand quality.

Trivial attributes refer to product attributes that have a "trivial and/or subjective relationship to perceived quality as well as objectively irrelevant attributes" [12]. Trivial attributes can cues the quality or value of a product [13]. Carpenter et al. [14] found that having a trivial attribute (e.g. "alpine class down fill" for down jackets) instead of a regular attribute (e.g. "regular down filling") made a choice option receive a significantly higher preference rating in a high-mental-load decision. This effect existed regardless of whether or not the subjects knew the meaning of the given trivial attribute. As a replication, Brown and Carpenter [12] confirmed that having a trivial attribute positively affected the selection rate of an alternative in a three-alternative set.

4.2.2 Effects of Product Body and Product Feature on Consumer Judgments. The body and the individual features of a product are two major constitutions of a product's visual design. They both play roles in influencing consumer judgments of the product, as investigated and identified by previous literature. As summarized in Fig. 1, much of the literature identified here focuses on either the body or the feature, and only four articles consider both the body and the feature.

Authors and References	Focuses	Consumer Judgments	Sample Stimuli
Reid et al. [1]	Body	Environmental friendliness, Preference, Inspired by nature	
Tseng et al. [15]	Body	Sportiness, Ruggedness, Aerodynamics, and Fuel efficiency	
Lai et al. [16]	Body	Feeling of the car on three scales: (1) Young ←→ Mature (2) Field ←→ City (3) Personal ←→ Family	
Dagher and Petiot [17]	Body, Feature	Impressions in terms of aggressive, elegant, intrepid, happy, confident, severe, and laughing separately; Preference	
Orbay et al. [18]	Body, Feature	Brand identification, fast, muscular, elegant, sophisticated, utility, and compact	
Yumer et al. [19]	Body, Feature	Cars (Luxurious, sporty, compact, muscular, and modern) Shoes (Fashionable, comfy, feminine, active, and durable) Chairs (Comfortable, ergonomic, elegant, antique, and sturdy) Airplanes (Civilian, sleek, stealth, fighter, and fast)	
Reid et al. [20]	Body	Preference, Stylishness, Objective evaluations (width, length, and height), Inferences (heat retention, recyclability, and fuel efficiency)	
MacDonald et al. [21]	Body	Flavor of the wine that the bottle contains	
She and MacDonald [22]	Feature	Trigger thoughts of sustainability in multiple judgments	
Kelly et al. [23]	Body	Preference	
Swamy et al. [24]	Feature	Preference	
Orsborn et al. [25]	Body, Feature	Preference	

Fig. 1 Previous literature has identified effects of the body and the feature on consumer judgments

A number of papers focus on consumer vehicles. Reid et al. [1] manipulated car bodies to identify shape-defining points that affect perception of a car's environmental friendliness. Subjects perceived boxy shapes as less environmentally friendly than smoother ones. Tseng et al. [15] varied car body shapes to test consumer judgments of aerodynamics, sportiness, fuel efficiency, and ruggedness, finding some interesting correlations. For example, windshield angle and rear window angle had significantly positive correlations with the judgments of aerodynamics and sportiness. Lai et al. [16] found that different car bodies evoke different

sentiments, as measured by the scales in Fig. 1. Dagher and Petiot [17] classified 13 existing cars into groups judged to have different semantic attributes such as aggressive, elegant, and intrepid. The stereotype they identified for each group had a distinct body shape and feature designs. Orbay et al. [18] tested consumer judgments of cars' abstraction models and full models. They observed that subjects associated body shapes with certain unobservable attributes; and detailed feature designs, which were related to brand recognition, also affected consumer judgments. Yumer et al. [19] proposed a method that enabled designers to continuously deform products' visual designs using sematic attributes. They demonstrated the method on cars, shoes, airplanes, and chairs. A mapping between the semantic attributes and the products' visual designs, learned from a survey, served as a foundation of their method.

Reid et al. [20] studied the consistency of consumers' inferences across different product representations. They observed that, regardless of the representation mode, their subjects tended to judge the "short and stout" kettle as having better heat-retention ability while judging the "tall and narrow" kettle as more recyclable. A study on wine bottles by MacDonald et al. [21] showed that the body shape of the bottle influenced the perceived flavor of the wine. She and MacDonald [22] found that purposefully-designed features of toasters triggered consumers to include sustainability in their purchase criteria. In addition, refs. [1, 17, 20, 23-25] demonstrate the effects of body and feature on consumer preferences in various ways. The literature here shows the potential of body and feature cues, but no comparisons between the effectiveness of the two has yet been offered.

4.2.3 Eye-Tracking. Eye-tracking technology allows researchers to obtain eye movement data that detail what and how people look at stimuli, which can indicate how people cognitively process the stimuli [26]. The data provided a different dimension of information

when used with other forms of data (for example, survey data). References [4, 20, 27, 28] summarize uses of eye movement data in product design research. We will use *fixations*, or "eye movements that stabilize the retina over a stationary object of interest" [29], to study *areas of interest* (AOIs), which refer to "areas of a given stimulus related to the research hypothesis" [27]. We will investigate percentage-fixation time (ratio of an AOI's fixation time to the total time a person looks at the stimulus)—referred to as *gaze data*.

4.3 Research Propositions and Associated Hypotheses

Note that for propositions 1 and 2 and their corresponding hypotheses presented below, we will test them after a mental association training exercise, detailed in Section 4.4.2.

PROPOSITION 1: *It is possible to build mental associations between a product's existing visual form and the product's environmental friendliness within a brief computerized experiment*. In other words, we propose that it is possible for consumers to learn a product's visual cues for environmental friendliness within a brief computerized experiment.

Studies reviewed in Section 4.2 show that the visual design of a product can affect consumer judgments of product attributes like environmental friendliness, and that consumers do use cues for their product judgments. This study expects that components of the visual form (cues), as the examples shown in Fig. 2, can be associated with a product's environmental friendliness through an association-building task and then affect consumer judgments of environmental friendliness in the desired direction. If the associations are built and consumers learn the cues, positive cues should have a positive effect on consumer judgments and negative cues should have a negative effect on consumer judgments.

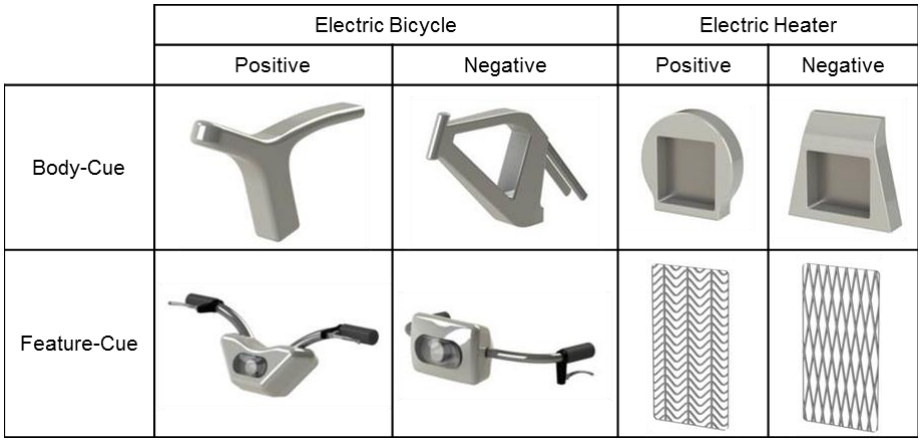


Fig. 2 Selected visual cues for the electric bicycle and the electric heater

PROPOSITION 2: *Body- and feature-cues affect judgments differently.* Body-cues are holistic and feature-cues are a detail or smaller portion of the product. Thus, body- and feature-cues may have different effects on consumer judgments of environmental friendliness. The following hypotheses test Proposition 1 and 2 at the same time by examining results for products (1) with both body- and feature- cues, (2) with only body-cues, and (3) with only feature-cues separately.

Hypothesis 1a: Subjects rate a product with only positive cues (P) as having a higher environmental friendliness rating (E) than that with no cues (∅).

$$\overline{E_P} > \overline{E_\varnothing} \tag{1}$$

And *Hypothesis 1b: Subjects rate a product with only negative cues (N) as having a lower environmental friendliness rating than that with no cues.*

$$\overline{E_N} < \overline{E_\varnothing} \tag{2}$$

PROPOSITION 3: *Building mental associations improves judgment efficiency*. We expect that, after the association-building task (A), subjects will rely on cues to judge environmental friendliness, resulting in more targeted evaluations during the testing task (T).

Thus, how subjects view cued AOIs (body and feature cue areas) versus uncued AOIs (rest of the areas) should change, as extracted from gaze data. We expect that cued vs. uncued AOI will have larger differences in viewing times after the associations are built.

Hypothesis 2: Difference in subjects' percentage-fixation times (Φ) between the cued AOIs (C) and the uncued AOIs (U) during the testing task (T) is greater than during the preceding association-building task (A).

$$\overline{\Phi_{TC} - \Phi_{TU}} > \overline{\Phi_{AC} - \Phi_{AU}} \tag{3}$$

4.4 Method

This study tested the research hypotheses through a computerized experiment, as illustrated in Table 1. The experiment had five parts. Hypotheses testing used the results from part I and II of the experiment. These two parts had the same tasks: a mental association-building task, a preference task, and a testing task. The preference task was used for another investigation. Its results are inconclusive and are not discussed in this paper. The experiment took about 20 minutes to complete. iMotions software [30] recorded and managed subjects' eye movements tracked by a Tobii T120 eye-tracker during the experiment. The software also recorded survey question results. Section 4.4.1 describes product stimuli used in the experiment. Section 4.4.2 and 4.4.3 detail the mental association-building task and the testing task, respectively. Section 4.4.4 introduces the experiment design. Section 4.4.5 introduces the subjects and data preparations for further statistical analysis.

4.4.1 Stimuli. Electric bicycles and electric heaters serve as case products. As product familiarity may affect mental association-building, considering these products, with different familiarity to the U.S. market, is warranted. Stimuli were generated in Solidworks. Preparing the

Table 1 Overview of the experiment design. "H" stands for "Hypothesis"

Part	Product	Task	Theme	Image Types (Explained in Section 4.4.1)	*H*	Gaze data	Survey data discussed here?
I	Electric Bicycle /Electric Heater	1	Association-building task	$\{Body^{P/N}, Feature^{P/N}\}$; $\{Body^{P/N}, Feature^{\varnothing}\}$; $\{Body^{\varnothing}, Feature^{P/N}\}$; $\{Body^{\varnothing}, Feature^{\varnothing}\}$	2	%-Fixation time	No
		2	Preference task	$\{Body^{\varnothing}, Feature^{\varnothing}\}$	—	—	No
		3	Testing task	$\{Body^{P/N}, Feature^{P/N}\}$; $\{Body^{P/N}, Feature^{N/P}\}$; $\{Body^{P/N}, Feature^{\varnothing}\}$; $\{Body^{\varnothing}, Feature^{P/N}\}$; $\{Body^{\varnothing}, Feature^{\varnothing}\}$	1a 1b 2	%-Fixation time	Yes
II	Electric Heater /Electric Bicycle	1	Association-building task	$\{Body^{P/N}, Feature^{P/N}\}$; $\{Body^{P/N}, Feature^{\varnothing}\}$; $\{Body^{\varnothing}, Feature^{P/N}\}$; $\{Body^{\varnothing}, Feature^{\varnothing}\}$	2	%-Fixation time	No
		2	Preference task	$\{Body^{\varnothing}, Feature^{\varnothing}\}$	—	—	No
		3	Testing task	$\{Body^{P/N}, Feature^{P/N}\}$; $\{Body^{P/N}, Feature^{N/P}\}$; $\{Body^{P/N}, Feature^{\varnothing}\}$; $\{Body^{\varnothing}, Feature^{P/N}\}$; $\{Body^{\varnothing}, Feature^{\varnothing}\}$	1a 1b 2	%-Fixation time	Yes
III	—	—	Identify maximizers and satisficers	—	—	—	No
IV	Electric Bicycle, Electric Heater	—	Comprehension test and post-task survey questions	$\{Body^{P/N}, Feature^{P/N}\}$; $\{Body^{P/N}, Feature^{\varnothing}\}$; $\{Body^{\varnothing}, Feature^{P/N}\}$	—	—	Yes
V	Electric Bicycle, Electric Heater	—	Collect demographic information	—	—	—	Yes

stimuli included two steps: determining the design of the visual cues and generating different images of the case products, each detailed below.

Determining visual cue design. For feature-cues, we sought important product features that consumers cannot help but look at when evaluating products. Based on our expertise in eye-tracking product design research, we determined that the handlebar of the electric bicycle and the grille of the electric heater would serve this purpose well, thus they were selected as the feature-cues. For body-cues, we focused on the largest portion of the physical form: the heater case and

bicycle frame. Based on web images and our design thoughts, we generated cue-candidates: eight different bicycle frames, nine handlebars, nine heater cases, and eight grilles. A pilot study tested the perceived environmental friendliness of these cue-candidates, which were presented as merged into the same base product image and printed on individual cards. We sought designs that were initially neutral in rating, perceived neither positively nor negatively. Twelve pilot-study subjects evaluated the cue-candidates. Subjects sorted and ranked the cue-candidates in each group from most environmentally friendly to least environmentally friendly. Two cues were selected from each group, ranking in the middle (or neutrally). The cues with slightly higher rankings (not significant) were assigned as the positive cues and those with slightly lower rankings as negative. Cue-candidates that had extremely low or high rankings were thrown-out. The other neutral-ranking cue-candidates were preserved as no-cue variants for the experiment. Figure 2 presents the selected cues for the two products.

Preparing the Stimuli. To prepare the experimental stimuli, design variants were generated for six other features (termed as *dummy features*) including the bicycle's rearview mirror, seat, and wheel; and the heater's handle, base, and control knob. Each dummy feature had five design variants. Each cued AOI (frame, handlebar, case, and grille) had four no-cue design variants and two cue variants. The no-cue variants all came from the pool of cue-candidates introduced above, except for one variant of the grille, which was generated afterwards. The environmental friendliness rankings of the no-cue variants did not statistically differ from the cues in the pilot study. For the variant of the grille that was generated afterwards, a second round of the pilot study, which had 12 new subjects and followed the same procedures as the first round, confirmed its neutral environmental friendliness ranking.

Five types of product images were prepared for each product. Figure 3 provides sample images for each type. $\{Body^{P/N}, Feature^{P/N}\}$ means the two cues in the image have the same sign (e.g., an image with both the positive body-cue and the positive feature-cue). $\{Body^{P/N}, Feature^{\varnothing}\}$ means the image has a positive or negative body-cue and no feature-cue. $\{Body^{\varnothing}, Feature^{P/N}\}$ means the image has no body-cue and a positive or negative feature-cue. $\{Body^{P/N}, Feature^{N/P}\}$ means the two cues in the image have different signs (e.g., an image with the positive body-cue and the negative feature-cue). $\{Body^{\varnothing}, Feature^{\varnothing}\}$ means the image has no cues, only neutral variants. Configurations of the design variants involved in the product images were randomly determined. This paper will refer to the images that have only the positive cue(s) as *positive images*, and will refer to the images that have only the negative cue(s) as *negative images*.

Image Type	Sample Image		Image Type	Sample Image	
	Electric Bicycle	Electric Heater		Electric Bicycle	Electric Heater
$\{Body^{P}, Feature^{P}\}$			$\{Body^{P}, Feature^{\varnothing}\}$		
$\{Body^{P}, Feature^{N}\}$			$\{Body^{\varnothing}, Feature^{P}\}$		

Fig. 3 Sample product images in different types

4.4.2 Mental Association-Building Task. The association-building task in this experiment referenced the feedback training approach from psychology, an approach commonly used in category learning [31]. We conducted a pilot study of the experiment to test effectiveness

of this approach and another possible approach that attempted to build associations through a set of questions on product preferences (see Section 4.6 for more information on this unselected approach). Results of the pilot study demonstrated that the feedback training approach was more effective at quickly building associations, and so it was selected.

The feedback training association-building task presented 20 images of a product (bicycle, heater) to the subjects in random order. Table 2 provides detailed information on the composition of the 20 images. The subjects saw four screens related to each image: a nametag screen; a product image screen; a survey question screen; and a feedback screen, see Fig. 4. Subjects were asked to evaluate the image and then rate its environmental friendliness on a five-point likert scale that ranged from "not environmentally friendly at all" to "very environmentally friendly". After the rating, the same image was presented again along with feedback information on the predetermined corresponding environmental friendliness rating (e.g., "Environmental friendliness of this electric bicycle: 5 out of 5"), as specified in Table 2.

Table 2 The association-building task used product images with predetermined environmental friendliness ratings

	$\{Body^P, Feature^P\}$	$\{Body^N, Feature^N\}$	$\{Body^P, Feature^\varnothing\}$	$\{Body^N, Feature^\varnothing\}$	$\{Body^\varnothing, Feature^P\}$	$\{Body^\varnothing, Feature^N\}$	$\{Body^\varnothing, Feature^\varnothing\}$
Number of Images	3	3	3	3	3	3	2
Predetermined Environmental Friendliness Rating	5	1	5	1	5	1	3

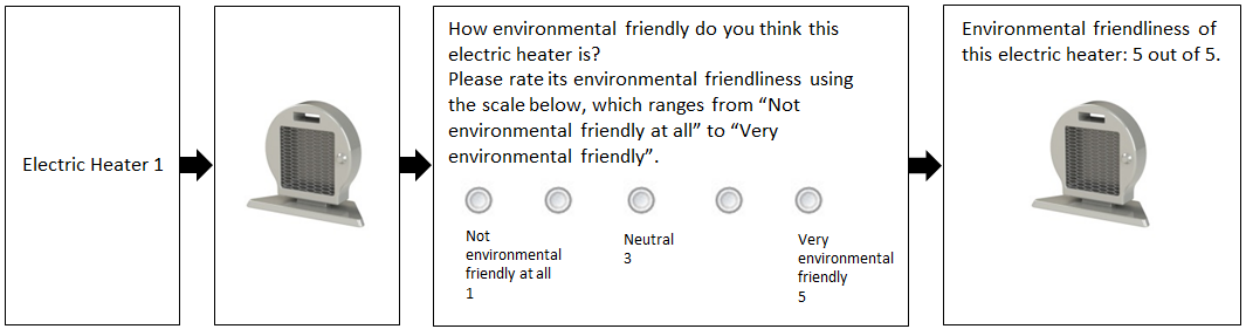


Fig. 4 Demonstration of the association-building task

4.4.3 Testing Task. The testing task provided 10 images of a case product, including all possible combinations of positive, neutral, and negative cues, to the subjects in random order as detailed in Table 3. For each product, we prepared two sets of 10 images for this task and each subject saw only one randomly-determined set for each product. The testing task had the same procedure as the association-building task, as shown in Fig. 4, except that the testing task did not provide any feedback information. The testing task asked the subjects to evaluate each given product image and rate its environmental friendliness using the five-point likert scale in Section 4.4.2.

Table 3 Product image composition in the testing task

	$\{Body^P, Feature^P\}$	$\{Body^N, Feature^N\}$	$\{Body^P, Feature^\varnothing\}$	$\{Body^N, Feature^\varnothing\}$	$\{Body^\varnothing, Feature^P\}$	$\{Body^\varnothing, Feature^N\}$	$\{Body^P, Feature^N\}$	$\{Body^N, Feature^P\}$	$\{Body^\varnothing, Feature^\varnothing\}$
Number of Images	1	1	1	1	1	1	1	1	2

4.4.4 Experiment Design. As shown in Table 1, the experiment had five parts. The experiment had no time constraint. The subjects advanced the stimuli manually and at their own pace. Part I asked questions about one randomly determined case product. Half of the subjects saw the electric bicycle in part I and the other half saw the electric heater in part I (the remaining product shown in part II). Part I had three tasks. In each task, the subjects completed a practice question about a practice product before seeing the questions about the case product. Task 1 of part I was the association-building task, as described in Section 4.4.2. Task 2 provided three pairs of $\{Body^\varnothing, Feature^\varnothing\}$ images of the case product in random orders to the subjects and asked them to indicate their preferences for each pair. Task 2 served as a break in between the association-building task and task 3, which was the testing task as detailed in Section 4.4.3.

Part II of the experiment asked questions about the remaining case product and had the same procedure as part I, except that it had no practice questions. Part III measured the subjects'

maximization tendency for another investigation that was inconclusive, and is not discussed in this paper. It presented six statements, referenced from a study of Nenkov et al. [32], to the subjects separately and asked them to rate their agreements with each statement on a seven-point likert scale that ranged from “completely disagree” to “completely agree”.

Part IV asked both comprehension test questions and post-task survey questions on both products. The comprehension test questions aimed to check if the subjects remembered the product images that appeared in the experiment as well as the feedback information given in the association-building task. Figure 5 summarizes and illustrates those questions using the electric bicycle as an example. An image-comprehension question, as shown in Fig. 5, showed six design variants to the subject, out of which only four variants appeared in the prior parts of the experiment. The post-task survey questions asked the subjects to (1) rate environmental friendliness for each frame, handlebar, case, and grille that had appeared in the experiment using the five-point likert scale introduced before; (2) rate their preference for each frame, handlebar, case, and grille that had appeared in the experiment using five-point likert scales that ranged from “don’t prefer at all” to “strongly prefer”; (3) specify what they think environmental friendliness means; and (4) explain how they decided the environmental friendliness of the given product images during the experiment. Part V, the last part of the experiment, collected some demographic information about the subjects.

4.4.5 Subjects and Data Preparation. The experiment had 80 subjects (49 females, 30 males, and 1 subject who preferred not to indicate gender), including students, faculty, and staff recruited from Iowa State University. Each subject was compensated with $5 cash. A subject under 18 years old participated in the study unexpectedly. Following the regulations of the Institutional Review Board for human subject studies, the following analysis has excluded the

Question ID	Objectives	Test Objects	Sample Questions
Image-comprehension Question	Test if the subjects remembered the handlebars and the frames that appeared in the experiment	Handlebar, Frame	Please check all handlebar designs for the electric bicycle that you saw in this survey. 
Information-comprehension Question	Test if the subjects remembered the feedback information from the association-building task	$\{Body^P, Feature^P\}$ vs. $\{Body^N, Feature^N\}$; $\{Body^P, Feature^\varnothing\}$ vs. $\{Body^N, Feature^\varnothing\}$; $\{Body^\varnothing, Feature^P\}$ vs. $\{Body^\varnothing, Feature^N\}$	In this survey, which of these electric bicycles has the higher environmental friendliness rating? 

Fig. 5 Summary and illustration of the comprehension test questions on the electric bicycle

data of that subject. All subjects passed an on-line screening survey about their eyes before coming to the experiment, as suggested by Pernice and Nielsen [33], which checks to make sure eye movements will not be difficult for the eye tracker to track due to such factors as wearing thick glasses. The subjects went through a calibration process provided by the iMotions software at the beginning of the experiment.

In order to analyze gaze data from the iMotions software, we manually defined AOIs for each product stimulus. Figure 6 shows example AOIs generated for the product images. Survey data were also exported from the software. Six subjects failed to pass the preliminary calibration and the eye tracker had difficulty tracking one subject's eye movements during the experiment. Therefore, gaze data analysis did not include these seven subjects. In addition, the software had difficulty identifying fixations for a few stimuli. The gaze data analysis did not include those stimuli and Fig. 8 lists the quantity of the excluded stimuli.

4.5 Analysis and Results

The analysis was performed separately for the two case products. Section 4.5.1and 4.5.2 detail the analysis and the results of Proposition 1, 2, and 3. As specified in Table 1, the analysis of Proposition 1 and 2 used the survey data from the testing tasks. The analysis of Proposition 3

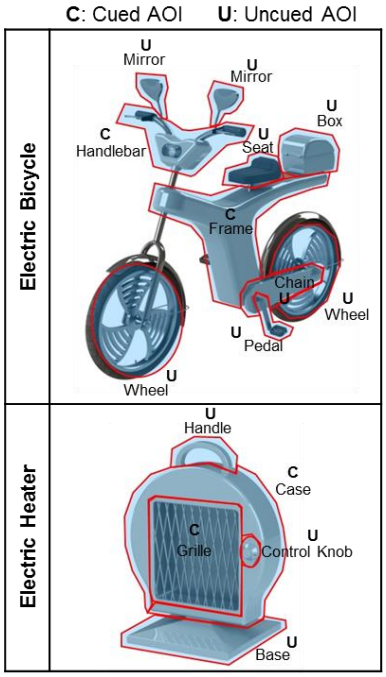


Fig. 6 Example AOIs generated for the experiment

used the gaze data from both the association-building tasks and the testing tasks. Section 4.5.3 presents the analysis and the results of Part IV of the experiment.

4.5.1 Analysis and Results of Proposition 1 and 2. Hypothesis 1a: An average environmental friendliness rating across the subjects ($\overline{E_P}$) was calculated separately for each positive image ($\{Body^P, Feature^P\}$, $\{Body^P, Feature^{\varnothing}\}$, and $\{Body^{\varnothing}, Feature^P\}$) following Eq. (4), where i stands for a subject and I stands for the number of subjects. As each subject saw two neutral (no-cue) images in the testing task, subject-level average environmental friendliness ratings ($\overline{E_{\varnothing i}}$) for $\{Body^{\varnothing}, Feature^{\varnothing}\}$ were calculated and then an average rating across the subjects ($\overline{E_{\varnothing}}$) was calculated following Eq. (5). Pairwise t-tests examined if the positive images had higher environmental friendliness rating than the neutral images. Figure 7 presents the results.

$$\overline{E_P} = \sum_{i=1}^{I} E_{Pi} / I \tag{4}$$

$$\overline{E_\varnothing} = \sum_{i=1}^{I} \overline{E_{\varnothing i}} / I \tag{5}$$

Hypothesis 1b: The same analysis was performed on the environmental friendliness ratings for the negative images (E_{Ni}). An average rating across the subjects ($\overline{E_N}$) was calculated for each type of negative images ({*Body*^{N}, *Feature*^{N}}, {*Body*^{N}, *Feature*$^\varnothing$}, and {*Body*$^\varnothing$, *Feature*^{N}}) separately following Eq. (6). Pairwise t-tests examined if the product with only the negative cue(s) had lower environmental friendliness rating than the product with no cues. Figure 7 presents the results.

$$\overline{E_N} = \sum_{i=1}^{I} E_{Ni} / I \tag{6}$$

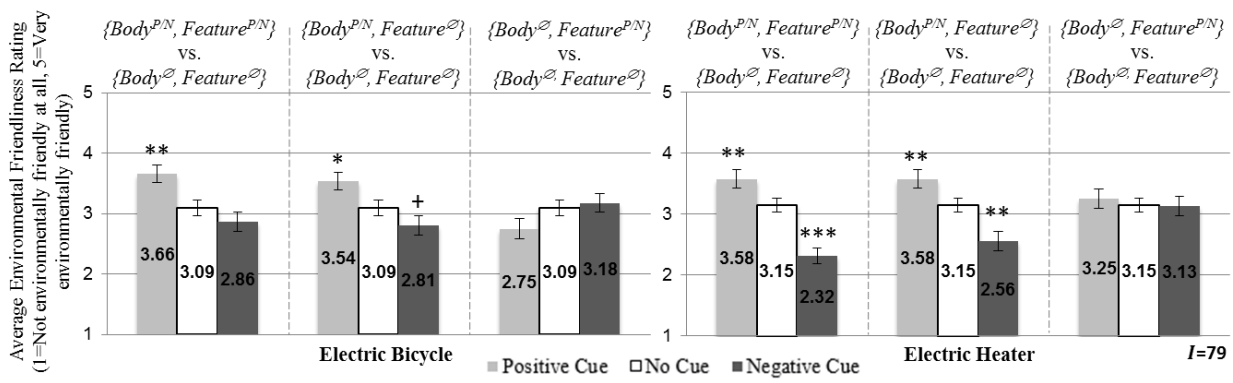


Fig. 7 After the association-building task, body and body + feature cues affect environmental friendliness ratings in the desired direction. Feature cues alone have no effect. Statistical significances obtained from pairwise comparisons with {*Body*$^\varnothing$, *Feature*$^\varnothing$} are specified as + p<0.1, * p<0.05, ** p<0.01, and *** p<0.0001. Error bars indicate +1 standard errors

4.5.2 Analysis and Results of Proposition 3. Hypothesis 2: The AOIs generated for a product image (j) were categorized into two groups: the cued- and the uncued-AOI-groups, as shown in Fig. 6. For the electric bicycle, the cued-AOI-group included the frame and the

handlebar; for the electric heater, the cued-AOI-group included the case and the grille. For both products, the uncued-AOI-group contained the rest of the product features.

For all the product images provided in the testing task of the experiment, average percentage-fixation time that each subject spent on the cued-AOI-group ($\overline{\Phi_{TCi}}$) and the uncued-AOI-group ($\overline{\Phi_{TUi}}$) was calculated separately following Eq. (7), where J represents the number of product images. Subject-level percentage-fixation-time differences between the cued- and the uncued-AOI-groups were first calculated by subject and then averaged across all the subjects ($\overline{\Phi_{TC} - \Phi_{TU}}$), as indicated in Eq. (8).

$$\overline{\Phi_{TCi}} = \sum_{j=1}^{J} \Phi_{TCij} /J, \overline{\Phi_{TUi}} = \sum_{j=1}^{J} \Phi_{TUij} /J$$
$$\overline{\Phi_{ACi}} = \sum_{j=1}^{J} \Phi_{ACij} /J, \overline{\Phi_{AUi}} = \sum_{j=1}^{J} \Phi_{AUij} /J \tag{7}$$

$$\overline{\Phi_{TC} - \Phi_{TU}} = \sum_{i=1}^{I} (\overline{\Phi_{TCi}} - \overline{\Phi_{TUi}})/I,$$
$$\overline{\Phi_{AC} - \Phi_{AU}} = \sum_{i=1}^{I} (\overline{\Phi_{ACi}} - \overline{\Phi_{AUi}})/I \tag{8}$$

The same analysis was performed on the product images in the association-building task. Subject-level average percentage-fixation times spent on the cued-AOI-group ($\overline{\Phi_{ACi}}$) and the uncued-AOI-group ($\overline{\Phi_{AUi}}$) were calculated separately following Eq. (7). Then, an average percentage-fixation time difference between the cued- and the uncued-AOI-groups ($\overline{\Phi_{AC} - \Phi_{AU}}$) was calculated following Eq. (8).

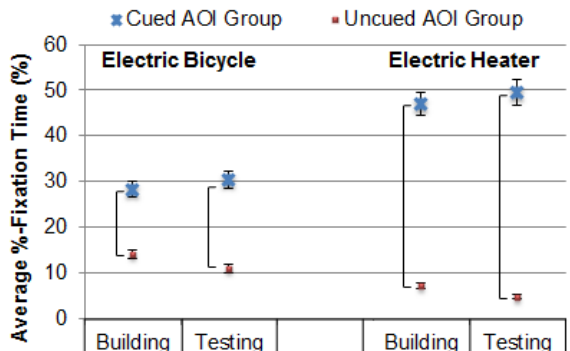
A pairwise t-test examined if the testing task had a significantly greater average percentage-fixation time difference between the cued- and the uncued-AOI-groups than the preceding association-building task. Two-way within-subject ANOVA tested if the AOI group and the task of the experiment (the testing task and the association-building task) had a

significant interaction effect on the percentage-fixation time. Table 4 provides the mean table associated with the ANOVA test. Figure 8 provides the results.

Table 4 Mean table of percentage-fixation time for the two-way within-subject ANOVA

Electric Bicycle		Task		Marginal Mean
		Building	Testing	
AOI Group	Cued AOI Group	28.20	30.38	29.29
	Uncued AOI Group	14.08	10.85	12.48
Marginal Mean		21.14	20.62	20.88 (Grand Mean)
Electric Heater		**Task**		**Marginal Mean**
		Building	Testing	
AOI Group	Cued AOI Group	47.03	49.63	48.33
	Uncued AOI Group	7.22	4.69	5.96
Marginal Mean		27.12	27.16	27.14 (Grand Mean)

4.5.3 Analysis and Results of Part IV of the Experiment. To examine subjects' memories of the product images, their answers to the image-comprehension questions were analyzed for the body and the feature separately. In an image-comprehension question, if a subject identified at least three out of the four design variants that appeared in the experiment, we considered the subject's answer as correct. Table 5 provides the percentage of subjects that answered each image-comprehension question correctly. To examine subjects' memories of the feedback information from the association-building task, their answers to the information-comprehension questions were analyzed. For each tested comparison, Table 5 provides the percentage of subjects that answered the question correctly.



	Building	Testing		Building	Testing
Stimuli (Excluded)	1440 (105)	720 (53)		1440 (84)	720 (42)
Number of subjects	72	71		72	72
Cued AOI Group	28.20	30.38		47.03	49.63
Uncued AOI Group	14.08	10.85		7.22	4.69
Difference (Cued – Uncued)	14.12 vs. 19.53 ***			39.81 vs. 44.94 ***	
Interaction Effect	F Ratio = 17.55 ***			F Ratio = 20.52 ***	

Fig. 8 After the association-building task, the subjects evaluate the products more efficiently by decreasing attention on the uncued AOIs and increasing attention on the cued AOIs (*** $p<0.0001$. Error bars indicate +1 standard errors)

Table 5 Majority of subjects remembered the design variants and the feedback information on environmental friendliness

I=79	Image-comprehension Question	
	Test Objects	Percentage of correct subjects
Electric Bicycle	Frame	84%
	Handlebar	87%
Electric Heater	Case	97%
	Grille	95%

I=79	Information-comprehension Question	
	Tested Comparisons	Percentage of correct subjects
Electric Bicycle	$\{Body^P, Feature^P\}$ vs. $\{Body^N, Feature^N\}$	76%
	$\{Body^P, Feature^\varnothing\}$ vs. $\{Body^N, Feature^\varnothing\}$	76%
	$\{Body^\varnothing, Feature^P\}$ vs. $\{Body^\varnothing, Feature^N\}$	53%
Electric Heater	$\{Body^P, Feature^P\}$ vs. $\{Body^N, Feature^N\}$	73%
	$\{Body^P, Feature^\varnothing\}$ vs. $\{Body^N, Feature^\varnothing\}$	65%
	$\{Body^\varnothing, Feature^P\}$ vs. $\{Body^\varnothing, Feature^N\}$	59%

The environmental friendliness ratings that subjects gave to each design variant in the post-task survey questions were averaged across the subjects separately. Figure 9 provides the average ratings. Each design variant's average preference rating was calculated similarly and was provided in Fig. 9. We also tested Pearson correlations between the environmental friendliness rating and the preference rating. Figure 9 presents the results.

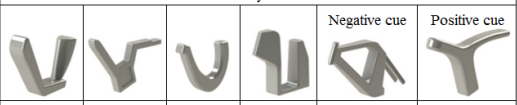
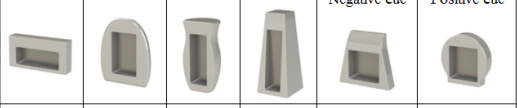
	Electric Bicycle - Frame						Electric Bicycle - Handlebar					
					Negative cue	Positive cue					Negative cue	Positive cue
												
Average Environmental Friendliness Rating (Standard Error)	3.12 (0.14)	3.18 (0.13)	3.83 (0.15)	2.65 (0.16)	2.53 (0.14)	4.23 (0.11)	3.08 (0.16)	2.90 (0.14)	3.29 (0.13)	3.09 (0.18)	3.19 (0.15)	3.55 (0.13)
Average Preference Rating (Standard Error)	2.94 (0.15)	3.05 (0.13)	3.32 (0.17)	2.05 (0.15)	2.68 (0.15)	3.80 (0.15)	3.01 (0.14)	2.67 (0.15)	3.09 (0.15)	2.67 (0.16)	2.95 (0.15)	3.69 (0.13)
Correlation	0.91*						0.93**					
	Electric Heater - Case						Electric Heater - Grille					
					Negative cue	Positive cue					Negative cue	Positive cue
												
Average Environmental Friendliness Rating (Standard Error)	2.65 (0.15)	3.42 (0.13)	3.19 (0.16)	3.22 (0.15)	2.70 (0.15)	3.86 (0.14)	3.52 (0.14)	3.14 (0.16)	3.30 (0.16)	3.42 (0.15)	2.87 (0.15)	3.51 (0.14)
Average Preference Rating (Standard Error)	2.32 (0.13)	3.19 (0.14)	3.33 (0.16)	3.27 (0.17)	2.34 (0.13)	3.22 (0.14)	3.23 (0.16)	3.38 (0.17)	2.95 (0.17)	3.03 (0.15)	2.76 (0.14)	3.47 (0.12)
Correlation	0.81*						0.59					

Fig. 9 Average environmental friendliness ratings and average preference ratings for the design variants obtained from the post-task survey questions (* $p<0.05$, ** $p<0.01$)

The majority of subjects' answers to the post-task question that asked about their definitions for environmental friendliness were normal and showed basic understandings of the concept, with some exceptions. For example, one subject answered "Marketing language in order to appeal to people that they are getting something they want and "helping" society. Mostly, it is disingenuous". One subject mentioned that they had no clue, and another subject thought environmental friendliness meant modern, trendy, and eye pleasing.

One of the authors coded the subjects' answers to the post-task question, which asked them to explain how they decided the environmental friendliness of the given product images during the experiment. Factors that the subjects considered in their decisions, indicated in their

answers, were categorized as: consumption of materials, shape/body shape, design/aesthetics, aerodynamics, size, effectiveness in use/ease of use, heating surface/grille, feedback information, and others (including factors that only a few subjects mentioned, such as weight, comfort, and manufacturability). Factors that the subjects mentioned most frequently were consumption of materials, shape/body shape, and design/aesthetics.

4.6 Discussion

Overall, the two case products have similar results. As summarized in Table 6, the results support the three research propositions in this paper.

Table 6 Hypotheses test results support the three research propositions (Hypothesis holds true at + 0.1 level, * 0.05 level, ** 0.01 level, or *** 0.0001 level)

H		Hypothesized Trend	Product					
			Electric Bicycle			Electric Heater		
1	a	$\overline{E_P} > \overline{E_\varnothing}$	$\{Body^P, Feature^P\}$	$\{Body^P, Feature^\varnothing\}$	$\{Body^\varnothing, Feature^P\}$	$\{Body^P, Feature^P\}$	$\{Body^P, Feature^\varnothing\}$	$\{Body^\varnothing, Feature^P\}$
			**	*		**	**	
	b	$\overline{E_N} < \overline{E_\varnothing}$	$\{Body^N, Feature^N\}$	$\{Body^N, Feature^\varnothing\}$	$\{Body^\varnothing, Feature^N\}$	$\{Body^N, Feature^N\}$	$\{Body^N, Feature^\varnothing\}$	$\{Body^\varnothing, Feature^N\}$
				+		***	**	
2		$\overline{\Phi_{TC} - \Phi_{TU}} > \overline{\Phi_{AC} - \Phi_{AU}}$	***			***		

Results of Hypothesis 1a and 1b indicate that mental associations between the body-cues and environmental friendliness formed using the selected association-building task in this experiment. Mental associations between the feature-cues and environmental friendliness did not form in this experiment or were not strong enough to affect consumer judgments. In detail: Hypothesis 1a holds true for $\{Body^P, Feature^P\}$ and $\{Body^P, Feature^\varnothing\}$ of both products. Note that, for brevity, when a hypothesis is said to "hold true" in this paper, it means that we fail to reject the null hypothesis at a certain significance level. After the association-building task, subjects rated the positive body-cue image as significantly more environmentally-friendly than

the neutral images. Hypothesis 1a is not accepted for *{Body$^{\varnothing}$, Feature^{P}}* of either product. The result of Hypothesis 1b for bicycle's *{Body^{N}, Feature^{N}}* follows the hypothesized trend but does not reach statistical significance. Hypothesis 1b holds true for heater's *{Body^{N}, Feature^{N}}*. It also holds true for *{Body^{N}, Feature$^{\varnothing}$}* of both products. After the association-building task, subjects rated the bicycle's negative body-cue only image as significantly less environmentally friendly than the neutral images; and subjects rated the heater's negative body-cue image as significantly less environmentally friendly than that having no cues. Hypothesis 1b is not accepted for *{Body$^{\varnothing}$, Feature^{N}}* of either product. As each of Hypothesis 1a and 1b involves multiple pairwise t-tests, to be conservative about the results, a Bonferroni correction is performed for each hypothesis. Table 7 provides the hypotheses test results after the Bonferroni correction, which still support Proposition 1 and 2.

Table 7 Results of Hypotheses 1a and 1b after Bonferroni corrections still support Proposition 1 and 2 (√: Hypothesis holds true at 0.0083 level)

H		Hypothesized Trend	Product					
			Electric Bicycle			Electric Heater		
1	a	$\overline{E_P} > \overline{E_\varnothing}$	*{Body^{P}, Feature^{P}}*	*{Body^{P}, Feature$^{\varnothing}$}*	*{Body$^{\varnothing}$, Feature^{P}}*	*{Body^{P}, Feature^{P}}*	*{Body^{P}, Feature$^{\varnothing}$}*	*{Body$^{\varnothing}$, Feature^{P}}*
			√			√	√	
	b	$\overline{E_N} < \overline{E_\varnothing}$	*{Body^{N}, Feature^{N}}*	*{Body^{N}, Feature$^{\varnothing}$}*	*{Body$^{\varnothing}$, Feature^{N}}*	*{Body^{N}, Feature^{N}}*	*{Body^{N}, Feature$^{\varnothing}$}*	*{Body$^{\varnothing}$, Feature^{N}}*
						√	√	

Results of Hypotheses 1a and 1b demonstrate different effectiveness of the body- and the feature- cues. This is further validated by (1) comparing *{Body^{P}, Feature$^{\varnothing}$}*-versus-*{Body$^{\varnothing}$, Feature$^{\varnothing}$}* difference in environmental friendliness rating from the testing task with *{Body$^{\varnothing}$, Feature^{P}}*-versus-*{Body$^{\varnothing}$, Feature$^{\varnothing}$}* difference; and (2) comparing *{Body^{N}, Feature$^{\varnothing}$}*-versus-*{Body$^{\varnothing}$, Feature$^{\varnothing}$}* difference with *{Body$^{\varnothing}$, Feature^{N}}*-versus-*{Body$^{\varnothing}$, Feature$^{\varnothing}$}* difference. In comparison (1), average *{Body^{P}, Feature$^{\varnothing}$}*-versus-*{Body$^{\varnothing}$, Feature$^{\varnothing}$}* difference in

environmental friendliness rating significantly differed from average $\{Body^{\varnothing}, Feature^{P}\}$-versus-$\{Body^{\varnothing}, Feature^{\varnothing}\}$ difference for electric bicycle (0.46 vs. -0.34, $p<0.01$ by pairwise t-test). In comparison (2), average $\{Body^{N}, Feature^{\varnothing}\}$-versus-$\{Body^{\varnothing}, Feature^{\varnothing}\}$ difference significantly differed from average $\{Body^{\varnothing}, Feature^{N}\}$-versus-$\{Body^{\varnothing}, Feature^{\varnothing}\}$ difference for both products (for electric bicycle: -0.28 vs. 0.09, $p<0.1$ by pairwise t-test; for electric heater: -0.59 vs. -0.02, $p<0.05$ by pairwise t-test).

Additional evidence from analyzing the environmental friendliness ratings that the subjects gave to $\{Body^{P}, Feature^{N}\}$ and $\{Body^{N}, Feature^{P}\}$ during the testing task also show different effectiveness of the body- and the feature- cues. For the electric bicycle, $\{Body^{P}, Feature^{N}\}$ obtained a significantly higher average environmental friendliness rating than $\{Body^{N}, Feature^{P}\}$ (3.23 vs. 2.71, $p<0.05$ by pairwise t-test). For the electric heater, $\{Body^{P}, Feature^{N}\}$ also obtained a significantly higher average environmental friendliness rating than $\{Body^{N}, Feature^{P}\}$ (3.32 vs. 2.76, $p<0.05$ by pairwise t-test). These results show that when the body-cue and the feature-cue contradict each other, the body-cue dominates the environmental friendliness rating.

The ineffectiveness of the feature-cue might be due to the fact that a feature accounts for a portion of the design and subjects may consider the feature-cue as a weak or secondary cue. Influences of the rest of the areas in the design may override any effects of the feature-cue. These results regarding the effectiveness of the body- vs. the feature- cues have some limitations. The comparisons here happen between the body and a particular feature for each product. If selecting another feature to cue the product's environmental friendliness, the results might change. But the consistent results between the two case products partially mitigate this concern. She and MacDonald [22] and She [34] used features to subliminally trigger preference for sustainability.

Our experiment did not test to see if feature-cues had a subliminal effect. Another limitation results from choosing environmental friendliness as the unobservable attribute. Some factors that determine a product's environmental friendliness such as material consumption and ease of manufacture need subjects to make holistic observations of product images. Therefore, this paper's choice of unobservable attribute may promote subjects to use body-cues, the holistic cues, rather than feature-cues. Results might change if choosing another unobservable attribute. In addition, the body-cue and the feature-cue have size differences. As the body-cue is larger than the feature-cue in this experiment, the body-cue should be more salient and easier to identify for subjects. This nature of the body-cue could contribute to its effectiveness observed in the experiment. Du and MacDonald [27] demonstrated that gaze data can differentiate noticeable and unnoticeable feature size changes. The experiment in this paper did not test the saliency of the size difference between the body-cue and the feature-cue.

Once subjects had built mental associations, they shifted their focus such that they spent more percentage of time looking at the cued vs. uncued areas of interest on the product images. This suggests that their decisions became more "frugal." The subjects' reliance on the cues to judge environmental friendliness after the association-building task results in more targeted evaluations during the testing task, which can reduce mental burden in decisions. Hypothesis 2 holds true for both products. The testing task of the experiment has a significantly greater percentage-fixation time difference between the cued and the uncued AOIs than that in the preceding association-building task. The AOI group and the experimental task have a significant interaction effect on the percentage-fixation time. After the association-building task, the subjects increased the percentage-fixation time spent on the cued AOIs and decreased the percentage-fixation time spent on the uncued AOIs. The association-building task allowed

subjects to learn visual cues and form associations between the cues and environmental friendliness. Learning the cues and the associations may increase importance of the cued areas and decrease importance of the uncued areas to subjects in the case of judging environmental friendliness. Therefore, in the testing task that followed, subjects redistributed their gaze attention by adjusting the percentage-fixation time they spent on the cued vs. the uncued AOIs as Hypothesis 2 validates.

The comprehension test results show that majority of subjects remembered the design variants appeared in the experiment as well as the given feedback information on the body-cue image's environmental friendliness. The results confirm the effectiveness of the association-building task. In the post-task questions, subjects rated the positive body-cues as significantly more environmentally friendly than the negative body-cues (by least significant difference test). The heater's positive feature-cue received significantly higher environmental friendliness rating than its negative feature-cue in the post-task question. This indicates that mental associations between the heater's feature-cue and environmental friendliness may have formed to certain extent. As the heater's feature-cue did not affect the heater's environmental friendliness rating according to Hypothesis 1a and 1b results, the associations, if formed, were weak.

For the bicycle's frame and handlebar and heater's case, preference ratings for the design variants correlated with the environmental friendliness ratings. It is possible that after rating environmental friendliness for 60 product images, subjects held a mind-set of considering environmental friendliness and so tended to prefer design variants that were perceived as more environmentally friendly.

When the post-task question asked subjects to specify how they decided the products' environmental friendliness during the experiment, they mentioned material consumption most

frequently. Their concern on material consumption may facilitate building mental associations between body-cues and environmental friendliness, as the body of a product plays a large role in determining material consumption. This may also help explain the ineffectiveness of the feature-cues in the experiment.

According to the experimental results, the association-building task selected in this experiment works in terms of building mental associations between the body-cues and the products' environmental friendliness. But the unselected approach, mentioned in Section 4.4.2, did not manage to build any mental associations as indicated by the results of the pilot study. Different outcomes of the two approaches indicate that the possibility to build mental associations depends on the nature of the association-building task. The unselected association-building task asked pilot-study subjects to evaluate 10 pairs of choice alternatives and indicate their preferences for each pair separately. Each choice alternative in a pair included a product image and four text-described product attributes. The choice alternative's predetermined environmental friendliness rating was given as one of the four attributes. The unselected association-building task had the same product images as the selected task. During post-task interviews, the pilot-study subjects mentioned that they did not think of the associations as each pair already had much information for them to process. As people need to process information in their minds in order to form mental associations, association-building tasks that impose extra mental burden like the unselected task may fail to serve their purposes.

4.7 Conclusion

This study demonstrates the possibility to build fast and frugal mental associations between a product's body-cues and its environmental friendliness. After building the

associations, the body-cues can affect consumer judgments either positively or negatively, depending on the associations that are built. If the findings can apply to other situations, they suggest that designers can construct mental associations between product visual cues and certain unobservable attributes intentionally within a short time frame and then use the visual cues to influence consumer judgments. Once subjects build the mental associations, they make more efficient, or frugal, judgments with less mental burden. The study also shows that it is more effective to cue through body shape than individual features.

It is important to put these findings in context with our previous findings. In [27], we found that time spent looking at a feature AOI could predict that features' importance in a preference decision. However, in this paper, the features do not stand-out to subjects, even when cued, thus suggesting they do not affect environmental friendliness ratings. An explanation for this is that subjects may tend to consider design details as much as possible in order to determine their preferences in the frame of the study in [27], but they may take some shortcuts when rating environmental friendliness and therefore only focus on design aspects that are most crucial to their decisions in the study here. Next, in [4] we were interested in how feature evaluations came together to make a whole. We found that when two products with some shared features were presented for choice, the configurations of those shared features were evaluated in the choice even though they were identical. This suggested a more holistic assessment of the product, rather than feature-by-feature. This is consistent with what this paper identifies. Body-cues can affect environmental friendliness ratings but feature-cues cannot, indicating a holistic evaluation strategy. In Reid et al. [20], we found that a difference presentation form (i.e. sketch vs. rendering) resulted in a difference in opinions and objective evaluations, but not inferences. An inference was defined as a judgment that could not be accurately made using the visual image

presented, but did have a 'correct' answer as opposed to an opinion. As rating environmental friendliness, or any unobservable product feature, must be an inference when presented with only visual information, then it is likely that the findings in this paper will hold across visual presentation style.

This study adds to existing research on visual working memory. In the study, body-cues affected environmental friendliness ratings in the desired direction after the association-building task, indicating that subjects managed to remember and recognize the body-cues after the task, although they also looked at other features and other design variants. This observation provides evidence for other researchers' investigations on relationships between visual working memory and visual attention. For example, this observation lends support to a conclusion made by Maxcey-Richard and Hollingworth [35], task-relevant objects remains in visual working memory regardless of the proceeding of attention. This study also provides evidence for top-down guidance in visual search. In the study, subjects redistributed their attention to cued and uncued areas after the association-building task. This follows a particular form of top-down guidance: knowledge or memory for the relevant object's visual characteristics (here, the cued areas are the relevant objects) [36].

This study has some limitations. Although reasonable in approach, it is unknown if the difference in effectiveness of body- vs. feature-cues was in-part influenced by the experiment, in terms of: (1) the selected association-building task; (2) the selected products; (3) the selected unobservable attribute; (4) the selected product designs; (5) the rating scale associated with them; and (6) their size differences. The duration of the mental associations built in the experiment is unknown—will subjects remember these associations a month or a year later? While not crucial to the tests at-hand, this is important for implementation of cues long term.

There may be a combination of limitations, for example, although features were not effective cues in this 'fast' experiment, it may be that they are even more effective than body-cues when given a long time to take-hold. This study only focuses on cuing a product's environmental friendliness. Whether or not designers or researchers can apply these results to other unobservable product attributes like safety, reliability, and quality is unknown, but seems likely given the existence of real-world examples of such cues. Researchers could also further the study by investigating how the visual cues will interact with text information, which can convey either consistent or contradicting information as the visual cues, to affect consumer judgments.

Acknowledgments

The authors would like to thank: iMotions Inc. for providing software and technical support; Dr. Eric Cooper and Dr. Robert West at Iowa State University for their advice; and Dr. Judy Vance at Iowa State University for all of her support for this study.

Nomenclature

E Environmental friendliness rating

P Positive cue

$\varnothing$ No cue

N Negative cue

B Body-cue

F Feature-cue

Φ Percentage-fixation time

C Cued areas of interest

U Uncued areas of interest

T Testing task

A Association-building task

i Index of the subject

I Number of subjects

j Index of the product image

J Number of product images

References

[1] Reid, T. N., Gonzalez, R. D., and Papalambros, P. Y., 2010, "Quantification of Perceived Environmental Friendliness for Vehicle Silhouette Design," Journal of Mechanical Design, 132(10), p. 101010.

[2] MacDonald, E. F., Gonzalez, R., and Papalambros, P. Y., 2009, "The construction of preferences for crux and sentinel product attributes," Journal of Engineering Design, 20(6), pp. 609-626.

[3] Gigerenzer, G., 2004, "Fast and frugal heuristics: The tools of bounded rationality," Blackwell handbook of judgment and decision making, D. J. Koehler, and N. Harvey, eds.Oxford, UK.

[4] Du, P., and MacDonald, E., 2015, "Products' Shared Visual Features Do Not Cancel in Consumer Decisions," Journal of Mechanical Design, 137(7), p. 071408.

[5] She, J. J., and MacDonald, E., 2014, "Priming Designers to Communicate Sustainability," Journal of Mechanical Design, 136(1).

[6] MacDonald, E. F., and She, J., 2015, "Seven cognitive concepts for successful eco-design," Journal of Cleaner Production, 92, pp. 23-36.

[7] Berkowitz, M., 1987, "Product Shape as a Design Innovation Strategy," Journal of Product Innovation Management, 4(4), pp. 274-283.

[8] Chandler, J. J., Reinhard, D., and Schwarz, N., 2012, "To judge a book by its weight you need to know its content: Knowledge moderates the use of embodied cues," Journal of Experimental Social Psychology, 48(4), pp. 948-952.

[9] Wänke, M., Bless, H., and Schwarz, N., 1998, "Context Effects in Product Line Extensions: Context is Not Destiny," Journal of Consumer Psychology, 7(4), pp. 299-322.

[10] Dawar, N., and Parker, P., 1994, "Marketing Universals: Consumers' Use of Brand Name, Price, Physical Appearance, and Retailer Reputation as Signals of Product Quality," Journal of Marketing, 58(2), pp. 81-95.

[11] Kirmani, A., 1990, "The Effect of Perceived Advertising Costs on Brand Perceptions," Journal of Consumer Research, 17(2), pp. 160-171.

[12] Brown, C. L., and Carpenter, G. S., 2000, "Why Is the Trivial Important? A Reasons-Based Account for the Effects of Trivial Attributes on Choice," Journal of Consumer Research, 26(4), pp. 372-385.

[13] Simonson, I., Carmon, Z., and O'curry, S., 1994, "Experimental Evidence on the Negative Effect of Product Features and Sales Promotions on Brand Choice," Marketing Science, 13(1), pp. 23-40.

[14] Carpenter, G. S., Glazer, R., and Nakamoto, K., 1994, "Meaningful Brands from Meaningless Differentiation: The Dependence on Irrelevant Attributes," Journal of Marketing Research, 31(3), pp. 339-350.

[15] Tseng, I., Cagan, J., Kotovsky, K., and Wood, M., 2013, "Form Function Fidelity," Journal of Mechanical Design, 135(1), p. 011006.

[16] Lai, H. H., Chang, Y. M., and Chang, H. C., 2005, "A Robust Design Approach for Enhancing the Feeling Quality of a Product: A Car Profile Case Study," International Journal of Industrial Ergonomics, 35(5), pp. 445-460.

[17] Dagher, A., and Petiot, J.-F., "Study of the Correlation between User Preferences and Design Factors: Application to Cars Front-end Design," The International Conference on Engineering Design, ICED'07, Paris, Aug. 28–31.

[18] Orbay, G., Fu, L. T., and Kara, L. B., 2015, "Deciphering the Influence of Product Shape on Consumer Judgments Through Geometric Abstraction," Journal of Mechanical Design, 137(8), p. 081103.

[19] Yumer, M. E., Chaudhuri, S., Hodgins, J. K., and Kara, L. B., 2015, "Semantic Shape Editing Using Deformation Handles," ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2015, 34(4).

[20] Reid, T., MacDonald, E., and Du, P., 2013, "Impact of Product Design Representation on Customer Judgment," Journal of Mechanical Design, 135(9), p. 091008.

[21] MacDonald, E. F., Lubensky, A., Sohns, B., and Papalambros, P. Y., 2009, "Product

Semantics and Wine Portfolio Optimisation," International Journal of Product Development, 7(1/2), pp. 73-98.

[22] She, J., and MacDonald, E., 2013, "Trigger Features on Prototypes Increase Preference for Sustainability," Proceedings of the ASME 2013 International Design Engineering Technical Conference & Computers and Information in Engineering ConferencePortland, Oregon.

[23] Kelly, J. C., Maheut, P., Petiot, J., and Papalambros, P., 2011, "Incorporating User Shape Preference in Engineering Design Optimisation," Journal of Engineering Design, 22(9), pp. 627-650.

[24] Swamy, S., Orsborn, S., Michalek, J., and Cagan, J., 2007, "Measurement of Headlight Form Preference Using Choice Based Conjoint Analysis," Proceedings of the ASME 2007 International Design Engineering Technical Conferences & Computers and Information in Engineering ConferenceLas Vegas, Nevada, pp. 197-206.

[25] Orsborn, S., Cagan, J., and Boatwright, P., 2009, "Quantifying Aesthetic Form Preference in a Utility Function," Journal of Mechanical Design, 131(6), p. 061001.

[26] Just, M. A., and Carpenter, P. A., 1976, "Eye Fixations and Cognitive Processes," Cognitive Psychology, 8(4), pp. 441–480.

[27] Du, P., and MacDonald, E., 2014, "Eye-Tracking Data Predict Importance of Product Features and Saliency of Size Change," Journal of Mechanical Design, 136(8), pp. 081005-081005-081014.

[28] Du, P., and MacDonald, E., 2015, "Eye-tracking Aids in Understanding Consumer Product Evaluations," The Psychology of Design: Creating Consumer Desire, R. Batra, D. Brei, and C. Seifert, eds., M.E. Sharpe, Inc.

[29] Duchowski, A. T., 2007, Eye Tracking Methodology: Theory and Practice, Springer, London.

[30] iMotions, Accessed Apr. 18, 2016, http://imotionsglobal.com/.

[31] Ashby, F. G., Maddox, W. T., and Bohil, C. J., 2002, "Observational versus Feedback Training in Rule-based and Information-integration Category Learning," Memory & Cognition, 30(5), pp. 666-677.

[32] Nenkov, G. Y., Morrin, M., Ward, A., Schwartz, B., and Hulland, J., 2008, "A Short Form of the Maximization Scale: Factor Structure, Reliability and Validity Studies," Judgment and Decision Making, 3(5), pp. 371-388.

[33] Pernice, K., and Nielsen, J., 2009, "How to Conduct Eyetracking Studies," Nielsen Norman Group, Fremont, CA.

[34] She, J., 2013, "Designing Features that Influence Decisions about Sustainable Products," Ph.D. thesis, Iowa State University, Ames, Iowa.

[35] Maxcey-Richard A. M., and Hollingworth A., 2013, "The Strategic Retention of Task-Relevant Objects in Visual Working Memory," Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(3), pp. 760-772.

[36] Hollingworth, A., Matsukura, M., and Luck, S. J., 2013, "Visual Working Memory Modulates Low-level Saccade Target Selection: Evidence from Rapidly Generated Saccades in the Global Effect Paradigm," Journal of Vision, 13(13): 4, pp. 1-18.

CHAPTER 5

CONCLUSION

This dissertation contributes to the product design field by providing useful insights in impacts of product visual designs on consumer judgments as well as how consumers evaluate these designs to make judgments. It borrows knowledge from psychology to study problems that concern product designers. It also contributes to incorporating eye-tracking in design research and laying foundations for interpreting eye-tracking data in the context of product evaluations.

In Chapter 2, Study 1 directly examines how subjects decode product designs. It proves linear correlations between feature importance in preference decisions and gaze data associated with the feature. Study 1 suggests a new approach to identify feature importance, which saves time and effort for both designers and subjects when comparing with other approaches like discrete choice analysis and traditional rating. Also, Study 1 demonstrates the ability of gaze data to differentiate noticeable and unnoticeable size changes of a product feature, which can help with optimizing the level of manufacturing imperfections with respect to geometrical variations for products. Additionally, both of these results can facilitate real time product design and product personalization.

In Chapter 3, Study 2 revisits the C&F model from psychology in the context of product design to seek guidance in designing product commonalities and differences. As predicted by the C&F model, the unique feature/attribute attracts more gaze attention than the shared feature/attribute in five of the six tested conditions, indicating the importance of product differences in consumer preferences. But, this evaluation strategy does not affect subjects' preference decisions as the model predicted when the test stimuli included product images.

Trends opposite to the model's predictions are observed in postpreference evaluations in the two image-only conditions, indicating that shared visual features were not canceled in consumer decisions, instead they can reinforce impressions. This finding highlights the importance of shared features in design and reminds designers to approach them with caution. When designing and determining shared features, designers should consider possible reinforcements the shared features may create. This can help designers avoid the risk that a new product or a brand portfolio leaves consumers with undesired impressions due to including inappropriately shared features.

In Chapter 4, Study 3 demonstrates the possibility of rapidly building mental associations between a product's body-cues and the product's environmental friendliness. These results reinforce the usefulness of visual cues for designers to deliver messages to consumers and affect consumer judgments. They also demonstrate the potential of implementing visual cues in real time product design. Study 3 finds it more effective to cue through body shapes than features, which points to a direction for designers to pursue and helps them spend their time and effort efficiently. Gaze data show that subjects' evaluations become more efficient and frugal after mental associations are built, indicating that they can make judgments with less mental burden.

As a whole, Study 1 found that a feature's gaze data correlated with the feature's importance in preference decisions, implying that subjects used a feature-by-feature evaluation strategy for product images. Study 2 observed that in the image-only conditions, shared features attracted less gaze attention than unique features, which further implies the feature-by-feature evaluation or comparison of product images. However, in Study 2, preferences and postpreference evaluations for product images did not follow the predictions of the C&F model, which relies on an assumption that decisions are based on attribute/feature-by-attribute/feature

comparisons. Additionally, shared visual features were not canceled in subjects' decisions. A possible explanation for the results of Study 2 is that the feature-by-feature evaluation of product images cannot dominant subjects' judgments (specifically, opinions) and the subjects employed a holistic evaluation strategy at the mean time, which can largely affect judgment. Study 3 validates this holistic evaluation, as it found that it was more effective to cue holistically through body shapes than by individual features. In addition, as visual cues are shared among product images, Study 3 also contributes to the research on product commonalities. The possibility of rapidly building mental associations between visual cues and environmental friendliness as well as the effect of visual cues on environmental friendliness ratings both support the finding of Study 2 that visual product commonalities do not cancel in consumer decisions.

All three studies provide evidence that consumers tend to reduce mental burden while they make decisions about product designs. In Study 1, subjects spent more gaze attention on more important features. This prioritizing behavior can be considered as a way to reduce mental burden. In Study 2, although subjects did not cancel shared features in their decisions, they did spend less gaze attention on the shared features in most cases, indicating that the subjects tried to save their efforts during decision-making. In Study 3, after mental associations between visual cues and environmental friendliness were built, subjects reduced their mental burden by increasing percentage of gaze attention spent on cued areas (areas that were more relevant to their decisions) and decreasing percentage of gaze attention spent on uncued areas (areas that were less relevant). Designers must take this tendency of consumers into consideration when designing products. By doing this, designers may better foresee how consumers will evaluate and perceive designs. Designers can then appropriately adapt their designs to ensure that the designs will be perceived as desired.

This dissertation also contributes to the communities that study decision-making, visual working memory, and visual search. Study 2 extends the test of the C&F model to product images. It finds that predictions of the C&F model on preference decisions and postpreference evaluations do not hold in the image-only conditions, which help demonstrate possible caveats for applying the C&F model. Study 3 adds to existing research on visual working memory and visual search as discussed in Chapter 4.

This dissertation has a limitation due to the type of stimuli it utilizes. In these three studies, restricted by the available eye-tracker, product stimuli are presented on screens. Therefore, applying these research findings to physical objects should be proceeded carefully and thoughtfully. Decoding physical objects involves more human sensors, which could lead to different results. However, this limitation can be mitigated by the current popularity of online shopping where products are shown on screens.

One major open question is regarding consumers' uses of feature-by-feature and holistic evaluations. Results from these three studies indicate that subjects use both types of evaluations, but holistic evaluation may have a larger effect on their judgment. This needs further validations. It is worth identifying, in detail, how these two types of evaluations function to affect consumer judgments. It is possible that consumers dynamically change their emphases on these two types of evaluations according to the judgments they need to make. Future research can also investigate if gaze data can signal a consumer's use of holistic evaluations. A thorough investigation into this open question may better guide product design decisions.

APPENDIX

ADDITIONAL RESULTS OF STUDY 2

Table A1 A summary of hypothesis testing results after Bonferroni corrections (√: Hypothesis holds true at 0.0083 level)

H	Hypothesized Trend	Conditions					
		ISeq	TSeq	ITSeq	ISBS	TSBS	ITSBS
1a: Choice rating	$\overline{V_G} > \overline{V_B}$		√				
1b: Satisfaction rating	$\overline{S_G} > \overline{S_B}$						
1c: Good-ness rating	$\overline{\Gamma_{GA}} > \overline{\Gamma_{BA}}$			√			√
	$\overline{\Gamma_{GR}} > \overline{\Gamma_{BR}}$		√			√	
2: Fixation time and count	$\overline{T_U} > \overline{T_H}$				√	√	√
	$\overline{Q_U} > \overline{Q_H}$		√		√	√	√
3: Transformed choice rating	$\overline{V'_G} > \overline{V'_B}$						