POLITECNICO DI MILANO
DEPARTMENT DESIGN
DOCTORAL PROGRAMME IN DESIGN

INTERFACE DESIGN FOR USER DECISION IMPROVEMENT IN E-COMMERCE

Doctoral Dissertation of:
Dongning Yan

Supervisor:
Prof. Alberto Vitale Colorni

The Chair of the Doctoral Program:
Prof. Francesco Trabucco

Year 2015 – Cycle XXVII

Acknowledgements

Here, I wish to express my heartfelt thanks to a host of people, without whose assistance the accomplishment of the thesis would have been impossible.

Firstly, I would like to express my deepest gratitude to my supervisor, Prof. Colorni Vitale Alberto. Thanks for his consistent guidance during these three years of my PhD study and solid support for the completion of the thesis.

I am also grateful to Prof. Chen Li at Hong Kong Baptist University. She provided me a valuable exchange opportunity and has given me many useful suggestions and advice.

Many thanks to my friends, such as Bao Qing at Hong Kong Baptist University and Hu Tingting at Hunan University, who have generously offered their help and time in listening to me and helping me work out my problems during the difficult course of the thesis. I would extend my thanks to all people who devoted time to participate in my user studies.

I greatly appreciate Chinese Scholarship Council for offering me a full scholarship to study abroad.

Finally, my thanks would go to my beloved family for their consistent support and encouragement all through these years!

February 2015

Abstract

Customer reviews in E-commerce are playing an important and unique role; a staggering 90 percent of people use and monitor reviews in their online purchasing process. However, the overwhelming number of reviews and inconsistent writing style require significant effort to read and tend to let important information slip by. To help users effectively and efficiently glean information from reviews, a number of systems have summarized customer reviews by extracting features and associate sentiment toward each feature. From the perspective of customers, online purchasing can be viewed as a decision making process. In light of human decision-making theory, we learn that the foundation of effective information displays for user decision improvement is gaining a deep understanding of user decision-making behavior. However, no clear picture exists to systematically elaborate on consumer decision-making behavior in E-commerce, in particular, with respect to customer reviews.

In this thesis, we take online hotel booking as an example to empirically investigate consumer decision-making behavior in three stages of online purchasing: (1) screening out interesting alternative(s) for further consideration, (2) evaluating alternatives in detail, and (3) comparing candidates to make the final choice. Interfaces that aggregate information from customer reviews have been developed to support the three alternative stages. Through analysis of the results, we identify the decision strategies users utilize to process information and the information they are inclined to seek at each stage. These findings lay solid groundwork to design E-commerce interfaces for consumer decision improvement.

Concerning user decision-making behavior in the stage of screening out interesting alternatives, we find that: (1) 94% of participants began by eliminating alternatives with values for an attribute below a cut-off to simplify the complexity of the choice; (2) 55.3% of participants eliminated alternatives by both static features (i.e., product specifications) and customer reviews. Moreover, the number of users who adopted opinion attributes (i.e., attributes extracted from customer reviews) is significantly higher than that using an overall review score; and (3) the cut-off values are determined by the value distribution of an attribute and correlation among attributes, in addition to stable preference. Grounded in these user decision-making behaviors, we framed two alternative designs for an opinion-attribute-embedded filter panel based on checkboxes and sliders. In the checkbox interface, the filter for each attribute is represented in the form of an array of N checkboxes, which is utilized in most E-commerce websites. In the slider interface, the filter for each numerical attribute (i.e., price and opinion attributes) is represented by a modified slider, which visualizes the distribution of an attribute via bars and the correlation among attributes via simultaneous change. Then, we performed a user study to compare the two alternative designs in the context of online hotel booking. The results show that people depended highly on opinion attributes to narrow down the range of options in both interfaces, which points to the effectiveness of incorporating opinion attributes in filters. And the slider interface achieves significantly higher user assessments in terms of perceived decision accuracy, cognitive effort, pleasantness to use and intention to return.

After narrowing down options to a smaller set, 40% of participants adopted a more compensatory strategy – Weighted Additive Difference, i.e., comparing the remaining alternatives on multiple attributes and selecting the alternative with the best overall value. More notably, significantly more participants compared alternatives by opinion attributes in comparison with those associated with an overall review score. Therefore, we developed a multi-attribute sorting panel embedded with opinion attributes. Furthermore, the multi-attribute sorting panel was expanded to three alternative designs that mainly differ in the way of eliciting relative importance for attributes: (1) direct assessment, asking users to directly assign weights to attributes; (2) indifference method, modifying one of two sets of stimuli until subjects feel that there is no difference between the two; and (3) indirect measurement, giving relative preference on a pair of alternatives. Through analysis of objective and subjective measures, the multi-attribute sorting was verified to be beneficial to consumer online purchasing. The direct way outperforms the indifference and indirect ways regarding perceived decision accuracy, cognitive effort, satisfaction and intent to use in an E-commerce environment.

Motivated by the results of the above user studies, we have derived a set of guidelines on how to design interfaces for consumer purchase decision improvement in E-commerce.

Keywords: human decision making; interface design; E-commerce; customer reviews; user study

Contents

Part 4: Conclusion and Future Work

List of Figures

List of Tables

Chapter 1

Introduction

This chapter provides an overview of the thesis, giving an account of the background and objectives of the research, the contribution of the study and the organization of the thesis.

1.1 Background

With the rapid growth of E-commerce, online users are provided with access to a greater variety of products and an enormous amount of information on the products. The huge number of customer reviews in E-commerce is playing an important and unique role; a staggering **90 percent** of people use and monitor online reviews in their online purchasing process [95].

From the perspective of the customer, his/her motivation for delving into an enormous quantity of online reviews is to find suitable items that meet his/her desire. However, because the overwhelming number of reviews and inconsistent writing style require significant effort to read and tend to let important information slip by, gaining insights becomes increasingly challenging for users.

To help users effectively and efficiently glean information from reviews to buy proper items, most E-commerce websites (e.g., Amazon) provide an overall review score for each entity to facilitate users to filter/ sort alternatives. However, given that people examine the attributes of a product to evaluate whether the product fits their desire, a number of systems have summarized customer reviews by extracting features and associate sentiment toward each feature, referred to as feature-sentiment summarization. Liu *et al.* (2005) [52] and Carenini *et al.* (2009) [15] used **Bar charts** to show the sentiment of summarized features, which supports a side-by-side and feature-by-feature comparison of competing products. Carenini *et al.* (2006) [14] summarized user reviews in the form of a **Tree map** by representing a feature as a rectangle with nested rectangles corresponding to the descendants of the feature. In addition to numerical ratings, which provide a quick understanding of how much an entity is liked or disliked, some recent systems use adjective-noun word pairs to summarize the sentiment (adjective) towards each feature (noun) to help users explore reviews in greater detail. Yatani (2011) [93] represented adjective-noun word pairs in the form of **Tag cloud**, with the font size and color indicating the number of occurrences and sentiments (positive/negative). Huang *et al.* (2013) [42] developed a system that can automatically **highlight sentences** that are related to relevant features to make a balance between reducing information overload and providing the original review context.

Because in most conditions of online shopping, customers identify the need for a product or service without specific requirement on which to buy [61], they need to select interesting choice(s) from a range of options that satisfy their desires. Therefore, online purchasing can be viewed as a decision-making process from the perspective of the customer. Moreover, Chen (2010) interpreted online purchasing behavior as a three-stage decision-making process: (1) narrowing down available alternatives and selecting one that is worth further consideration, (2) reading detailed information about the item selected in the preceding stage and deciding whether to take it as purchase candidate, and (3) comparing several candidates and making the final choice [17]. The transition between the three stages is iterative in nature but does follow an approximate sequence on the whole.

To design E-commerce interfaces with the objective to assist users in making more accurate online purchasing decisions with less effort, we consult decision-making theory and learn that human decision behavior in reality is adaptive in nature. The same decision maker appears to selectively adopt information and use a wide variety of strategies contingent on decision properties [66]. The information display, to a large extent, can impact not only information acquisition (i.e., what information is noticed) but also information combination (i.e., which decision strategy is utilized to process the information), leading to higher/lower decision accuracy and less/more cognitive effort [70]. In general, an effective information display depends on two matches: on the one hand, the match between the importance of information for the decision maker and the salience of the information display and, on the other hand, the congruence between the information format and the way information is processed. Hence, the foundation of designing information displays for user decision improvement is gaining a deep understanding of human decision-making behavior: (1) what information decision makers are inclined to seek, and (2) what decision strategy they utilize to process the information.

Some researchers investigated how people use online rating to make choices. For example, Lelis S. and Howes A. (2011) suggested that people gather more information for the best alternative, and they take more time to inspect reviews of lower rating [50]. However, no clear picture exists to systematically elaborate on consumer decision-making behavior in E-commerce, in particular, with respect to customer reviews. Most relevant studies developed a summary of customer reviews without placing strong emphasis on human decision-making behavior.

1.2 Objectives

1. **Investigating customer decision-making behavior in E-commerce that include customer reviews**

   Firstly, we investigate user decision-making behavior at each stage of online shopping process. More concretely, answer the following questions:

   Analysis of the decision strategy that decision makers adopt

   A central distinction among strategies is the extent to which they make trade-offs among attributes. Decision strategies (such as Weighted Additive) that explicitly consider trade-offs are called compensatory strategies, whereas strategies (like Lexicographic) that do not make trade-offs are called non-compensatory strategies.

   -*RQ1*: which kind(s) of decision strategies do customers adopt to process information, compensatory strategies or non-compensatory strategies?

   Analysis of the information that decision makers seek

   In the E-commerce environment, each entity is described by diverse information. In general, the information can be classified into two types – static features (such as price and specifications) and customer reviews.

   -*RQ2:* which kind(s) of information do decision makers seek, static features or/and customer reviews?

   To further explore the role of each type of information in more detail, we investigate the following research question:

   -*RQ3:* which kind(s) of information in static features and customer reviews do decision makers seek?

The format in which the sentiment corresponding to attributes extracted from customer reviews is presented can also be different. Numerical values (e.g. average rating, number of reviews) provide an easy proxy for opinions, whereas verbal values (e.g. adjective-noun word pairs) provide reasons underlying the scores.

-*RQ4:* which kind(s) of values do decision makers refer to concerning the sentiments of attributes extracted from customer reviews, numerical or/and verbal?

2. Designing interfaces for user decision improvement in E-commerce

Based on the analysis of the results of user decision-making behaviors in an E-commerce environment, especially with respect to customer reviews, we can gain insight on the data requirements and functional requirements that connect the gap between user research and interface design. For example, the results of this formative study show that in the process of eliminating alternatives whose values on an attribute are below a certain value, people determine the cut-off in terms of attribute correlation. This finding suggests that the interface should support low-cost access to information about correlation among attributes.

Motivated by the extracted requirements, we hope to put forward specific solutions to customer-review-embedded E-commerce interface design with the objective of helping users effectively and efficiently glean information to make better purchasing decisions, i.e., reducing the effort required while augmenting the decision accuracy. More specifically, figure out two design questions in terms of both data and functional requirements: (1) what information should be represented and (2) how to represent the information by taking advantage of knowledge on information visualization, ensuring that important data elements can be quickly perceived and the information format is in congruence with the forms of information processing,

1.3 Contribution

1. **Constructing the models of user decision-making behavior in E-commerce**

   We conduct a formative study to empirically investigate consumer decision-making behavior in the area of online-hotel booking. Participants are asked to perform three tasks corresponding to the three-stage decision process: (1) choosing interesting hotels from a list of 10 for further consideration, (2) reading detailed information about the selected hotels, and (3) comparing candidates to confirm the final choice. To monitor what information is acquired and how the captured information is processed, two process tracing methods are employed: verbal protocols and computerized process tracing.

   Through the analysis of the results of the formative study, we classify online customers into 4 types in terms of the decision strategies they used and identify the similarities and differences between different types of users in the decision strategy they utilize to process information and the information they are inclined to seek, in the three stages of online purchasing. Then, we create a set of personas encapsulating the key findings, many of which have not yet been exploited in the design space. The user models lay solid groundwork for generating new, previously unconsidered designs for customer purchasing decision improvement in E-commerce.

2. **Translating knowledge of user decision-making behavior into interface design**

   Concerning user decision-making behavior in the stage of screening out interesting alternatives, we find that: (1) the majority of customers began by eliminating alternatives with values for an attribute below a cut-off to simplify the complexity of the choice; (2) in addition to static features (i.e., product's specifications), 68.1% of participants (31/47) eliminated options in terms of customer reviews; and the number of users who adopted opinion attributes (i.e., attributes extracted from customer reviews) is significantly higher than that using the overall review score; (3) the cut-off value of an attribute is conditional on the value distribution of the attribute and the correlation among attributes. Grounded in these decision-making behaviors, we derived design requirements and framed two alternative designs for an opinion-attribute-embedded filter panel based on checkboxes and sliders. In the checkbox interface, the filter for each attribute is represented in the form of an array of N checkboxes, which is utilized in most E-commerce websites. Each checkbox denoting a certain value range on this dimension is followed by the number of entities with values within the range. In the slider filter, the filter for each numerical attributes (e.g., price and opinion attributes) is represented by a modified slider, which visualizes the distribution of an attribute via bars and the correlation among attributes via simultaneous change.

After narrowing down alternatives to a smaller set, 40% of participants (20/50) adopted a more compensatory strategy – Weighted Additive Difference, i.e., comparing the remaining alternatives on multiple attributes and selecting the alternative with the best overall value. More notably, both participants who selected alternatives by a single attribute and those making decisions by multiple attributes depended highly on opinion attributes. Grounded in these user decision-making behaviors, we developed a multi-attribute sorting embedded with opinion attributes for online shopping. Furthermore, the multi-attribute sorting panel was expanded to three alternative designs that mainly differ in the way of eliciting relative importance for attributes: (1) direct assessment, asking users to directly assign weights to attributes; (2) indifference method, modifying one of two sets of stimuli until subjects feel that there is no difference between the two; and (3) indirect measurement, giving relative preference on a pair of alternatives.

3. **Testing and verifying the effectiveness of the design solutions**

We empirically proved the superiority of our design solutions in terms of measuring whether they could be ideally beneficial to E-commerce websites regarding improving customer decisions by asking users to perform several tasks requiring interaction with the system.

The first user study compared two alternative designs for an opinion-attribute-embedded filter panel, called checkbox interface and slider interface. The results show that in both interfaces, people depended highly on opinion attributes to narrow down the range of alternatives, which points to the effectiveness of incorporating opinion attributes in filters. In addition, the slider interface, which visualizes the value distribution via bars and attribute correlation via simultaneous change, obtained significantly higher user assessments in terms of perceived decision accuracy, perceived ease of use, satisfaction and intention to use.

Three alternative designs for an opinion-attribute-embedded sorting panel that primarily differ in the way of eliciting relative importance (i.e., weight) of attributes, called direct assessment, indifference method, and indirect measurement, were compared in the second user study. Through analysis of objective and subjective measures, the effectiveness of multi-attribute sorting function was verified, whereby, on average, more than 80% of participants sorted options by more than one attribute in three interfaces. Users gave significantly higher ratings to the direct way than to the indifference method and indirect measurement regarding perceived decision accuracy, cognitive effort, satisfaction and intent to use.

1.4 Overview of the Thesis

The content is organized as shown in Figure 1.1. We first introduce related research on the summary of customer reviews and human decision-making theory (chapter 2). We then describe the details of the formative study (chapter 3), including the experiment setup, materials and participants. The analysis of consumer decision-making behavior follows (chapter 4), which provides practical implications on E-commerce interface design. In terms of the implications, we develop concrete design solutions to the E-commerce that involve customer reviews to improve customer purchasing decisions (Chapter 5). Finally, through two user studies (chapter 6 and 7), we test and verify the effectiveness of the design solutions. A brief introduction for each chapter is listed below:

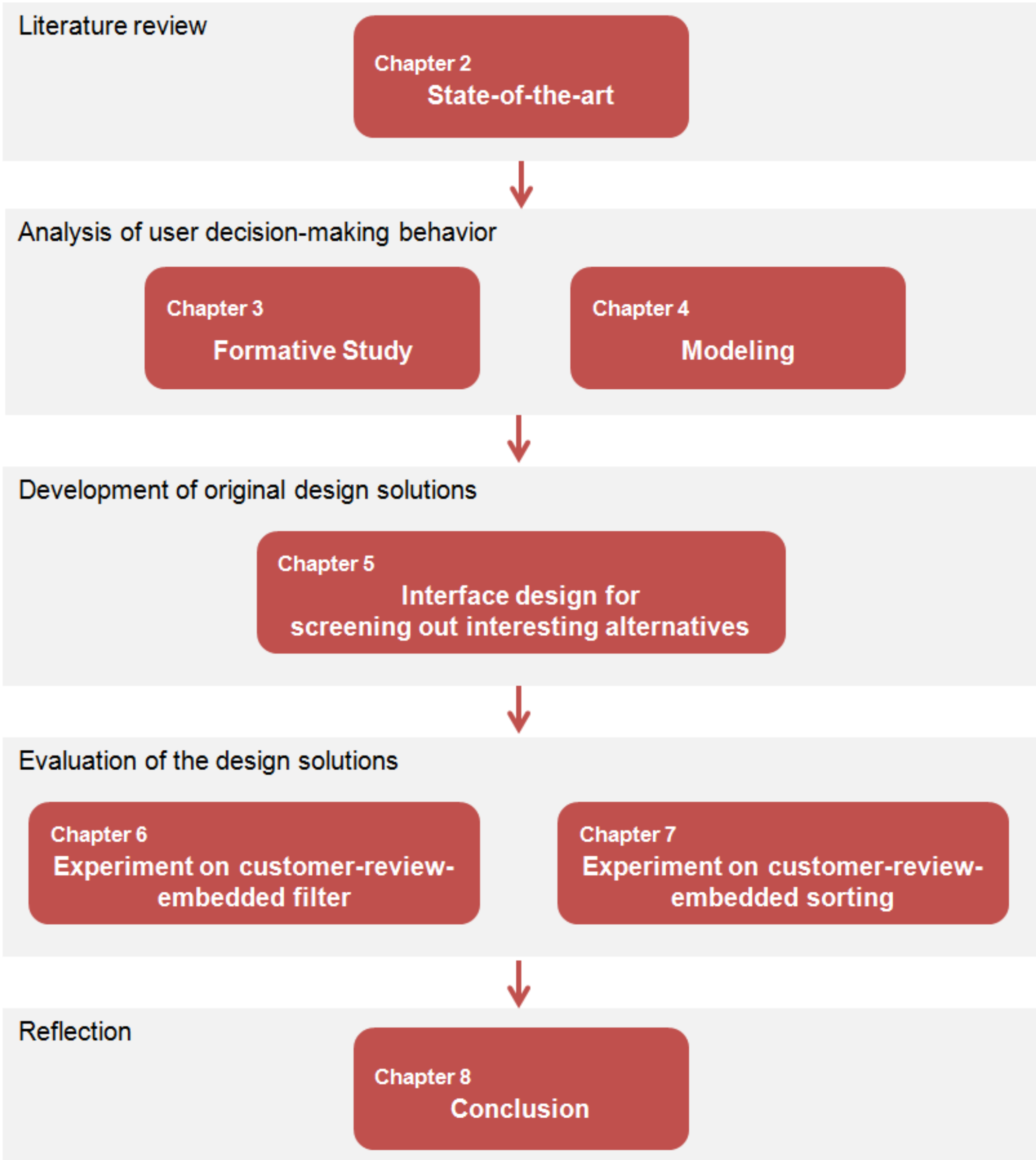


Figure 1.1 *Framework of the thesis*

Chapter two

In this chapter, we begin with the study of representations of customer reviews in current E-commerce websites and relevant research. Because online purchasing can be interpreted as a three-stage decision making process, in light of human decision-making theory, we propose that an in-depth understanding of customers' decision making behaviors serves as the foundation of effective information displays for helping users make better purchasing decisions.

Chapter three

We conduct a formative study to empirically investigate customer decision-making behavior in the three-stage decision process in the environment of online hotel booking. 50 participants are involved in performing three tasks corresponding to the three-stage decision process. To monitor what information is acquired and how the captured information is processed, we use two process tracing methods: verbal protocols and computerized process tracing. Interfaces that aggregate information from customer reviews are developed to support the three different tasks.

Chapter four

In this chapter, we begin to analyze the collected data to understand user decision making behavior in E-commerce. Firstly, we carry out single-case analysis, i.e., transcribing individual behavior by coding observed action in terms of a set of Elementary Information Processes (EIPs) and corresponding verbal protocols. Then group and compare the individual cases to identify user decision making patterns in the three-stage process, known as cross-case analysis. Finally, we create four personas that are composite models encapsulating the critical findings.

Chapter five

This chapter focuses on the interface design for screening out interesting alternatives. The design process is organized as follows. Firstly, based on the personas built in preceding chapter, we depict idealized interactions between personas and system, referred to as context scenarios, from which design requirements are extracted. Further, we introduce the status quo and limitations of interfaces in current E-commerce websites and then go on to describe in detail the steps of generating design solutions to three parts: the filter panel, sorting panel and details panel.

Chapter six

We performed a user study to compare two alternative designs for an opinion-attribute-embedded filter panel in the area of online hotel booking, called checkbox filter and slider filter. 52 participants were recruited to perform two tasks which were randomly assigned to the two compared interfaces. Both objective measures (e.g., task time) and users' subjective perceptions were collected in the experiment. It verifies the effectiveness of incorporating attributes extracted from customer reviews in the filter panel, and the modified slider filter (visualizing the distribution of an attribute via bars and the correlation among attributes via simultaneous change) achieved significantly higher user assessments in terms of perceived decision accuracy, perceived use of use, pleasantness to use and intention to return.

Chapter seven

In this chapter, we compare three alternative designs for a multi-attribute sorting panel that primarily differ in the way of eliciting relative importance of attributes, called direct assessment, indifference method, and indirect measurement. Online hotel booking is adopted as the test domain. Through analysis of objective and subjective measures, the multi-attribute sorting function embedded with opinion attributes is verified to be effective. Moreover, the direct way outperforms the indifference and indirect ways regarding perceived decision accuracy, cognitive effort, satisfaction and intention to use in an online environment.

Chapter eight

Motivated by the results of the above user studies, we have derived a set of guidelines on how to design interfaces for user decision improvement in E-commerce. Then, we point out the limitations in our research and propose several relevant research topics that we can work on in the future.

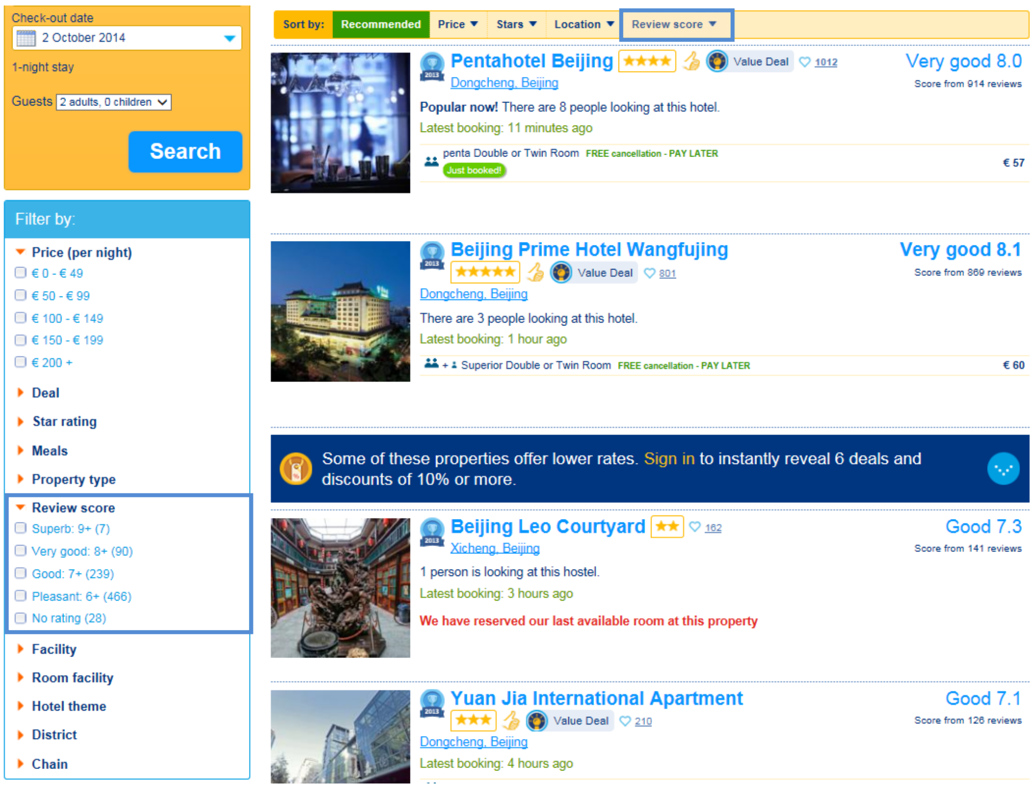
Chapter 2

State of the Art

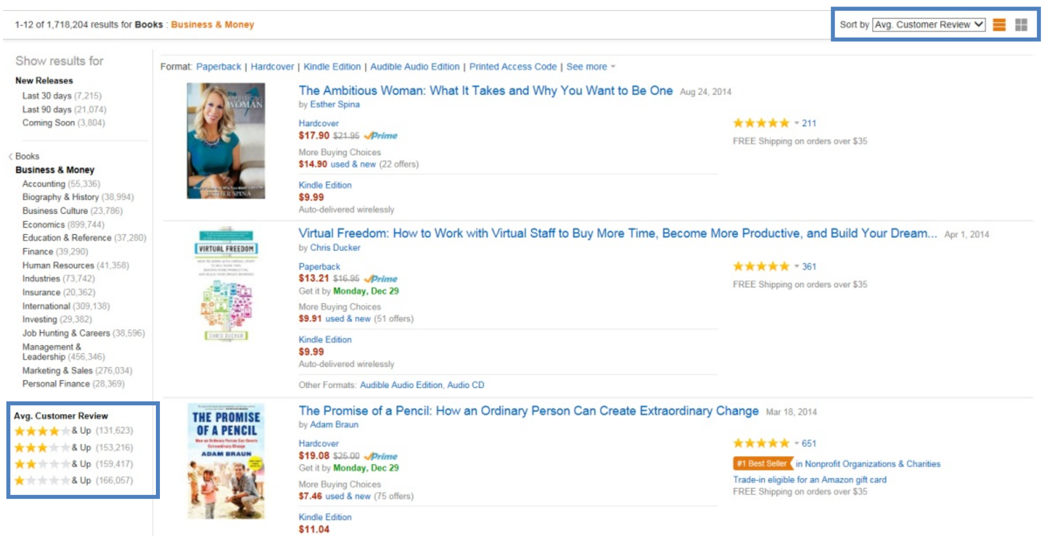
In this chapter, we begin with the study of representations of customer reviews in current E-commerce websites and relevant research. Because online purchasing can be interpreted as a three-stage decision making process, in light of human decision-making theory, we propose that an in-depth understanding of customers' decision making behaviors serves as the foundation of effective information displays for helping users make better purchasing decisions.

2.1 Customer Reviews in E-commerce Websites

Most E-commerce websites, such as Booking and Amazon, provide an overall review score for each entity. Users can filter or sort alternatives to screen out interesting choice(s) for further consideration (see the blue boxes in Figure 2.1).



(a) Booking



(b) Amazon

Figure 2.1 *FrontPage of E-commerce websites*

After entering the detail page, in addition to an overall review score of the hotel, Booking provides an average score for each attribute, e.g., cleanliness and comfort (see Figure 2.2). Prospective buyers can manually read a few reviews sorted by date to form a decision regarding the entity [31].

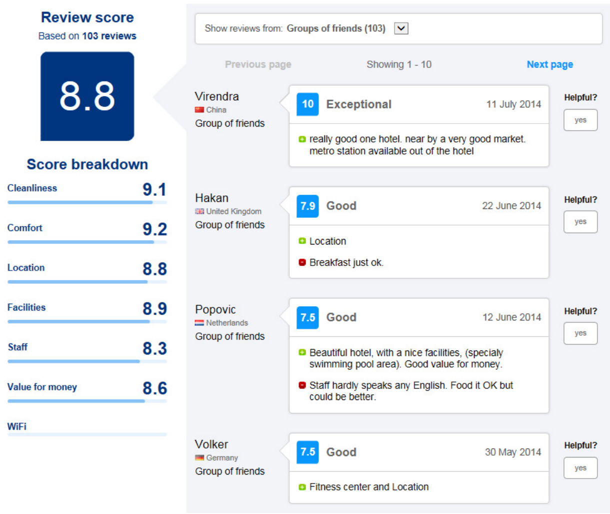


Figure 2.2 *The detail page in Booking*

Amazon allows readers to vote on the usefulness of posted reviews, facilitating users to sort customer reviews by users' votes on usefulness (see Figure 2.3). However, there are several limitations of ranking reviews in this manner: (1) the helpful votes are accumulated over a long period of time; hence, the important information in recent reviews tends to be missed [31], and (2) the amount of data can still be overwhelming and users cannot obtain an overall summary but rather must read them one by one.

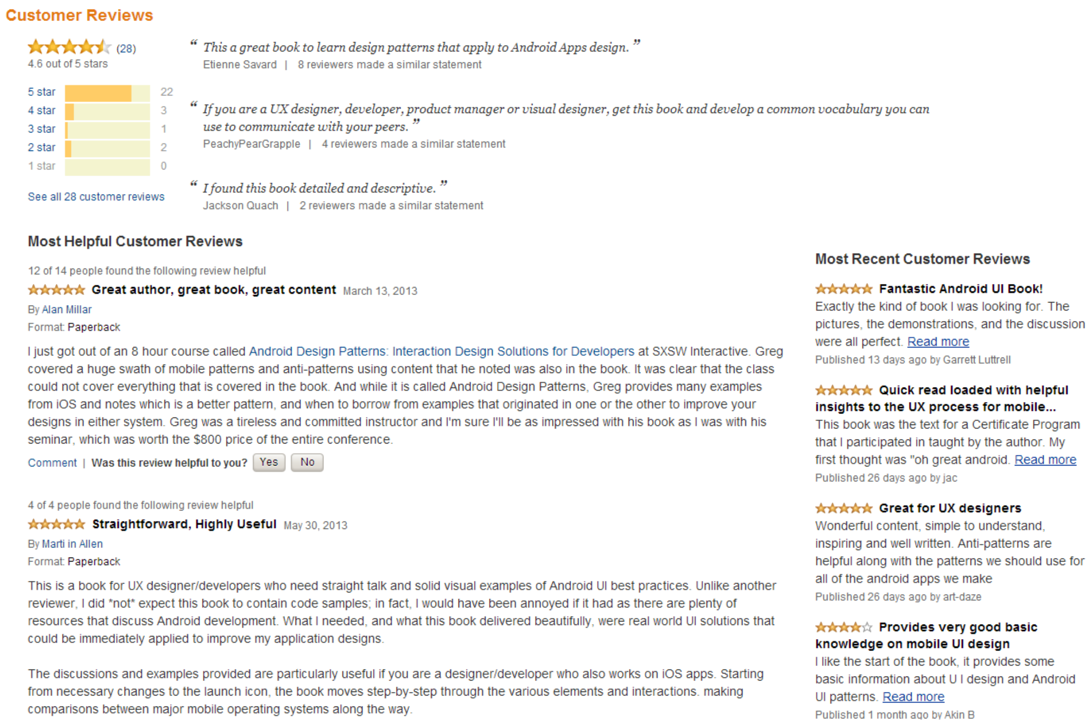


Figure 2.3 *The detail page in Amazon*

2.2 Literature Research on the Summary of Customer Reviews

The overall score does not necessarily reveal a product's quality and may provide misleading information [39]. An attribute of a product that is important for customer A and thus has an important impact on the total influence might be irrelevant for customer B [64]. Thus, people process information in an attribute-driven manner, which means, people examine the attributes of a product to evaluate whether the product fits their desire. In addition, the sentiment expressed by reviewers toward each attribute helps demonstrate the suitability of the product.

To facilitate the attribute-driven evaluation of a product, a number of systems summarize the information in user reviews by extracting features and associate sentiment toward each feature, referred to as feature-sentiment summarization. Researchers have presented different approaches for conveying that knowledge to users. Most systems describe sentiment toward each feature by an average rating or rating distribution. Because people tend to verbalize their impression with short descriptive phrases, several recent systems provide a summary of reviews using adjective-noun word pairs, which allows users to quickly explore the reviews in greater detail.

2.2.1 Bar chart

Liu *et al.* (2005) used **bar charts** to show the percentages of reviews that express positive (above x-axis) and negative (below x-axis) opinions on the features of a camera, with the bar's height representing the number of mentions [52]. It supports a side-by-side and feature-by-feature comparison of consumer opinions on competing products (Figure 2.4).

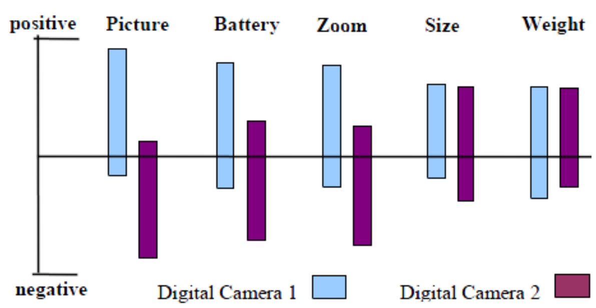


Figure 2.4 *Bar Chart customer review summary*

Carenini *et al.* (2009) developed a stacked bar chart based on Liu *et al.*'s work to visualize the opinions on each summarized feature [15]. Each bar corresponds to a polarity category (from -3 to 3) and its height represents the quantity of that opinion. It facilitates users to compare opinions on single feature, as well as the overall entity (Figure 2.5). However, the bar chart visualization is more favorable for comparison across fewer entities (around two), which might be limited in real cases. Further, these bar chart designs do not convey verbal information (i.e., opinion words associated with features).

Figure 2.5 *Stacked Bar Chart customer review summary*

2.2.2 Tree map

Carenini *et al.* (2006) also summarized user reviews in the form of a **Tree map** which represents a feature as a rectangle with nested rectangles corresponding to the descendants of the feature, so user can explore the hierarchy and details of a feature by zooming in corresponding rectangle [14]. Simultaneously, the frequency and sentiment of features are represented by varying sizes and colors (green for positive & red for negative), respectively (see Figure 2.6). If a feature has a larger size and darker red color, it suggests that a larger number of customers are not satisfied with this feature.

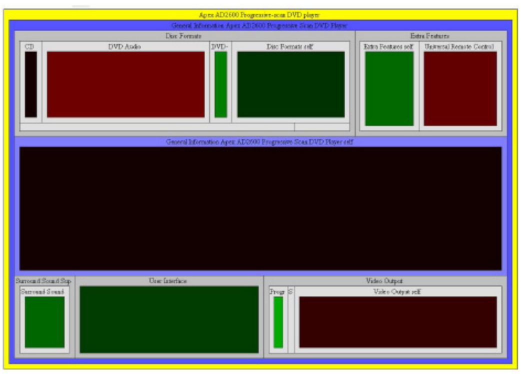


Figure 2.6 *Tree Map customer review summary*

2.2.3 Tag cloud

Although the rating gives readers a quick understanding of how much an entity is liked or disliked, it does not offer information about why that rating was given. Several recent systems use adjective-noun word pairs to summarize the sentiment (adjective) towards each feature (noun), which allows users to explore reviews in greater detail [41]. Yatani *et al.* (2011) provided a brief overview of reviews in the form of adjective-noun word pairs, using the font size and color of each word pair to represent the number of occurrences and sentiment. The user interface does not arrange word pairs in specific order but based on tag cloud. Users can quickly explore the original reviews by clicking the adjective-noun (see Figure 2.7). The results of a user study show that participants could form detailed impressions and make decisions significantly faster [93].

Figure 2.7 *Tag Cloud customer review summary*

2.2.4 Highlighting

Huang *et al.* (2013) presented sentiment toward each feature in the form of an average rating, allowing users to rank entities based on a feature rating. The system also automatically highlights sentences that are related to relevant features, to create a balance between reducing the information overload and providing the

Figure 2.8 *Highlighting customer reviews*

original context expressed by review writers (see Figure 2.8) [42].

2.2.5 Hybrid

Hu *et al.* (2013) developed a system (see Figure 2.9) that is made up of three parts: (a) a set of aspect-based blocks, (b) a system-extracted review snippet, and (c) the full text of the review. The aspect-based block includes the aspect name, adjective-noun word pairs describing the aspect (in the form of tag cloud), and a set of colored squares, each of which represents a review using color to indicate sentiment (positive, neural and negative). Referring to aspect-based blocks, users can obtain a visual summary by hovering over adjective-noun word pairs. Then, by clicking relevant highlighted squares, users can see the associated snippets and even the full context of original reviews [40]. The results of the user study showed that the majority of users correctly understood the aspects the system provided and their associated sentiment and found specific details of the product.

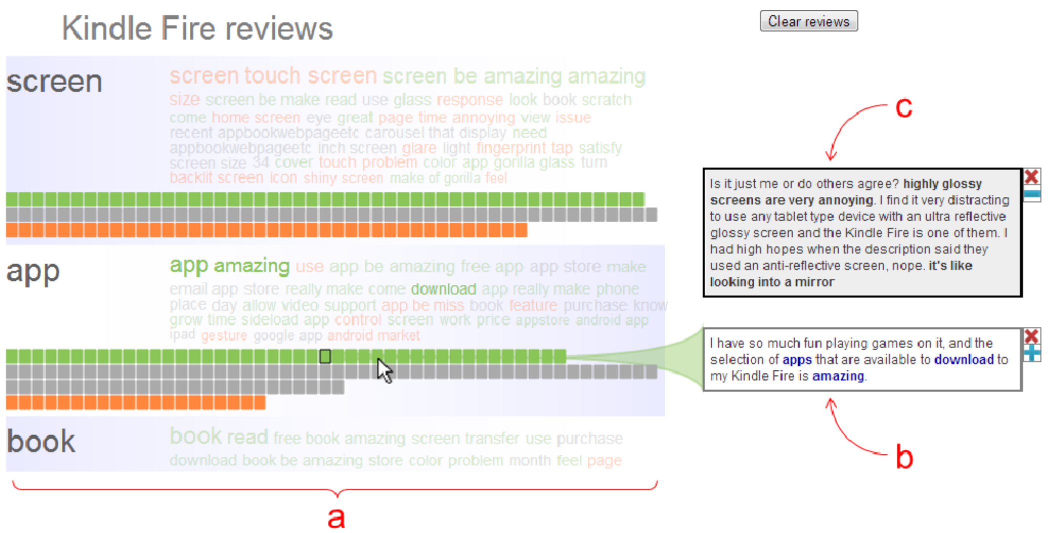


Figure 2.9 *Hybrid customer review summary*

In Table 2.1, the above works on summary of customer reviews are listed. For each, we indicate the format in which sentiment is presented (e.g., average rating, rating distribution, or adjective-noun word pairs), and ways of linking to raw reviews.

Table 2.1 *A summary of research on the representation of customer reviews*

	Summary of customer reviews	**Link to raw reviews**		**Raw reviews**
		Filter	Sorting	
Goodreads Amazon	rating distribution	rating	usefulness release date	raw reviews
Carenini *et al.* (2009)	Sentiment / Feature — 7-point scale: -3, -2, -1, 0, 1, 2, 3 summarized feature 1 → frequency (presented by bar chart) …… summarized feature n ……			
Carenini *et al.* (2006)	Sentiment / Feature — Average rating summarized feature 1: average rating is represented by color …… summarized feature n ……	summarized aspect (by zooming in)		raw reviews
Yatani (2011)	Sentiment / Feature — positive, neutral, negative adj. … … … adj. noun → frequency (presented by font size) …… noun ……	adjective-noun word pair		raw reviews
Huang *et al.* (2013)	Sentiment / Feature — Average rating summarized feature 1: average rating is represented by 5-point scale …… summarized feature n ……	summarized feature	rating score	highlight feature- related sentences
Hu *et al.* (2013)	Sentiment / Feature — positive, neutral, negative adj. … … … adj. summarized feature 1: noun, … → frequency (presented by bar chart) …… …. summarized feature n: noun Frequency (presented by font size)	adjective-noun word pair		highlight sentences that are related to relevant adjective-noun word pair

2.3 Three-stage Decision-Making Process of Online Purchasing

From the perspective of customers, online purchasing can be viewed as a decision-making process that requires a decision maker to make a choice between two or more alternatives (note that choosing nothing can be viewed as making a choice) [49]. In most conditions, customers identify the need for a product or service without specific requirements on which one to buy [61]. For example, a buyer who wants to purchase a case of wine for dinner party has not yet specified which type of wine is required; accordingly he/she needs to select interesting one(s) from a range of options that satisfy his/her desire. In this paper, we will focus on online purchasing behavior under such conditions.

Chen (2010) interpreted the above decision-making process in an E-commerce environment as a precise, three-stage decision process: (1) narrowing down available alternatives and selecting one(s) that is worth further consideration, (2) reading detailed information about the item selected in the preceding stage and deciding whether to take it as a purchase candidate, and (3) comparing several candidates and making the final choice [17]. The transition between the three stages does not follow a rigorous linear order; it is iterative in nature. However, on the whole, the process does follow an approximate sequence (see Figure 2.10).

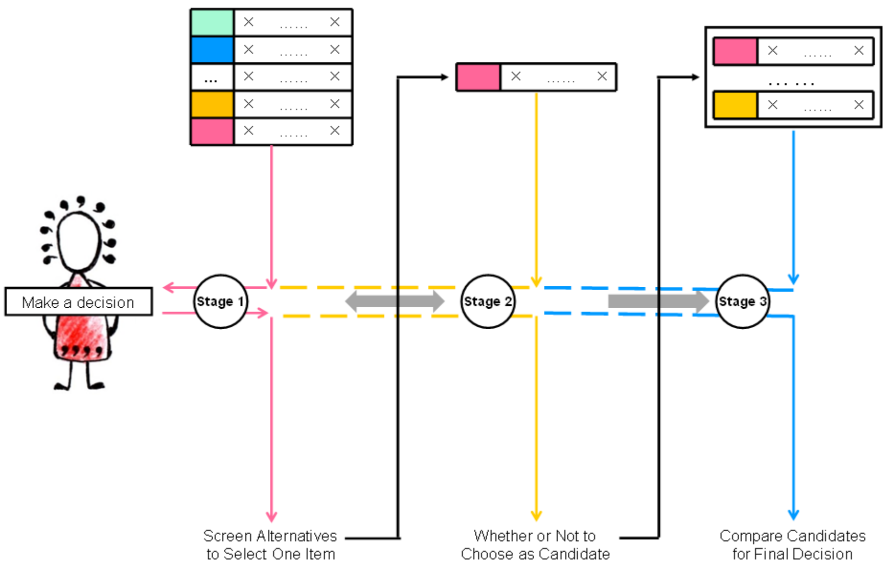


Figure 2.10 *Three-stage decision-making process of online purchasing*

We take consumer's decision-making behavior in book online stores as an example. At the start, a consumer has some initial preferences in mind, e.g., looking for a book about graphic design. A list of books in this category is presented; he/she selects interesting one(s) for in-depth reading (Stage 1). When the detailed information of a book is shown to the consumer, he/she evaluates the book and decides whether to save it as a purchase candidate, e.g., adding it to the shopping cart (Stage 2). Then, the user would return to stage 1 for another selection; the iterative cycle between stage 1 and stage 2 continues until a set of candidates is determined. At this point, the user compares the candidates and makes the final choice (Stage 3).

2.4 Human Decision Making Theory

Because online purchasing can be viewed as a decision-making process, we refer to human decision-making theory to study how to display information to improve purchasing decisions.

A decision making process can be viewed as multiple, interacting stages that fall between decision task representation and expression of a decision, including information acquisition and interpretation, and information combination leading to an evaluation (see Figure 2.11) [69]. Decision makers cycle between the two stages – information acquisition and interpretation (i.e., noticing aspects of the choice set) and information combination (i.e., deciding how to exploit those aspects) [70].

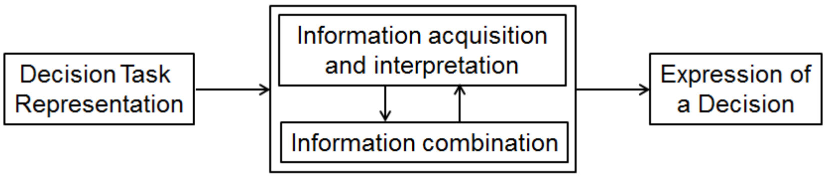


Figure 2.11 *Decision making process*

2.4.1 Classical decision theory

Classical decision theory, which is developed from normative models in economics and statistics, focuses heavily on the notion of rationality.

Decision makers are assumed to properly process all relevant information about the available alternatives and explicitly consider trade-offs among values, i.e., explicitly decide the extent to which a good value on one attribute can make up for a bad value on another, to choose an optimal alternative on the basis of an invariant strategy such as expected utility maximization.

2.4.2 Adaptive decision behavior

However, many works have demonstrated the lack of invariance in human decision behaviors in reality, which violates the prescriptions of classical decision theory. Individuals display a great deal of flexibility in making decision. The same decision maker appears to selectively adopt information and use a wide variety of strategies contingent upon decision properties (e.g., the complexity of problem, the response mode, the information display, etc.) [66, 80]. For example, decision maker tends to be more selective in the use of information and adopts non-compensatory strategies when faced with a larger number of alternatives, attributes or greater time pressure.

Given the extensive evidence on adaptive decision behavior, it is important to identify framework that can help explain why adaptive processing occurs. Two major frameworks, within which adaptive decision behavior might be understood, the perceptual framework and accuracy-effort framework are summarized [68]. The two frameworks complement each other by providing insights into different stages of the decision-making process. The perceptual approach has much to say about which information is noticed, whereas the accuracy-effort framework is more relevant for the decision strategy that decision makers utilize to process information.

1. **Perceptual framework**

   In classical decision theory, it is assumed that people notice all properties of the decision task. Unfortunately this may not always be the case. Due to limited cognitive abilities, users can not process all available information. Hence, selectivity is necessary.

   Two major determinants of which information is selected are the voluntary goal-driven aspect and involuntary perceptual aspect [9]. The voluntary goal-driven aspect means that decision makers devote more attention to information that they believe to be helpful to current goals (i.e., exemplifying the “diagnosticity” aspect in Feldman and Lynch’s accessibility-diagnosticity framework). Attention also may be involuntarily captured by perceptually salient information, referred to as the involuntary perceptual aspect (i.e., labeled as the “accessibility” aspect in accessibility-diagnosticity framework [25]).

Particular information representation can make some information more salient and accessible for decision makers. For example, Felferig *et al.* (2007) indicated that product explanation positioned at the beginning and at the end of a list of explanations are remembered more often than those in the middle of a list [27]. Mandel and Johnson (1999) provided an example that one version of online furniture selling had a coin-background and the second version had a cloud-background. Users who interacted with the first version chose significantly less expensive products than those who interacted with the second version [56]. Phenomena that are relevant to the impact of information format on information acquisition and interpretation are listed in Table 2.2.

Table 2.2 *Impact of information display on information acquisition*

Effect	Content
Context effect[81]	Decision is always made depending on the context in which alternatives are compared. Additional irrelevant items (e.g. asymmetric dominated one) significantly trigger change in decision, which is against optimal decision model.
primacy effect [27]	Information units at the beginning and at the end of a list are analyzed more often than information units in the middle of a list.
Priming [56]	If special decision attribute is made more accessible in memory (e.g. emphasizing through webpage background), which can manipulate attribute weights and ultimately influence decision.
Framing effect [51,89]	The different way in which decision alternatives are represented (e.g. positive/negative framing) causes a shift in preference for identical alternatives, i.e. risk aversion in positive and risk seeking in negative framing condition.

2. Accuracy-effort framework

Rather than one invariant approach to solving choice problems, the same individual appears to utilize a variety of strategies when making decisions. The accuracy and effort characteristics are different across strategies for a given decision environment and different across environment for a given strategy [9]. Strategies yielding more accurate decision are often more effortful, and strategies that need less effort tend to yield lower levels of accuracy. Figure 2.12, developed by Bettman *et al.*, shows the relative accuracy (normalized by the accuracy of weighted adding) and effort (in total Elementary Information Process) for five strategies: weighted adding (WADD), equal weight (EQW), lexicographic (LEX), eliminate by aspect (EBA), and random choice (RC) [9].

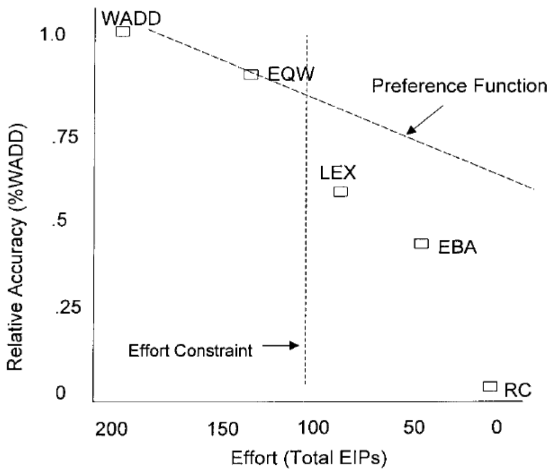


Figure 2.12 The *Accuracy-Effort framework of decision strategies*

A decision maker selects a strategy as a result of multiple, potentially conflicting goals (i.e., maximizing accuracy and minimizing effort) depending on the relative weight (importance) he/she places on the goal of making an accurate decision versus saving cognitive effort, denoted as the preference function in Figure 2.12. In other words, the choice between decision strategies can be assumed as a trade-off between benefits (in terms of decision accuracy) and costs (in terms of cognitive effort) [34]. Some of the most commonly used strategies are listed in Table 2.3.

Table 2.3 *Decision Strategies*

Strategy	Content
Weighted Adding	Assess the importance of each attribute, assign a subjective value to each attribute level, then multiply each attribute's subjective value by its weight and sum all attributes to obtain an overall value for each alternative.
Lexicographic	Select the alternative with the best value on the most important attribute.
Satisfying	Alternatives are considered sequentially, if values of all attributes of an alternative meet predetermined level, then the alternative is selected instead of comparing all alternatives.
Elimination-by-a spect	Eliminate alternatives that do not meet the cutoff value for the most important attribute, the elimination process is repeated for the second most important attribute.
Majority of confirming dimensions	Two alternatives are compared in terms of the value of each attribute, the alternative with a majority of better attribute values is retained, and then the retained alternative is compared to the next alternative.
Combination of strategies	Decision makers also use combinations of strategies. One frequently observed combination is an initial use of EBA to reduce to several alternatives, which are then analyzed by more compensatory strategy.

2.4.3 Effects of information display on decision behavior

The adaptive nature of decision making behavior provides the insight that information display can impact not only information acquisition and interpretation (i.e., what information is noticed) but also information combination (i.e., which decision strategy is utilized to process information), leading to higher/lower decision accuracy and less/more cognitive effort.

In the following, we elaborate the effects of information display on decision accuracy and effort at each stage and derive design implications to help decision makers increase the accuracy of choices and reduce the effort required by changing the information display.

1. Information acquisition and interpretation

Firstly, referring to the completeness of an information display, an insufficient information display can blind a decision maker to myopic, uninformed decision [69].

Moreover, merely presenting all necessary information to decision makers is not enough. An information display may make it difficult for decision makers to assess some properties of decision information. MacGregor and Slovic, (1986) provided evidence that decision makers tend to ignore important information simply because the most salient information is not diagnostic or important for decision makers, leading to bias in information selectivity. When more important information is more salient in display features, the judgmental accuracy is greater [55]. In addition, making important information more salient can help users pay enough attention to the subset of information that is crucial, which can mean that a decision maker is not overloaded with too much information. Accordingly, it is important to consider the match between the relative importance of information for decision makers and the salience of information display, in other words, making important information more salient. Otherwise potential problems in information acquisition and interpretation will arise.

Implication: based on the impact of information displays on decision accuracy and effort during information acquisition, we learn that making better decision is guaranteed by both a completeness of information and a match between the importance of information for decision makers and the salience of the information display, i.e., providing important information using formats that make it salient and easy to process. Hence, it is crucial to learn what kind(s) of information decision makers are inclined to seek, and which kind(s) of information are particularly important for decision makers when making decisions.

2. Information combination

The effort associated with a particular operation would certainly be expected to vary as the information format varies [83]. Some Information displays can exacerbate difficulties individuals have in executing certain decision strategy. Coupey (1994) provided an example considering how difficult it would be to compare attribute values across alternatives if information were represented in different units within an attribute (e.g., warranty in weeks, months or years) or information on an attribute appeared in different columns for each alternative [20]. Russo (1977, 1991) showed that sorting the available brands by increasing unit-price helped consumers make more accurate decisions with less effort compared with the unit price displayed on separate tags under each item because it made price comparison easier for shoppers [74,78]. Hence, decision effort can be reduced by improving the congruence between the format and organization of information and the way in which users process information and make decisions (known as reactive approach).

Because different information formats can make some forms of processing easier and less effortful than others, processing encouraged by the format will be more likely. If the encouraged form of processing is not an improper one for the particular environment, the effect of the information display will be lessened, e.g., resulting in potential accuracy losses. At the information combination stage, decision makers often avoid making explicit trade-offs, relying instead on an array of non-compensatory decision strategies. Todd and Benbasat (1991) provided evidence that decision maker can be directed toward the use of compensatory processing by reducing the relative effort needed to execute certain operations (e.g., calculation of differences between options on various attributes) [87], referred to as proactive approach.

Implication: we are able to (1) understand how decision makers are currently processing information and arrange information displays to make that information processing easier, and (2) if the decision strategy adopted is not efficient or proper, reduce the effort needed to execute certain operations that lead to more accurate decisions [70].

To recap, to explore what information should be represented and how to represent information to augment decision accuracy while reducing effort required, our primary task is to investigate consumer decision-making behavior: (1) which kind(s) of information customers are inclined to seek, which kind(s) of information are particularly crucial for decision makers when making decisions, and (2) which kind(s) of decision strategies do they utilize to process the information.

Chapter 3

Formative Study

To explore what information should be represented and how to represent information to enhance users' decisions in E-commerce (i.e., reducing the effort to make purchasing decision while augmenting the decision accuracy), we take online hotel booking as an example to investigate customer decision-making behavior in the three-stage decision-making process. 50 participants are involved in performing three tasks corresponding to the three stages. To monitor what information is acquired and how the captured information is processed, we use two process tracing methods: verbal protocol and computerized process tracing. Interfaces that aggregate information from customer reviews are developed to support the three different tasks.

3.1 Research Questions

3.1.1 Analysis of the decision strategy that decision makers adopt

A central distinction among strategies is the extent to which they make trade-offs among attributes. Decision strategies (such as Weighted Additive) that explicitly consider trade-offs are called compensatory strategies, whereas strategies (like Lexicographic) that do not make trade-offs are called non-compensatory strategies. We first study the following research question:

-*RQ1*: which kind(s) of decision strategies do customers adopt to process information, compensatory strategies or non-compensatory strategies?

3.1.2 Analysis of the information that decision makers seek

In the E-commerce environment, each entity is described by diverse information. In general, the information can be classified into two types – static features (such as price and specifications) and customer reviews.

-*RQ2:* which kind(s) of information do decision makers seek, static features or/and customer reviews?

To further explore the role of each type of information in more detail, we investigate the following research question:

-*RQ3:* which kind(s) of information in static features and customer reviews do decision makers seek?

The format in which the sentiment corresponding to attributes extracted from customer reviews is presented can also be different. Numerical values (e.g., the average rating and number of reviews) provide an easy proxy for opinions, whereas verbal values (e.g., adjective-noun word pairs) provide reasons underlying the scores.

-*RQ4:* which kind(s) of values do decision makers refer to concerning the sentiment of attributes extracted from customer reviews, numerical or/and verbal?

3.2 Tasks

To examine decision-making behavior in an E-commerce environment, we took online hotel booking as the test domain for two reasons. First, it is feasible to recruit appropriate and sufficient subjects to participate in the study. Most of the participants are students at university, who would be familiar with or interested in hotel booking for vacation. Second, the hotel domain contains abundant online customer reviews that are written with multiple attributes in mind (e.g. cleanliness and service).

All hotel information and corresponding customer reviews used in the formative study were crawled from Tripadvisor.com in May, 2014.

The same individual has different decision-making behaviors while interacting with E-commerce, for example, a choice task (e.g., selecting one from a set of alternatives) offers a more selective decision process; however, when making a judgment (e.g., whether to save as a candidate), decision makers tend to use a less selective, alternative-based process to obtain an overall evaluation of alternative. We design three decision tasks corresponding to the three interaction stages.

Task 1: imagine that you will have a trip to Beijing with your friends in the summer holiday and need to book a hostel online. The top 10 Beijing Bed and Breakfast (B&B) are presented. Please choose interesting one(s) for further consideration.

Task 2: please read detailed information of the hostel you selected in the preceding task and decide whether to save it as candidate, e.g. adding it to the shopping cart.

Task 3: compare the candidates you have selected to choose one as the final choice.

3.3 Research Methods

Two process tracing methods that have proven especially valuable in monitoring what information individuals acquire and how that information is processed to make a choice are verbal protocols and information acquisition techniques. Eye tracking, computerized process tracing tools (CPT), and information boards [1] can be used to observe Information acquisition behavior. Although they are fairly straightforward in data collection, internal cognitive processes are not directly observed. Investigators must infer the underlying cognitive strategy from the information acquisition data, and sometimes those inferences are not correct. In contrast, verbal protocols have been used to measure information acquisition and processing stages directly [30]. From the concurrent verbal protocol, experts can make it clear what information is processed in a certain manner [67]. However, the most important point about verbal protocols is that they are difficult to analyze formally [76]. Thus, verbal protocols and information acquisition techniques are concurrently employed to complement each other to avoid misinterpretation.

3.3.1 Verbal protocols

Subjects are simply asked to give continuous verbal reports, "to think aloud", while performing the decision task. It is a straightforward method for obtaining process data, which treats the verbal protocol as the ongoing behavior of the subject.

3.3.2 Information acquisition techniques

Eye tracking: Russo and his associates used a photoelectric sensing device and a computer for recording and analyzing eye fixations and movement to study decision behavior [77].

Computerized process tracing: this is done by setting up a decision task so that all relevant information is hidden in boxes until a subject moves mouse to click a box. At that time, the box opens and the information is revealed. Only one item of information is visible at a time [70].

Information board: this consists of a matrix of envelopes on a poster board. The subject can pull a card out of the appropriate envelope to obtain information [66].

These procedures eliminate the possibility of the subject acquiring information from peripheral vision, which might be the case in a more normal visual information environment.

As to different information acquisition techniques, eye tracking requires 0.2-0.4s per acquisition [75], whereas computerized process tracing and information boards take 1-2s and 15-20s, respectively, to acquire the same piece of information [12, 76]. In other words, the process underlying eye tracking is the most similar to real-world process. Compared with information board, computerized process tracing outperforms in terms of reducing the effort associated with the physical acquisition of information through the use of mouse. Lohse and Johnson (1996) found that CPT increases the amount of time needed to acquire information compared with eye tracking. As a result, subjects using CPT tend to have a fewer number of fixations and use a more alternative-based search pattern than those associated with eye tracking [53]. We do not have eye tracking equipment and the differences in the decision process caused by the two information acquisition techniques are not likely to substantively influence our research questions. For example, the nature of our research conclusion on which information users are inclined to seek will not be changed because all information is equally accessible. We therefore employ verbal protocols and computerized process tracing tool to examine individuals' decision making behavior.

With the aid of the two research methods, we can collect an array of data, comprising recordings of participants' interaction process with the system and comments behind each behavior.

3.4 System Design for the User Study

To trace information acquisition behavior, we design the system with all relevant information hidden in boxes until the subject moves mouse to click them.

3.4.1 Review summary

The systems in related research summarized customer reviews based on features and associate sentiments. The feature-sentiment summarization has proven to be an effective way to help users digest the massive quantity of customer reviews [40]. However, the variances in other elements of customer reviews, for example the temporal dimension, are not taken into account. Customer reviews of an entity may change overtime, such as "*if the restaurant improves service, the recent reviews indicate that as opposed to old reviews in which people complain about the service*" [93]. People use customer reviews differently depending on their helpful votes and reviewer prestige; usually customers prefer to read reviews with a high number of helpful votes or written by knowledgeable reviewers, which to some extent guarantees higher review quality [60, 65].

Hence, each review can be viewed as a tuple including two parts: metadata and review content. The metadata contain three elements {post date, reviewer, helpfulness}, where the post date is the date when the review is commented, the reviewer is the customer who posted the review, and the usefulness is the number of customers who found the review helpful. Content includes two elements {feature, sentiment}, where feature is frequently mentioned aspects of a product/service and sentiment is the positive or negative description toward each feature, as shown in Figure 3.1.

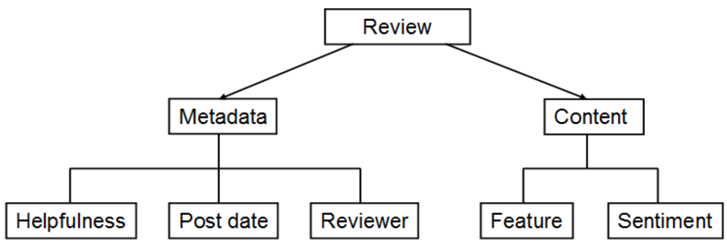


Figure 3.1 *The structure of customer reviews*

In our study, we provide multiple-level exploration of customer reviews (see Figure 3.2), which incorporates three more parameters into our review summary: post date, usefulness and reviewer, in addition to feature and associate sentiment. More concretely, reviewers are classified into 3 types: novice, junior contributor, and senior contributor in terms of his/ her prestige, i.e., the number of reviews he/she has posted.

Feature: because people process information in an attribute-driven manner in the absence of actual products (e.g., online shopping) [48], we extract customer reviews from several main aspects {location, service, value, cleanliness} in addition to an overall summary, to facilitate attribute-driven evaluation of products.

Sentiment: the average score does not necessarily reveal the product's true quality [39]. We present both the average rating and rating distribution toward each attribute, which give users a quick understanding of how many customers liked or disliked the entity.

Moreover, adjective-noun word pairs are presented to offer information about why that rating was given and allow users to quickly explore the reviews in greater detail [93]. We use the size and font color of each word pair to represent the occurrence frequency and sentiment, respectively.

Distribution: above and beyond the feature-sentiment summary, users can examine the time, usefulness and reviewer distribution of a selected subset of reviews. The distribution is shown in the form of bar charts, with bar's height representing the number of reviews. For example, when users learn that there are 23 5-star reviews for location, they can inspect their usefulness distribution (from 0 to 1 vote), time distribution (from Jan. 2011 to May 2014), and reviewer distribution (the majority of the 23 reviews were commented by junior contributors), as shown in the red boxes of Figure 3.2.

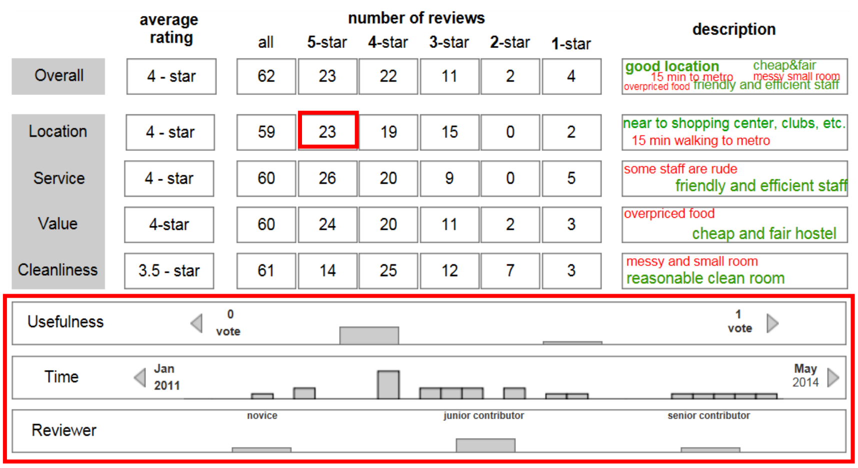


Figure 3.2 *Screenshot of customer review summary with all boxes uncovered*

3.4.2 Interface for choosing interesting hotels from a list

There is evidence that subjects tend to use non-compensatory strategies when faced with complex (multi-alternative) decision tasks, such as elimination-by-aspects, to avoid trade-offs among attributes [88]. Thus, in the interface for task one (see Figure 3.3), in addition to information on each hotel, namely, static features and review summary, there are sorting and filtering to facilitate the lexicographic heuristic (selecting the alternative with the best value on the most important attribute) and elimination-by-aspect heuristic (eliminating all alternatives with values for an attribute below a cut-off). In addition, attributes extracted from customer reviews are added to complement conventional attributes in sorting and filtering.

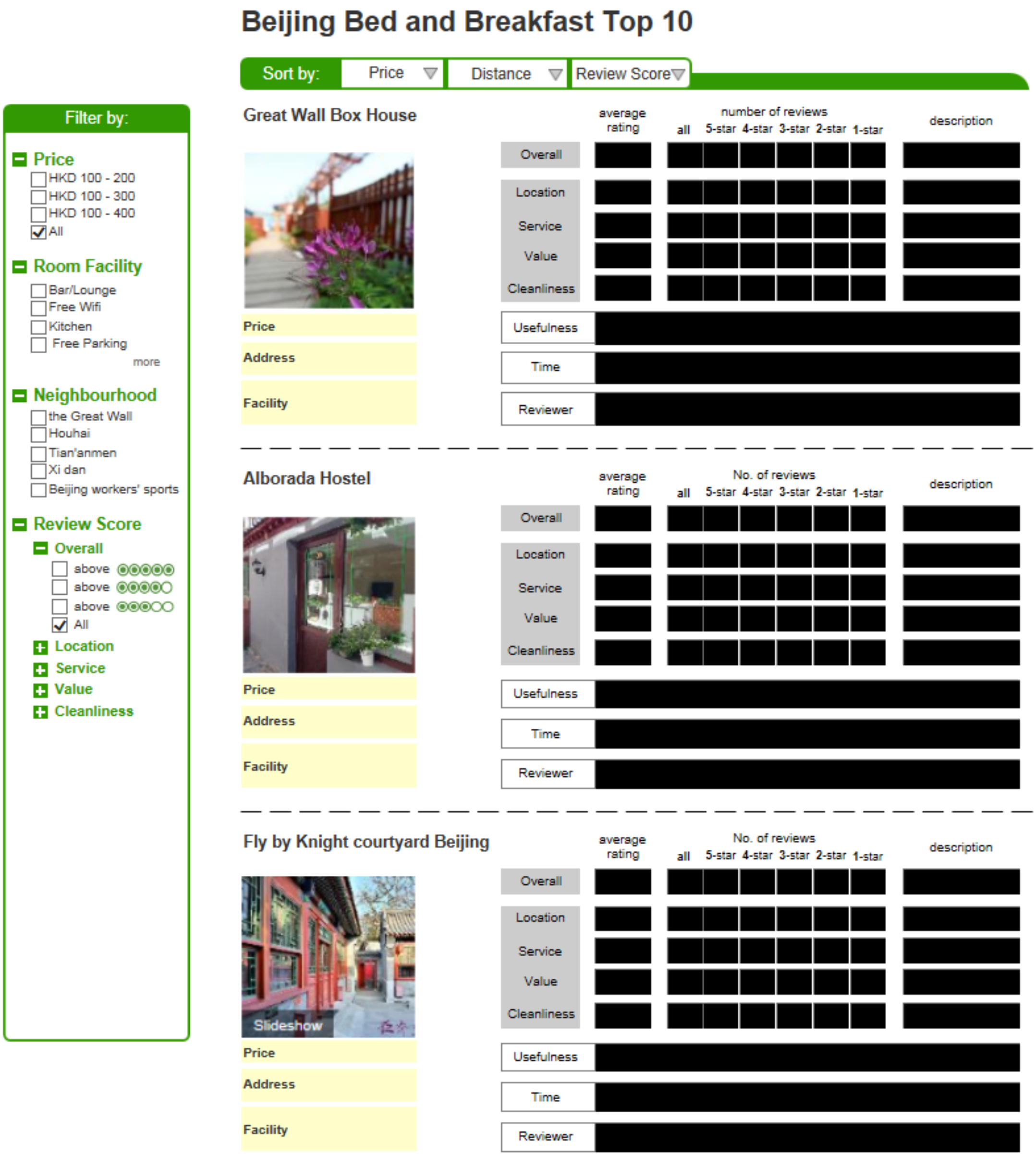


Figure 3.3 *Screenshot of the Interface for task one*

3.4.3 Interface for evaluating hotels in detail

Sinha and Swearingen used a music system as an example and noted that the information that comes into play in user decision making during this stage can be classified into three categories: basic item information, social opinion and item sample [82, 85]. In the hotel system, hotel name, price, address and facilities are included as basic information. Social opinion is customer reviews from a large community of travelers. In addition to basic information and social opinion, we use traveler photos as the item sample to enable hotel preview (see Figure 3.4).

Huang *et al.* (2013) found that users prefer original context rather than purely summarized feature descriptions through a comparison between adjective-noun pair summarizations and review highlighting [42]. Therefore, in this interface, user can explore the raw reviews (i.e., the textual comments) in addition to review summaries, for example, filtering raw reviews from a large list by feature/sentiment and sorting reviews by date/usefulness.

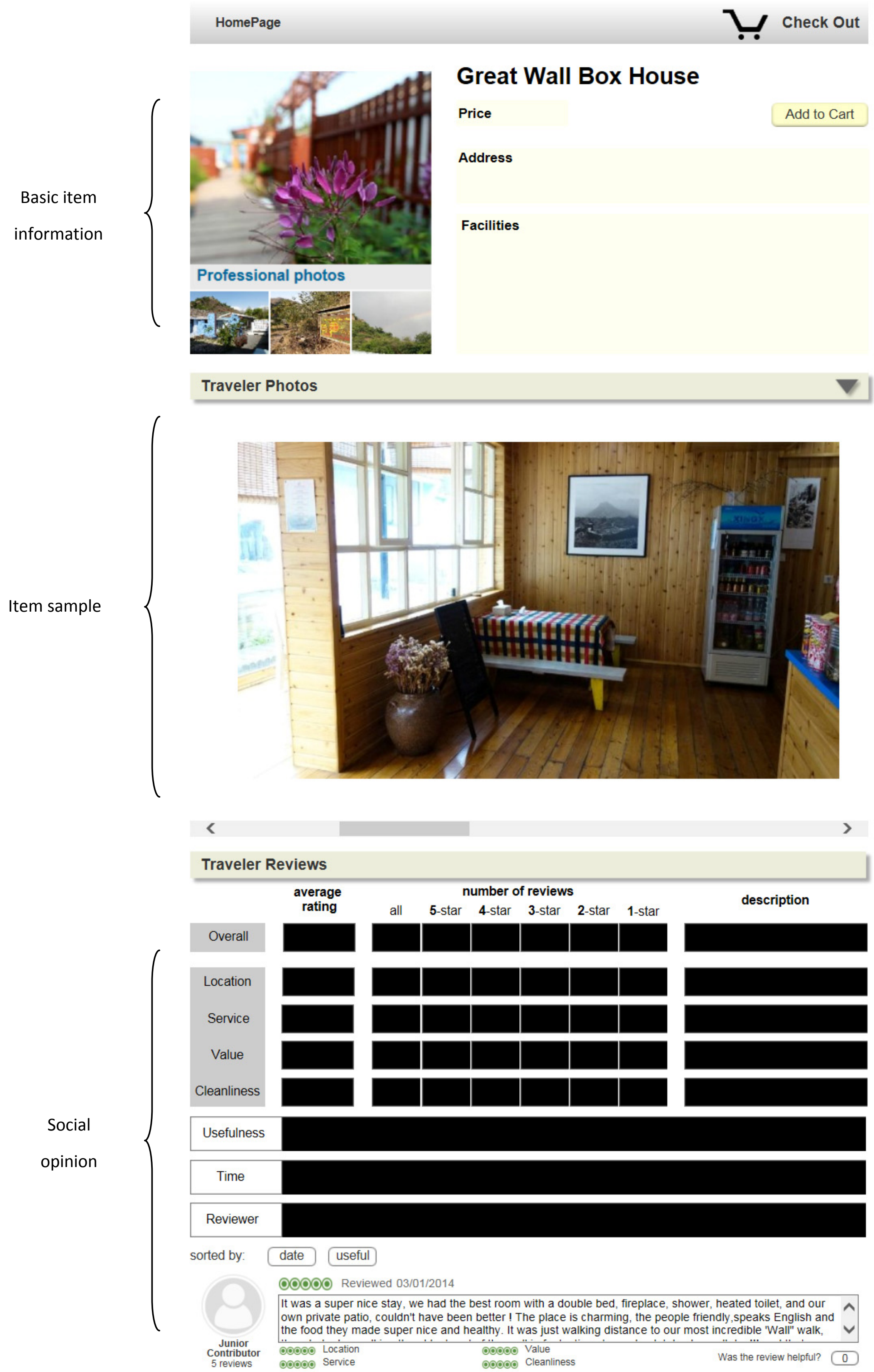


Figure 3.4 *Screenshot of the interface for task two*

3.4.4 Interface for comparing candidates to confirm the final choice

`At this stage, the shopping cart provides a comparison matrix in the form of alternatives (columns) and attributes (rows), with which users can perform side-by-side comparison between multiple products. This method has been demonstrated to improve decision quality compared with its absence [35]. Moreover, the attributes (rows) are not limited to brief static features (e.g., hotel price, address and facility); the {feature-sentiment} pairs extracted from customer reviews are embedded to complement {static feature- value} pairs, to facilitate users' comparison. Figure 3.5 provides a comparison matrix of two alternatives.

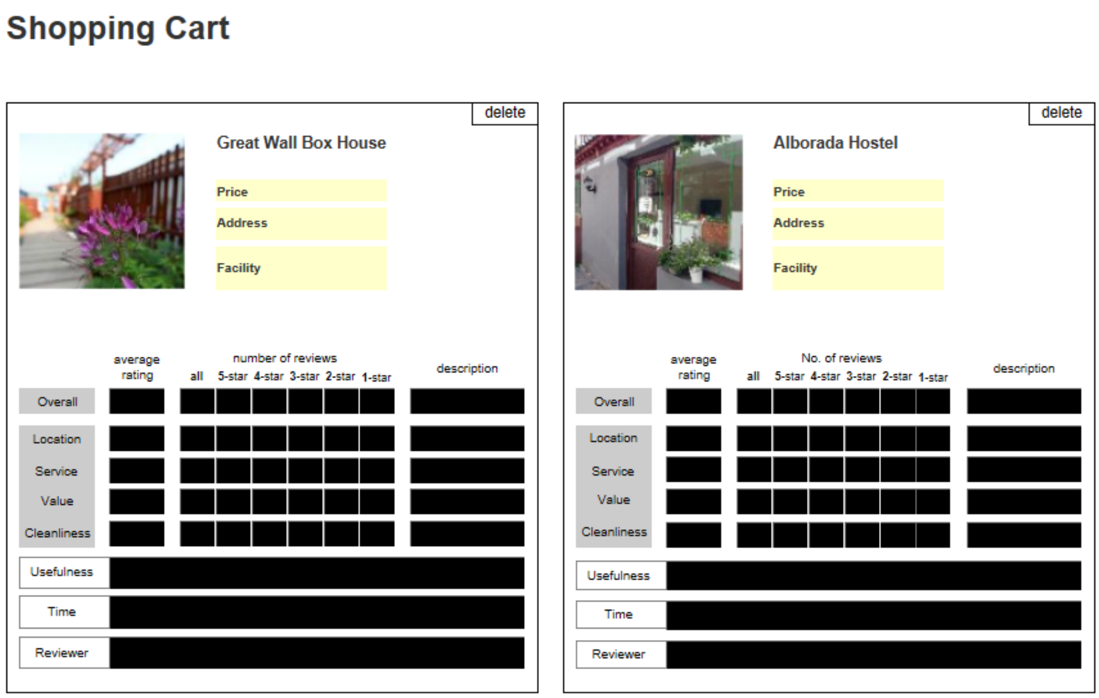


Figure 3.5 *Screenshot of the interface for task three*

3.5 Procedure

The main procedure for the formative study can be divided into three steps.

Step one: At the beginning of the study, each participant was required to fill in his/her personal background and E-commerce experience (see Appendix 1 for detail). Then, we gave a brief introduction on the experiment and explained the interfaces to participants.

For instructional purpose, all boxes within a given screen were uncovered, displaying their normally hidden contents to familiarize subjects with the experimental stimuli.

Step two: before conducting the task, we allowed the participants to use the interfaces and asked participants several testing questions (e.g., which box to click if you want to learn the average rating of location) to make sure that they understand the hidden content of each black box and would not randomly click.

Step three: we let participants perform the three tasks. While they performed the set of tasks, participants were asked to think aloud and verbalize their thinking processes. The entire session was recorded by screen recording software; all the mouse click events, the time participants spent and verbal protocols were recorded for further analysis.

3.6 Participants

50 participants were recruited to take part in the experiment. They are students at Hong Kong Baptist University pursuing Bachelor, Master or PhD degrees, from different departments, such as Computer Science, Chemistry, Education and Management. Table 3.1 lists the demographic profiles of these participants. In the pre-study questionnaire, they specified their frequency of Internet use (on average 4.96 ‘daily/almost daily’, S.D. =.23), e-commerce shopping experience (on average 3.5 ‘1-3 times a month’, S.D. =.56), and online hotel booking experience (on average 2.42 ‘1-3 times’, S.D. =.45). Thus, most of them are frequent E-commerce users and target customers of online hotel booking.

Table 3.1 *Demographic profile of participants in the formative study*

Gender:	female (26); male (24)
Average age	21-30 (46); >30 (4)
Major	Computers, chemistry, management, design, etc.
Internet usage	4.96 (daily/almost daily)
E-commerce shopping experience	3.5 (1-3 times a month)
Online hotel booking experience	2.42 (1-3 times)

The average scores for ‘internet usage’ and ’ e-commerce shopping experience’ are given on a five-point Likert scale from 1, ”least frequent” to 5, ”very frequent”, and ‘online hotel booking experience’ is given on a three-point Likert scale from 1, “never” to 3 “more than 3 times”.

Chapter 4

Modeling

After the formative study, we had dozens of pages of notes, photos and many hours of video recordings; therefore, we had to condense and massage them to make sense of data, i.e., understand human behavior patterns. The entire process is known as modeling.

Figure 4.1 provides an overview of the modeling phase, which begins with understanding what we have heard and observed (i.e., single case study); once this task is accomplished, we group and compare individual cases to identify patterns and relationships and interpret what those patterns and relationships mean (i.e., cross-case analysis); finally it is essential to develop a set of personas. The process is iterative, rather than strictly linear, but does follow an approximate sequence on the whole.

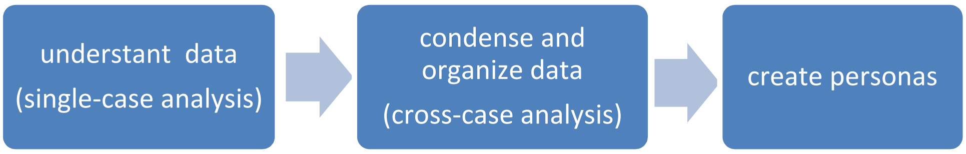


Figure 4.1 *Modeling process*

4.1 Single-case Analysis

Single-case analysis focuses on understanding what we heard and observed with one individual at a time.

4.1.1 Coding

To begin the single-case analysis of decision making behavior, we assign a category to each observed behavior, referred to as coding. Bettman *et al.* (1990) proposed that decision strategies can be decomposed into more detailed components, i.e., a set of Elementary Information Processes (EIPs) [8], as shown in Table 4.1. In turn, based on a specific collection and sequence of EIPs, the decision strategy participants adopted can be inferred.

Table 4.1 *Elementary information processes*

Read	read an alternative's value on an attribute into STM (short-term memory)
Compare	compare two alternatives on an attribute
Difference	calculate the size of the difference of two alternatives for an attribute
Add	add the values of an attribute
Product	weight one value by another (multiply)
Eliminate	remove an alternative or attribute from consideration
Move	go to next element of external environment
Choose	announce the preferred alternative and stop the process

Thus, we transcribe individual cases by coding each observed behavior in terms of Elementary Information Processes (EIPs) and corresponding verbal protocols (i.e., supporting commentary). An example of a formally coded data transcript in our study is recorded in Table 4.2. Because move and read are perfectly correlated, we will only consider read in our experiment. The action of acquiring information from previous experience is shown in brackets, e.g., deciding which attribute is the most important according to previous experience.

Table 4.2 *A sample of formally coded data transcript*

Verbal protocols	Elementary Information Processes
Stage 1: Screening out interesting alternatives	
1. *I think accessibility to attractions is the most important, so I would like to choose the hotel nearest to "houhai".*	(acquire the most important attribute) **Compare** hotels on their distance to destination
2. *Then, I will go to the detail page of the first hotel.*	**Choose**
Stage 2 Evaluating alternatives in detail	
3. *From the photo, I learn it is a typical courtyard, which is my favorite.*	**Read** the photos and **Compare** with the aspiration
4. *Then, I am concerned about the customer review on cleanliness. The average rating is high, and travelers said it is neat and clean, although slightly small.*	**Read** the sentiment of cleanliness and **Compare** with the aspiration
5. *I will save it to the shopping cart.*	**Choose**
Stage 3 Comparing candidates to confirm the final choice	
6. *Ok, first I see the price of the two hotels One is HKD195, the other Is HKD275.*	**Read**, and **Compare** the prices across alternatives
7. *The difference is HKD 80.*	Calculate the **Difference** between alternatives on the attribute
8. *Then, I compare the rating of cleanliness. One is 4-star, and the other is 5-star.*	**Read** and **Compare** the values on another attribute across alternatives
9. *The expensive hotel is given higher rating on cleanliness.*	Calculate the **Difference** between the values on the attribute
10. *I think the extra cost (HKD 80) is worth the higher score on cleanliness.*	**Add**
11. *My final choice is determined.*	**Choose**

4.1.2 The reliability of coding

Neuendorf (2002) noted [62], "given that a goal of content analysis is to identify and record relatively objective characteristics of messages, reliability is paramount. Without the establishment of reliability, content analysis measures are useless". To guarantee the reliability of coding, in addition to transcribing all individual cases myself, I invited another PhD candidate in Design to transcribe a sample set that consists of 15 randomly selected cases. Our work was done independently. Then, based on the comparison of our separate coding for the 15 cases, the reliability of the coding can be calculated.

For the index, to calculate reliability, there are several recommendations for Cohen's Kappa. Dewey (1983) argued that despite its drawbacks, Kappa should still be "the measure of choice" [22], and this index appears to be commonly used in research that involves the coding of behavior [4]. Hence, we calculated Cohen's Kappa. For example, with respect to the variable that measures whether users eliminate hotels by price, there are two sets of judgments from coder1 and coder2, where "1" means "yes" and "0" means "no" (see Figure 4.2). The measure of agreement Kappa is .857, $p<.05$. In other words, the transcriptions of the two coders are correlated for this variable.

	name	price_coder1	price_coder2
1	鲍青	1	1
2	蔡震尧	0	0
3	周燕	0	1
4	张扶桑	0	0
5	张同学	1	1
6	俞川	1	1
7	余璐	0	0
8	尹昭	0	0
9	吴丽晶	1	1
10	王峰	0	0
11	王剑涛	0	0
12	彭勤牧	0	0
13	娄坚	0	0
14	梁凤凤	0	0
15	兰伟霞	1	1

Figure 4.2 *Screenshot of the comparison of coding*

The measure of agreement of Kappa for each variable is above 0.7, which is appropriate in exploratory studies [54], suggesting a good level of consistency between the two coders. Disagreements in the coding were resolved by discussion.

4.2 Cross-case Analysis

Cross-case analysis involves grouping and comparing the individual cases to identify behavior patterns and formulate explanations for why respondents behaved as they did and how various aspect of their behaviors are related.

4.2.1 Stage 1: Screening Out Interesting Alternatives

In stage one, users narrow down available alternatives and select interesting items that are worth for further consideration.

4.2.1.1 Decision strategy at stage one

The decision strategies users adopted at stage one can be interpreted as four categories: (1) Eliminate-by-aspect (EBA), (2) Lexicographic (LEX), (3) Eliminate-by-aspect plus Lexicographic (EBA+LEX), and (4) Eliminate-by-aspect plus Additive difference (EBA+ADDIF).

1. **Eliminate-by-aspect:** participants retrieve the cut-off value for an attribute. Then, all alternatives with values for that attribute below the cut-off are eliminated. The process continues with the second attribute, and then the third, until a smaller set of alternatives remains. They go to the detail page of the remaining options one by one. A sample of the EBA process is recorded in Table 4.3.

Table 4.3 An *example of Eliminate-by-aspect*

Verbal protocols	Elementary Information Processes
1. *There should be free Wi-Fi in the hotel, so I eliminated all hotels without Wi-Fi.*	(Acquire the cut-off value for facility) **Eliminate** hotels that do not have Wi-Fi
2. *And I prefer to choose hotels with a score on cleanliness above 4-star, so eliminate all hotels with cleanliness below 4-star.*	(Acquire the cut-off value for cleanliness) **Eliminate** hotels whose score on cleanliness is lower than 4-star
3. *Ok, there are four hotels left. Then, I will evaluate the detailed information of the four hotels one by one.*	**Choose**

2. **Lexicographic:** with lexicographic heuristic, participants determine the most important attribute and then examine the values of all alternatives for that attribute. The alternative with the best value for the most important attribute is selected. Table 4.4 shows an example of the LEX process.

Table 4.4 *An example of Lexicographic*

Verbal protocols	Elementary Information Processes
1. *I think accessibility to attractions is the most important, so I would like to sort hotels by their distance to "houhai".*	(Choose the most important attribute) **Compare** the distance to destination of the hotels
2. *Then, I will go to the detail page of the first hotel.*	**Choose**

3. **Eliminate-by-aspect** plus **Lexicographic**: this procedure is a combination of Eliminate-by-aspect and Lexicographic. Participants eliminate alternatives with values for an attribute below the cut-off. The process continues until a smaller number of alternatives remains. Then, they select the alternative with the best value on the most important attribute.

4. **Eliminate-by-aspect** plus **Additive difference**: firstly, participants narrow down the set of alternatives in terms of Eliminate-by-aspect, i.e., setting cut-offs on attributes. Then, the remaining alternatives are compared on multiple dimensions because no one option best meets all of objects. Participants progressively sum the weighted differences between alternatives on each attribute and select the alternative with the best overall value (see Table 4.5).

Table 4.5 *An example of Eliminate-by-aspect plus Additive difference*

Verbal protocols	Elementary Information Processes
1. *First, the hotel should be near "Houhai" because there are many attractions nearby.*	(Acquire the cut-off value for district) **Eliminate** hotels in other places
2. *And the score of cleanliness should be above 4-star.*	(Acquire the cut-off value for cleanliness) **Eliminate** hotels whose cleanliness score is

Verbal protocols	Elementary Information Processes
	lower than 4-star
3. *When the above requirements are met, I am concerned with the values on price, address and cleanliness for the remaining alternatives.*	**Read** and **Compare** the values of each attribute across alternatives
4. *The third one is slightly more expensive but scores higher on cleanliness, and the difference in distance to destinations is unnoticeable.*	**Difference** the value of several important attributes across alternatives
5. *I think the extra cost on price is worth the cleanliness.*	**Add** the difference on each attribute
6. *Then, I will go to the detailed information page of the third hotel.*	**Choose**

3 /50(6%) participants adopted Lexicographic, 9/50 (18%) participants made use of Eliminate-by-aspect plus Lexicographic, 18/50 (36%) participants screened out alternatives by Eliminate-by-aspect, and 20/50 (40%) participants used Eliminate-by-aspect plus Additive difference (see Figure 4.3 (left)). In a nutshell, at stage one, 30 of 50 participants (60%) adopted non-compensatory strategies (i.e., conflict avoiding), whereas 20 of 50 participants (40%) changed to more compensatory strategies (i.e., conflict confronting [37]) with the decrease of the number of alternatives (see Figure 4.3 (right)). In the following, we use EBA, LEX, EBA+LEX, EBA+ADDIF to denote participants who adopted corresponding decision strategies.

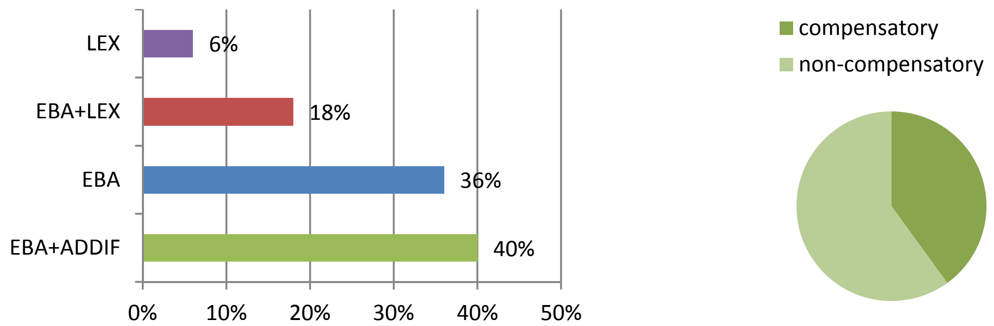


Figure 4.3 *Decision strategies utilized in stage one*

4.2.1.2 Information Acquisition in Eliminate-by-aspect

1. 47 participants began by narrowing down the range of options by Eliminate-by-aspect (9 with EBA+LEX, 18 with EBA, and 20 with EBA+ADDIF) to simplify the complexity of choice. 34% (16/47) of participants eliminated alternatives by static features (e.g., address, price and facility), 10.7% (5/47) did so by customer reviews, and 55.3% (26/47) of participants relied on both static features and customer reviews (see Figure 4.4 (left)). Significantly more users narrowed down the range of options in terms of both, $\chi^2(2) = 14.09$, $p<.05$. For different type of participants, the type of information they used is listed in Figure 4.4 (right). There is no significant association between the type of participants and the type of information utilized, $\chi^2(4) = 2.79$, $p>.05$.

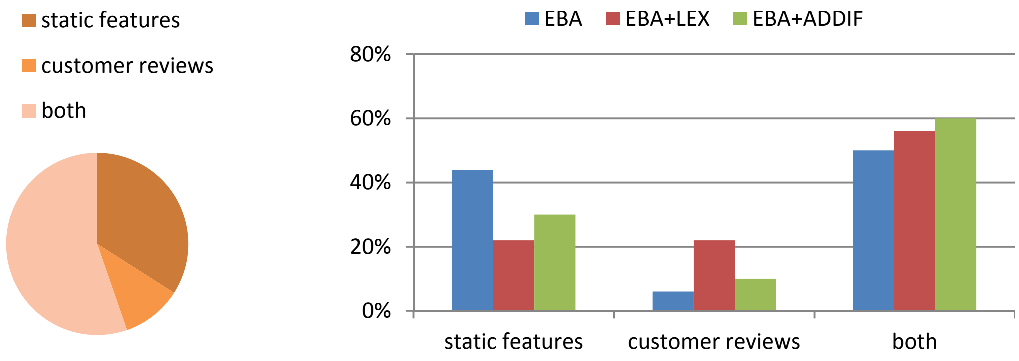


Figure 4.4 *Information acquisitions in Eliminate-by-aspect*

2. In greater detail, we list the specific information of static features and customer reviews to which participants referred (see Figure 4.5). On average, 2.62 attributes (S.D. =1.22) were utilized, to which static features and customer reviews respectively contribute 1.5 and 1.12. Concerning customer reviews, the number of participants who eliminated alternatives by attributes extracted from customer reviews (denoted by opinion attributes) is more than that using an overall review score (26/47 vs. 5/47). The difference is significant, $\chi^2(1) = 14.23$, $p<.05$. For example, people mentioned "*I would like to book hotel with higher user rating on cleanliness and location, while for other aspects, like service, I do not mind*".

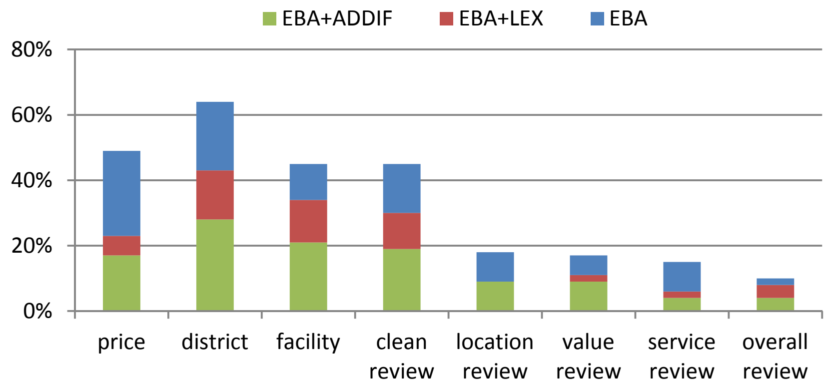


Figure 4.5 *specific information acquisitions in Eliminate-by-aspect*

3. The cut-off level determines which alternatives pass the initial screen and can be further examined, i.e., the minimum level of each attribute the chosen alternative will possess. Only when people are familiar or experienced with the preference object is the cut-off value determined by stable preference, which is well-defined by previous experience and relatively constant across situations. In most condition, the process of generating cut-offs may itself be adaptive, determined by the joint effect of task and context contingencies [43]. According to our observation, the cut-off value of an attribute is highly conditional on two characteristics of choice: the value distribution of the attribute and the correlation among attributes.

   The value distribution of an attribute: Participants would like to generate the cut-off value in terms of both the value range and the number of alternatives within a certain range (i.e., the value distribution of an attribute). For example, when there few/many available options satisfied the cut-off value, participants relaxed/tightened the constraints. Some participants browsed the attribute values of all remaining hotels and then decided the cut-off value: "*Ok, there are 7 hotels left. The cleanliness score of the 7 hotels is from 3.5/5 to 5/5, I will choose hotels with cleanliness above 4.5-star.Then, 4 hotels are left*".

   The correlation among attributes: Attribute correlation affects the adaptation of the cut-off value. In the hotel example, price and quality are generally thought to be negatively correlated because higher-quality hotels tend to have higher rents. One might explore a hotel with a price above one's original price limit just to see how much better it is. If it greatly exceeds expectations, the cut-off may be shifted; however, if prior expectations are confirmed, the cut-off is reinforced. For example, "*the cleanliness of hotels cost € 20 is 4-point, while the cleanliness of hotels at €30 is 4.5-point. The extra €10 is not worth the higher rating on cleanliness, so I will consider hotels cost around €20*".

	Price	Attribute
Alternative 1	390	5
Alternative 2	275	4.5
Alternative 3	325	4.5
......	......	......
Alternative n	170	3.5

Figure 4.6 *An example of the correlation among attributes*

The determinants of cut-off values for different participants are listed in Figure 4.7. There are no significant differences in the references of stable preference ($\chi^2(2)$ =.98, p>.05) or the distribution of an attribute ($\chi^2(2)$ =4.29 p>.05) between participants. Participants who explicitly considered trade-offs among values in product selection (i.e., EBA+ADDIF) significantly more frequently referred to attribute correlation to determine the cut-off value compared with other participants, $\chi^2(2)$=6.41, p<.05.

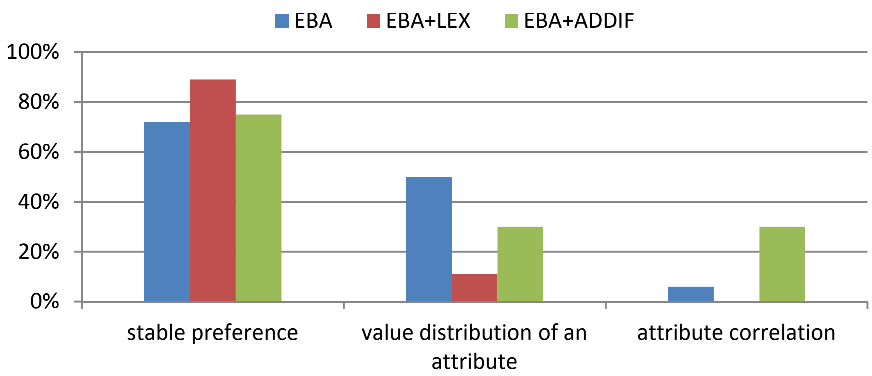


Figure 4.7 *Determinants of cut-off values*

4.2.1.3 Information Acquisition in Lexicographic

1. 12 participants selected alternatives by LEX (3 with LEX, 9 with EBA+LEX). 58.3% (7/12) of subjects chose the entity with the best value on a certain static feature, and 41.7%

(5/12) chose based on customer reviews (see Figure 4.8 (left)). The two proportions are not significantly different, $\chi^2(1)$ =.33, p>.05. In addition, there is no significant relationship between the type of participants and the type of information used to sort alternatives, $\chi^2(1)$ =1.03, p>.05 (see Figure 4.8 (right)).

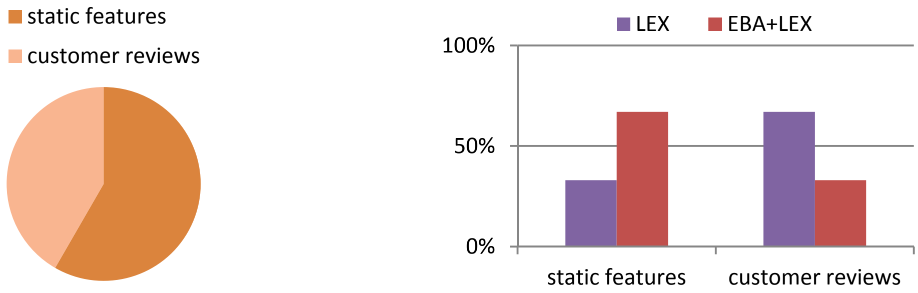


Figure 4.8 *Information acquisitions in Lexicographic*

2. The frequency of each attribute considered the most important is listed in Figure 4.9. As to sorting by customer reviews, the proportion of participants who selected hotels in terms of an opinion attribute is not significantly different from that using an overall review score (2/12 vs.3/12), $\chi^2(1)$ =.20, p>.05.

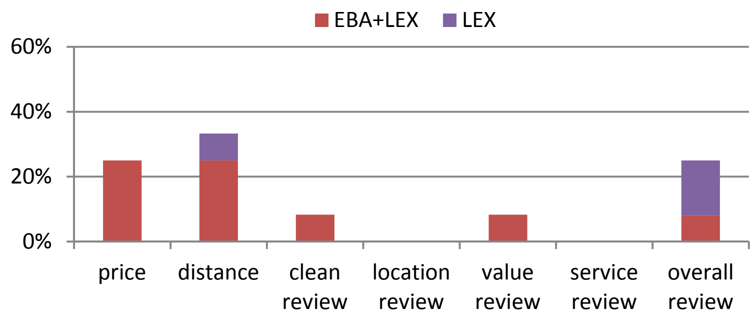


Figure 4.9 *Specific information acquisitions in Lexicographic*

3. During the process, the weight of an attribute is determined not only by previous experience but also by the value range of an attribute. If several alternatives are within a just-noticeable difference (JND) on an attribute, they are thought to be tied, and then the second most important attribute is considered. Meyer and Eagle (1982) and Goldstein (1990) found that the importance weight given to an attribute is a function of attribute ranges [32, 59]. As the variance in the values on one attribute across alternatives increases, the importance weight on that attribute becomes higher.

4.2.1.4 Information Acquisition in Additive difference

1. 20 participants screened hotels with the EBA+ADDIF strategy. That is, after narrowing down alternatives to a relative smaller set by EBA, participants would compare the remaining alternatives on multiple attributes to select one for further consideration. Because price and quality are generally thought to be negatively correlated (i.e., higher-quality hotels tend to have higher rents), all 20 participants referred to both price and customer reviews to make decisions. In addition, 45% participants added address into comparison (see Figure 4.10).

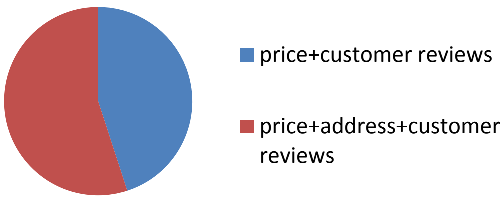


Figure 4.10 *Information acquisitions in Additive difference*

2. Considering the information of customer reviews, 4/20 participants (20%) compared alternatives on cleanliness, 6/20 (30%) did so on location and cleanliness, 7/20 (35%) relied on every opinion attribute (i.e., location, value, cleanliness and service), and 3/20 (15%) selected based on the overall review score (see Figure 4.11). In general, significantly more participants compared alternatives by opinion attributes in comparison with those associated with an overall review score (17/20 vs. 3/20), $\chi^2(1)=9.80$, $p<.05$. Moreover, during product comparison, the extent to which one is willing to trade off more of one valued attribute for less of another valued attribute is different. In other words, people gave different relative importance (i.e., weight) to attributes.

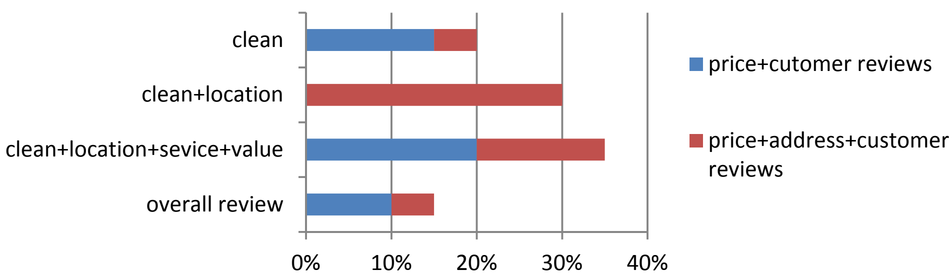


Figure 4.11 *specific information acquisitions in Additive difference*

3. The format in which the sentiment of an opinion attribute is evaluated can be different, for example, numerical (i.e., average score, the number of reviews) versus verbal (i.e., word pairs). The sentiment utilized to accomplish the comparison is listed in Figure 4.12. Overall, the majority of participants (10/20) relied on a combination of numerical and verbal values, followed by just numerical values (9/20). The smallest proportion relied on only adjective-noun word pairs (1/20).

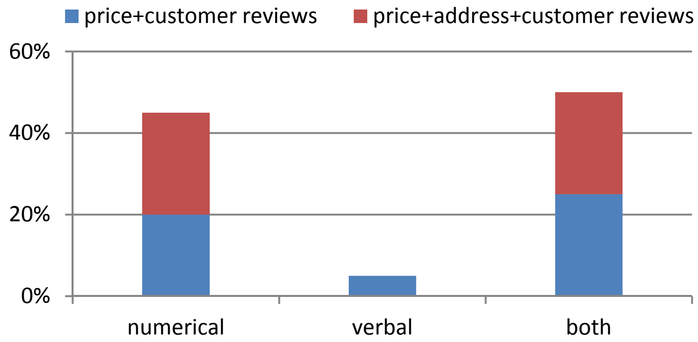


Figure 4.12 *Values for opinion attributes in Additive difference*

4.2.2 Stage 2: Evaluating Alternatives in Detail

In this stage, only one alternative is considered at a time. Participants read detailed information about the option to decide whether to save it as a purchase candidate.

4.2.2.1 Decision Strategy at stage two

Participants tend to use alternative-based (also called as holistic) manner, which means users evaluate multiple attributes of a single alternative and compare them with a predefined cut-off value, often thought of as an aspiration level. If any attribute value is below the level, the alternative is rejected. When the values of all attributes meet the aspiration level, the alternative is chosen as a purchase candidate.

Table 4.6 *An example of the decision strategy utilized in stage two*

Verbal protocols	Elementary Information Processes
1. *The price and facilities of the hotel are ok for me.*	**Read** the static features of an alternative and **Compare** them with aspirations

Verbal protocols	Elementary Information Processes
2. *From the photo, I learn it is a typical courtyard, which is my favorite.*	**Read** the photos and **Compare** them with aspirations
3. *Then, I am concerned about the customer review on cleanliness. The average rating is high and travelers said it is neat and clean, although slightly small.*	**Read** the sentiment of cleanliness and **Compare** it with aspirations
……	……
4. *Ok, I will save the hotel in shopping cart*	**Choose**

4.2.2.2 Information Acquisition at Stage Two

1. After selecting a product from the first stage, the information that users evaluate at stage 2 can be classified into three types: static features, photos and customer reviews. On the whole, 50% (25/50), 84% (42/50) and 88% (44/50) of participants evaluated static features, photos and customer reviews, respectively. For different types of participants, the type of information they evaluated is shown in Figure 4.13. There is no significant association between the type of participants and whether they would evaluate static features $\chi^2(3)$ =1.53, p>.05, photos $\chi^2(3)$ =2.26, *p*>.05, and customer reviews $\chi^2(3)$ =2.18, *p*>.05.

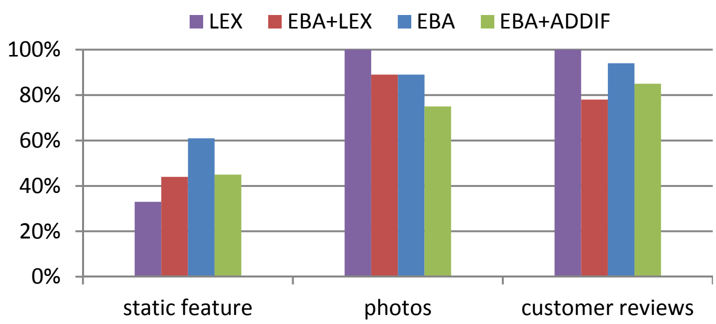


Figure 4.13 *Information acquisitions in stage two*

2. More specifically, Figure 4.14 shows which aspects of customer reviews that participants would inspect. The majority of participants read customer reviews in a feature-driven manner instead of holistic manner, 38/50 vs. 6/50. For example, people

mentioned "*I mainly concern about information important for me, e.g., cleanliness and location, while others are indifferent…. (reading reviews)… but I cannot find content on cleanliness, most of them are talking about location, service"*. Through statistical analysis, the number of participants reading reviews in a feature-driven manner is significantly larger than the number of participants doing so in a holistic manner, $\chi^2(1) = 23.27$, $p<.05$.

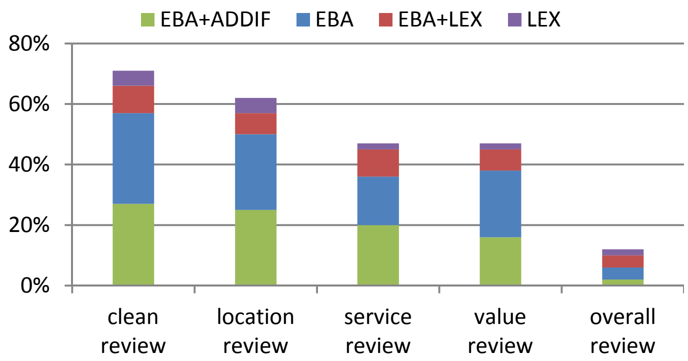


Figure 4.14 *Information acquisitions for customer reviews in stage two*

3. In general, 75% of participants obtained information for each opinion attribute in terms of both numerical values (e.g., average rating, the number of reviews) and verbal values (e.g., adjective-noun word pairs and raw reviews), whereas the other 25% of participants relied only on verbal values (see Figure 4.15 (left)). The information format that different participants evaluated at this stage is listed in Figure 4.15 (right). There is no significant association between the type of participants and the format of the sentiment, $\chi^2(3) = 2.68$, $p>.05$.

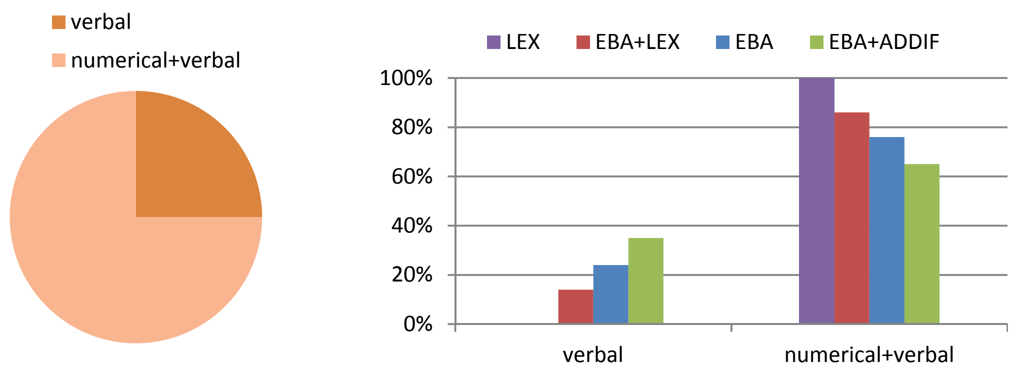


Figure 4.15 *Values for opinion attributes in stage two*

Numerical value: Concerning the numerical values of customer reviews, 24% and 39% of participants evaluated the average rating and average rating plus the number of reviews, respectively, whereas the other 37% also read the time distribution of all and the 5-point customer reviews to examine whether there is a download trend for customer reviews (see Figure 4.16). For example, people noted "*the trend of customer reviews often change over time, like the hotel may improve its service, so that the recent reviews may be opposite to old reviews in which people complained about the service. In other words, the average score extracted from customer reviews of all time can not exactly reflect its real value* ".

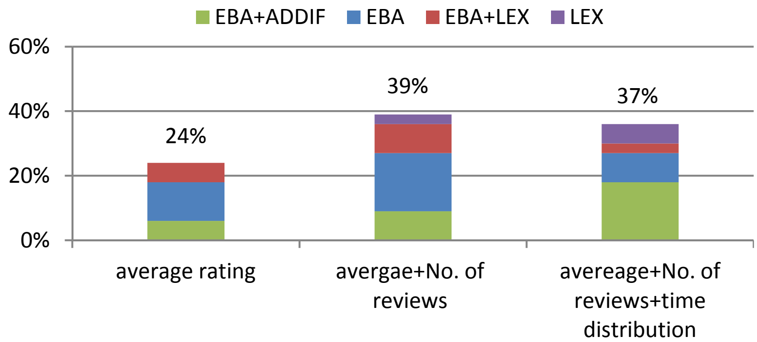


Figure 4.16 *Numerical values for opinion attributes in stage two*

Verbal value: In addition to the numerical values of customer reviews, 7%, 56.5% and 36.5% of participants referred to the verbal values in terms of summarized adjective-noun word pairs, raw reviews and both, respectively. Overall, 93% of participants read raw reviews (i.e., the textual comments) to assist in the context understanding behind positive/negative reviews.

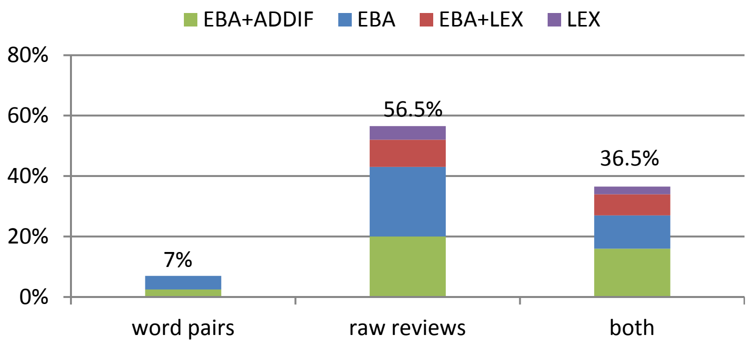


Figure 4.17 *Verbal values for opinion attributes in stage two*

4. Due to the large quantity of raw reviews, 41 participants who read raw reviews performed two types of behavior: seeking the latest and/or the most negative customer reviews.

   The timeliness of customer reviews

   33 participants (80%) sorted customer reviews by date, i.e., reading the latest reviews. People mentioned that "*as things change over time, the renovation, equipment would change within months, so I will follow the newest reviews*", "*I would like to read the newest reviews… especially the reviews written by those who just lived in there last night… I think it will be closer to the real condition and more credible*".

   More attention to negative reviews

   Participants did not treat positive and negative reviews equally. More than half of the participants clearly indicated that they favored negative comments compared with positive ones. For example, people said "*I would like to read negative ratings, and learn the reasons why other customers gave lower rating to see if I have the same concern.",* and *"The reason of adding it in shopping cart is not only because how good it is, but also whether I can stand its drawbacks*".

4.2.3 Stage 3: Comparing Candidates to Confirm the Final Choice

For the number of candidates added to the shopping cart, which ranged from 1 to 4, the primary choices (84% total) were two and three (see Figure 4.18). At this stage, **43** participants who saved more than one option engaged in this stage, namely, comparing the candidates to confirm the final choice.

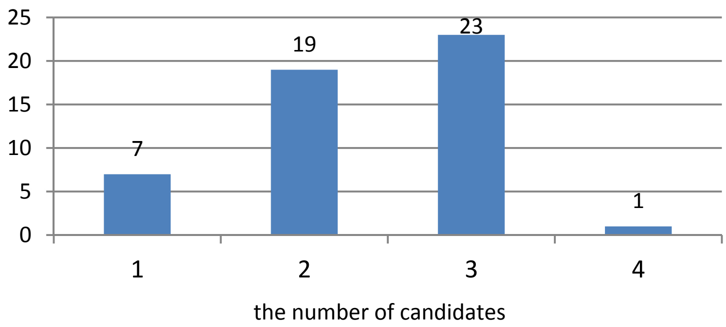


Figure 4.18 *The number of candidates in stage three*

4.2.3.1 Decision Strategy at Stage Three

The decision strategy involved at this stage can be interpreted as calculating the value difference across alternatives on one attribute. Moving to another attribute, the process of comparison is repeated. The differences are summed to obtain an overall evaluation for each entity. Finally, the alternative with better evaluation is retained as the final choice.

Table 4.7 *An example of the decision strategy utilized in stage three*

Verbal protocols	Elementary Information Processes
1. *Ok, first I see the price of the two hotels. One is 195, the other is 275.*	**Read, Compare** the value of one attribute across alternatives
2. *The difference is HKD 80.*	Calculate the **Difference** between the values on the attribute
3. *Then, I am concerned about the average rating of cleanliness. One is 4-star, the other is 5-star.*	**Read, Compare** the value of another attribute across alternatives
4. *the expensive hotel is given higher rating on cleanliness.*	Calculate the **Difference** between the values on the attribute
5. *One is a little more expensive but with higher cleanliness score, and the difference of distance to destination is unnoticeable.*	**Add** the difference on attributes
6. *I think the extra cost (HKD 80) is worth the higher score on cleanliness*	**Compare**
7. *So I will choose the hotel with HKD 275*	**Choose**

4.2.3.2 Information acquisition at stage three

1. There are three types of information that participants would compare at this stage: (1) static attributes (e.g., price, address and facility), (2) item sample (e.g., hotel photos), and (3) customer reviews (e.g., cleanliness, service extracted from customer reviews). Overall, of the 43 participants who saved more than one candidate and made decisions

at stage three, 36 participants (83.7%) compared static features, 20 (46.5%) compared hotel photos, and 27 (62.8%) emphasized customer reviews.

Figure 4.19 lists the type of information to which different participants referred. Through statistical analysis, there is no significant difference in the static feature comparison between participants χ^2 (3) =1.32, *p*<.05; there are significant associations between the type of participants and whether they compare customer reviews (χ^2 (3) =8.21, *p*<.05) and whether they compare photos (χ^2 (1) =11.87, *p*<.05).

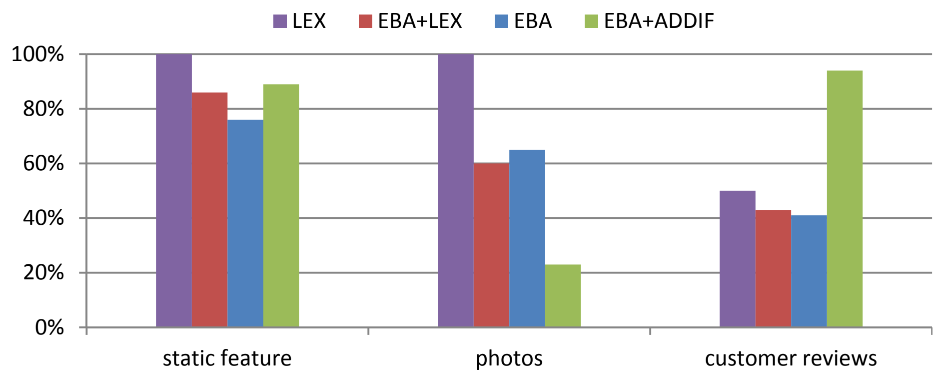


Figure 4.19 *Information acquisitions in stage three*

More notably, we found that participants who adopted a compensatory strategy at stage one, i.e. EBA+ADDIF, (denoted as compensatory) focused significantly more on customer reviews (χ^2 (1) =16.59, *p*<.001) and less on photos (χ^2 (1) =7.34, *p*<.01) compared with participants who adopted non-compensatory strategies, i.e. EBA, LEX and EBA+LEX, (denoted as non-compensatory), as shown in Figure 4.20. This seems to represent the fact that based on the odds ratio, the odds of participants comparing customer reviews were 25.5 times higher if they adopted compensatory strategies than if they adopted non-compensatory strategies in the first stage. The odds of participants comparing photos were 6.22 times higher if they adopted non-compensatory strategies than if they adopted compensatory strategies in the first stage.

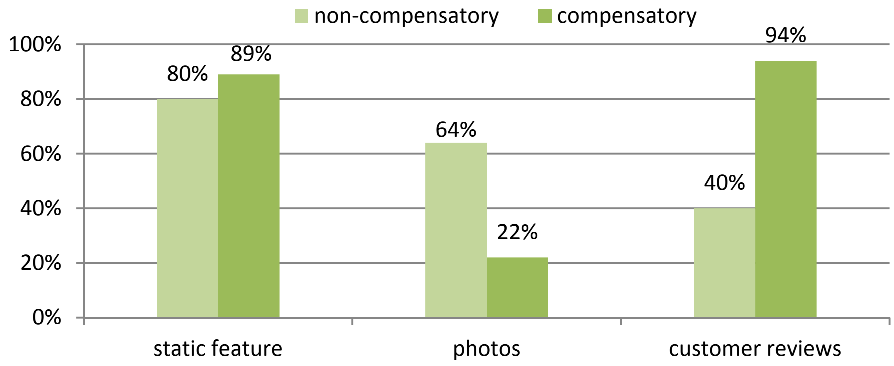


Figure 4.20 *Comparison on Information acquisitions in stage three*

The reason for the difference is that the relative weight placed in maximizing accuracy and minimizing effort is different. Participants who prefer non-compensatory strategies more greatly emphasized minimizing effort, rather than referring to extensive amount of information to make an optimal decision: "*I would compare the photos, as it can bring me more intuitive impression, and choose the most attractive one*".

2. Figure 4.21 illustrates the frequency of each attribute utilized in the product comparison at this stage. On average, participants compared 1.56 static features (S.D. = 1.08) and 1.58 opinion attributes (S.D. = 1.61). For all types of participants, price is the most frequently compared, which means people treat price as a crucial factor in online purchasing. Moreover, significantly more participants used {opinion attribute, sentiment} pairs extracted from customer reviews to perform feature-by-feature and side-by-side comparison between products compared to those merely referring to an overall review score (22/43 vs. 5/43), χ^2 (1) =10.7, p<.001.

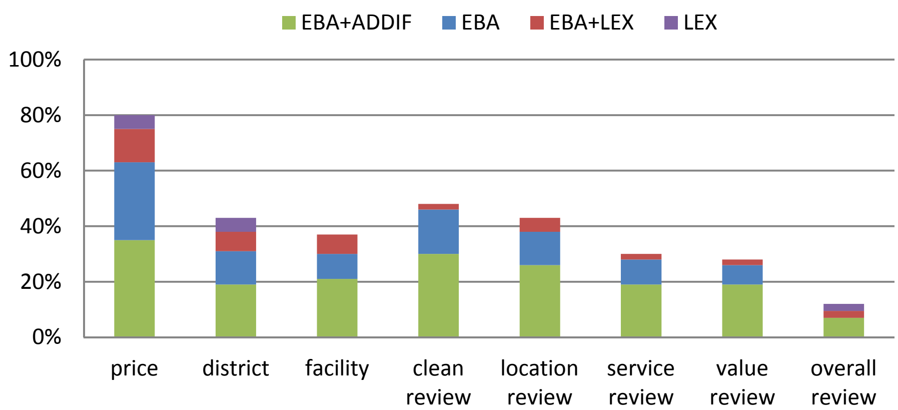


Figure 4.21 *Specific information acquisitions in stage three*

3. Out of the 27 participants who compared customer reviews, 13/27 participants made their decisions based on numerical values, 3/27 participants relied on the adjective-noun word pair description, and 11/27 participants accomplished the comparison based on a combination of numerical and verbal values. There is no significant difference between the type of participants and the value format they inspected, χ^2 (6) =4.84, *p*>.05.

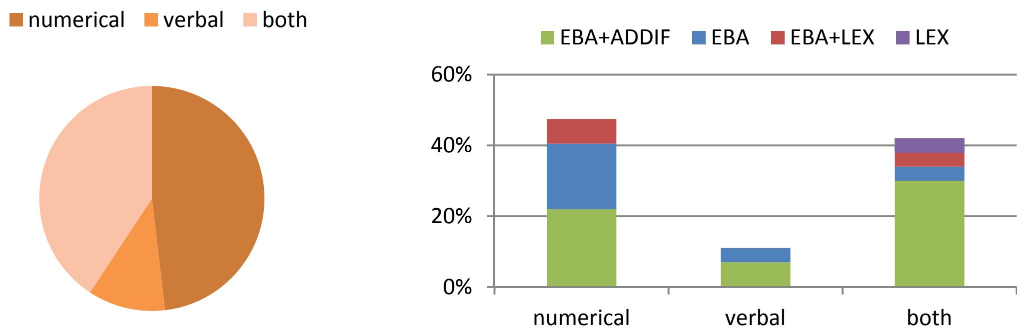


Figure 4.22 *Values for opinion attributes in stage three*

4.3 Personas

Persona is a design technique that "models" the target audiences. The above findings lay the groundwork to create personas. The creating process involves encapsulating the key findings from user research, identifying the critical behavior pattern and turning them into a set of useful characterizations. A persona encapsulates the most critical behavior data of the users in that group, in a way that is clear, memorable, and likely to invoke empathy.

1. **Persona 1**

Zhang is a graduate student who likes to travel during holiday. For hotel reservations, he tends to screen out interesting choices in terms of eliminate-by-aspect. More concretely, he eliminates alternatives with values for an attribute below a cut-off; the process continues with the second attribute and then the third, until a smaller set remains. Static features and attributes extracted from reviews are used for eliminating alternatives. Moreover, the choice of cut-off value for an attribute is adaptive, depending on the value distribution of the attribute, in addition to stable preferences obtained from previous experiences.

Zhang went to the detail page to evaluate a single alternative in terms of static features, the item sample (e.g., photos) and customer reviews. If no value is below his aspiration level, the alternative is saved in his shopping cart; otherwise, it is rejected. More specifically, he reads customer reviews in a feature-driven manner and gains information on the sentiment toward each attribute by both numerical and verbal values. Numerical values give a general understanding, for example, of how much the attribute is liked or disliked through an average rating and the number of reviews. In addition, because opinions always change over time, the time distribution of positive/negative customer reviews is also considered. Meanwhile, the verbal values (adjective-noun word pairs and raw reviews) provide information on why that rating was given; the latest and negative raw reviews are more frequently inspected considering the enormous number of raw reviews.

After saving several candidates in his shopping cart, Zhang compares the values on multiple attributes across alternatives and calculates the value difference between alternatives. He then sums the differences to purchase the alternative with a relatively better overall evaluation. In the process of making attribute-based comparison, Zhang more greatly emphasizes static features (e.g., price) and traveler photos which provide an intuitive impression, focusing less on attributes extracted from customer reviews that require an extensive amount of effort to address.

2. Persona 2

Betty is a white-collar employee who likes travelling alone and takes this time as a period of relaxation outside of busy work. She does not like to spend too much effort on hotel reservations; rather, she simply chooses one with the best value on the most important attribute. This process is known as Lexicographic. The most important attribute can be a static feature, some aspect of customer reviews (e.g., cleanliness) or the overall user review, which is determined by the value ranges of attributes and stable preference.

Assuming that she enters the detail page of the hotel with the highest user rating on service, the processes of evaluating alternatives in detail to decide whether to add it to her shopping cart and subsequently comparing the candidates saved in the shopping cart to confirm a final choice are roughly the same as with persona 1.

3. **Persona 3**

   Tom, an employee at a foreign trade corporation, often travels for business. When booking a hotel, he first eliminates unsatisfying options in terms of static features and aspects of customer reviews (i.e., eliminate-by-aspect), for example eliminating hotels whose price is higher than the reimbursement limitation. More notably, the cut-off level of each attribute is determined not only by existing preferences but also by the value distribution of the attribute. Then, from the remaining options, he selects the one with the best value on a single attribute considered the most important (i.e., Lexicographic). The attribute can be a static feature, an aspect of customer reviews or the overall user rating depending on previous experience and the value ranges of attributes.

   For example, Tom thinks distance to his destination is the most important; he opens the detail page of the nearest hotel. The process of evaluating the alternative in detail is not distinguished from personas 1 and 2. Moreover, the process of comparing candidates to confirm a final choice is roughly the same as with personas 1 and 2.

4. **Persona 4**

   Monica is a PhD student of computer science with a rigorous personality. With the purpose of finding the most suitable hotel, she first narrows down the set of options to simplify the complexity of choices in terms of eliminating alternatives by static features and aspects of customer reviews (i.e., eliminate-by-aspect). Different than the above three personas, she decides the cut-off value not only by previous experience and the value distribution of an attribute, but also by considering the trade-off between attributes, i.e., the correlation among attributes. When a relatively smaller set of alternatives remain, she compares them on both static features and attributes extracted from customer reviews by explicitly considering how much she is willing to give up the value on one attribute for value improvement in another (i.e., weighted additive difference). More specifically, the value of attributes extracted from reviews is evaluated in terms of both numerical values (i.e., average rating, the number of reviews) and verbal values (i.e., adjective-noun word pairs). Then, the alternative with a relatively better evaluation is selected for further consideration.

The process of evaluating single alternative in detail is more or less the same as with personas 1, 2, and 3.

After saving several candidates in her shopping cart, Monica also contrasts alternatives based on the values of multiple attributes and calculates the value difference between alternatives on each attribute. She then sums the differences to select the alternative with the better overall evaluation. Similar with the above three personas, static features, especially price, are most often used. However, Monica relies more on attributes extracted from customer reviews and less on traveler photos, which is contrary to the three personas. Concerning the value of the attributes of customer reviews, both numerical values (i.e., average rating, the number of reviews) and verbal values (i.e., adjective-noun word pairs) are utilized for product comparison.

Grounded in the analysis of the results the formative study, the critical behavior patterns of the four personas at each stage are summarized in Tables 4.8, 4.9 and 4.10, which briefly depict the decision strategy personas adopt and the information they seek to implement the strategies.

Table 4.8 *Personas in stage one*

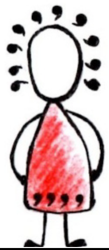
		Persona 1	**Persona 2**	**Persona 3**	**Persona 4**
Stage one: screening out interesting alternatives					
Decision strategy		EBA	LEX	EBA+LEX	EBA+ADDIF
EBA	Attributes utilized to eliminate alternatives	Static features *and/or* attributes extracted from customer reviews		Static features *and/or* attributes extracted from customer reviews	Static features *and/or* attributes extracted from customer reviews
	Determinants of cut-off values	Stable preferences The value distribution of an attribute		Stable preferences The value distribution of an attribute	Stable preferences The value distribution of an attribute The attributes correlation
LEX	The most important attribute		A static feature/ an attribute extracted from customer reviews/ the overall user review	A static feature/ an attribute extracted from customer reviews/ The overall user review	
	Determinants of weight		Stable preferences The value range of an attribute	Stable preferences The value range of an attribute	
ADDIF	Attributes utilized to select alternatives				*Both* static features *and* attributes extracted from customer reviews
	Values of the attributes extracted from customer reviews				Numerical values (average rating, the No. of reviews) and verbal values (adjective-noun word pairs)

Table 4.9 *Personas in stage two*

<table>
<tr><th colspan="2"></th><th>Persona 1</th><th>Persona 2</th><th>Persona 3</th><th>Persona 4</th></tr>
<tr><td colspan="2">Stage two:
Evaluating alternatives in detail</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Decision strategy</td><td colspan="4">Satisfying</td></tr>
<tr><td colspan="2">Attributes evaluated in stage two</td><td colspan="4">Static features
Item sample (i.e. traveler photos)
Attributes extracted from customer reviews</td></tr>
<tr><td rowspan="2">Values of the attributes extracted from customer reviews</td><td>Numerical</td><td colspan="4">The average rating
The number of reviews
Time distribution of positive/negative reviews</td></tr>
<tr><td>Verbal</td><td colspan="4">Adjective-noun word pairs extracted from reviews
(the latest and negative) Raw reviews</td></tr>
</table>

Table 4.10 *Personas in stage three*

<table>
<tr><th colspan="2"></th><th>Persona 1</th><th>Persona 2</th><th>Persona 3</th><th>Persona 4</th></tr>
<tr><td colspan="2">Stage three:
comparing candidates to confirm final choice</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Decision strategy</td><td colspan="4">Additive difference</td></tr>
<tr><td colspan="2">Attributes utilized in comparison</td><td colspan="3">More frequent: Static features and Item sample (e.g. traveler photos)
Less frequent: Attributes extracted from customer reviews</td><td>More frequent: static features and Attributes extracted from customer reviews
Less frequent: Item sample (e.g. traveler photos)</td></tr>
<tr><td rowspan="2">Values of the attributes extracted from customer reviews</td><td>Numerical</td><td colspan="4">The average rating
The number of reviews</td></tr>
<tr><td>Verbal</td><td colspan="4">Adjective-noun word pairs extracted from customer reviews</td></tr>
</table>

Chapter 5

Interface Design for

Screening out Interesting Alternatives

This chapter focuses on the interface design for screening out interesting alternatives, to facilitate the process of narrowing down available alternatives and selecting interesting ones for further consideration.

The design process is organized as follows. Firstly, based on the personas built in preceding chapter, we depict idealized interactions between personas and system, referred to as context scenarios, from which design requirements are extracted. Further, we introduce the status quo and limitations of interfaces in current E-commerce websites and then go on to describe in detail the steps of generating design solutions for three parts: the filter panel, sorting panel and details panel.

5.1 Context Scenarios

A context scenario is a plausible textural description of an idealized typical interaction with a future system from the personas' points of view. Each scenario begins with a specific situation and then describes the interaction between personas and the system from the beginning of a task through its completion [33]. In the following, we depict different context scenarios from the perspective of each persona.

1. **Scenario for persona one (EBA)**

   Zhang prefers hotels with Wi-Fi to facilitate him sharing photos with friends and families after a day of travel and at reasonable price because he is a student without much money; in addition, cleanliness is also thought of as important.

   In the process of screening interesting hotels, he was used to a non-compensatory strategy – eliminate-by-aspect. He first eliminated all hotels without Wi-Fi. For price, Zhang did not have an explicit cut-off value. He learned that of the remaining 7 hotels, four hotels cost approximately 40 Euros. The other three hotels cost more than 50 Euros, so he decided to eliminate all hotels with prices higher than 50 Euros. Then, he went on to narrow down his choices in terms of cleanliness scores given by travelers. Four hotels remained, two of them with 5 points for cleanliness and the other two hotels with 4.5 points. Thus he eliminated hotels with values on cleanliness below 5 points. Taken together, Zhang determined the cut-off value for an attribute in terms of both its value range and the number of alternatives within a certain range, i.e., the value distribution of the attribute, in addition to stable preference obtained from previous experience.

2. **Scenario for persona two (LEX)**

   Betty recently planned to take a trip to the seaside in the coming summary vacation. With the goal of having a nice period of relaxation away from busy work, she treats hotel service as the most important factor in hotel reservations.

   Thus, she sorted all options by user rating on services and went to the detail page of the hotel with the highest score on service for further consideration.

3. Scenario for persona three (EBA+LEX)

Tom, an employee at a foreign trade corporation, is going to Beijing for a business trip. In hotel reservations, there is a price limit considering reimbursement.

He eliminated the hotels whose price was higher than the reimbursement limitation. He observed that the cleanliness ratings of the remaining hotels ranged from 5 points to 4 points, which is not noticeably different for him. However, their locations are widely dispersed. As a business person, time is very tight, so Tom regarded convenience to the destination as the most important attribute and selected the nearest one from the remaining hotels. For choosing the most important attribute, Tom determined the weight of each attribute in terms of its value range to some extent.

4. Scenario for persona four (EBA+ADDIF)

To reserve a hotel with the best value for money, Monica first narrowed down the hotels to simplify the complexity of choice by eliminating alternatives with values for an attribute below a certain vale. The cut-off of an attribute was determined not only by previous experience and value distribution but also by correlation among attributes. For example, in acquiring the cut-off for price, originally she wanted to book a hotel costing less than 20 Euros. However, price and user ratings are usually negatively correlated. She found that the cleanliness of hotels costing less than 20 Euros is approximately 3.5 points, but the cleanliness of hotels with prices at 30 Euros is 4.5 points. She believed the extra 10 Euros is worth the higher rating on cleanliness, so she effectively adapted the cut-off of price to 30 Euros**.**

With a relatively smaller set of alternatives, she was used to explicitly considering trade-offs among attributes and choosing the option with the best overall evaluation. For example, rather than treating all attributes equally, she thought cleanliness played a more important role compared with other features. Thus, Monica gave cleanliness a higher weight and sorted alternatives by multiple attributes. Next, she browsed the hotels' information and went to the detail page of the hotel with the highest weighted additive value.

5.2 Requirements

Requirements (i.e., product definition), which connect the dots between user research and design, are implied in the above context scenarios. Requirements can be clustered into 2 types: functional requirements and data requirements. In the following, the main functional requirements and their associated data requirements are described (see Table 5.1).

5.2.1 Eliminating alternatives with values for an attribute below a certain value

In the process of screening out interesting alternatives, 94% (47/50) of participants began by narrowing down the set of alternatives by means of eliminating alternatives with values for an attribute below a certain value to simplify the complexity of the choice.

1. Including both static features and opinion attributes (i.e., attributes extracted from customer reviews) in the filter

   Our formative study revealed that the majority of participants (26/47) eliminated alternatives based on both static features and customer reviews. In addition, more users selected opinion features (e.g., cleanliness and service) to eliminate alternatives, the proportion of which is significantly higher than those selecting based on an overall review score, 26/47 vs. 5/47.

2. Displaying the value distribution of each attribute

   When determining the cut-off value for an attribute, people refer to its value range and the number of remaining alternatives within a selected range (i.e., the value distribution of an attribute), which can avoid invalid filter; for example, too many/few options meet requirements due to a loose or strict cut-off.

3. Representing the correlation among attributes

   Price and customer reviews are generally thought to be conflicting. For example, when the cut-off value on cleanliness is set at 5 stars, the minimum price of the remaining hotels may be beyond the acceptable range. Thus, representing the correlation among attributes (especially negatively correlated attributes) can help users explicitly consider the trade-off when determining the cut-off of an attribute.

5.2.2 Selecting an alternative with the best value on a single attribute

24% (12/50) of participants tend to screen out interesting alternatives using Lexicographic, namely determining the most important attribute and then selecting the alternative with the best value on that attribute.

1. Incorporating opinion attributes (i.e., attributes extracted from customer reviews) in sorting

   Instead of a static feature (e.g., price), 41.7% of participants would select the best alternative in terms of customer reviews. Moreover, rather than an overall rating, some participants regarded a certain opinion attribute as more important, e.g., selecting the hotel with the highest rating on cleanliness.

2. Representing the ranges of attribute values

   The weight of an attribute, to some extent, is determined by its value range of available options. Generally, the more diverse in values one attribute is across alternatives, the heavier the weight that is assigned to that attribute.

5.2.3 Choosing an alternative with the best weighted additive value

After narrowing down the set of alternatives under consideration, 40% (20/50) of participants adopted a more compensatory strategy, namely Additive Difference, i.e., comparing the remaining alternatives on multiple attributes to select the one with the best overall value. Moreover, the extent to which one is willing to compensate the value of one attribute with another varies. In other words, the relative importance (i.e., “weight”) of an attribute given by different people is different, and the relative importance is different across attributes for a given decision maker.

Hence, users should be provided access to sort alternatives by multiple attributes and give different weights to the attributes to facilitate users to select an alternative with the best weighted additive value.

1. Involving both static features and opinion attributes in multi-attribute sorting

   Participants select the alternative with the best overall evaluation based on both static features and customer reviews. Concerning customer reviews, 85% of participants performed feature-by-feature comparisons of alternatives, i.e., calculating the differences on opinion attributes, rather than merely taking the overall rating.

2. Representing both numerical and verbal values toward each opinion attribute

   45% (9/20) of participants compared opinion attributes in terms of numerical values, i.e., the average rating and the number of reviews; 5% (1/20) compared with verbal values, i.e., adjective-noun word pairs; and 50% (10/20) of participants used both numerical and verbal values.

Table 5.1 *Requirements of interface design for screening out interesting alternatives*

Scenario	Functional Requirements	Data Requirements
Scenario 1, 3, 4	Eliminate alternatives with values for an attribute below a certain value.	Both static features and opinion attributes should be included in the filter.
		For each attribute, the value range and the number of alternatives within a selected range should be represented.
		Enable users to learn attribute correlation, especially the correlation between negatively correlated attributes.
Scenario 2, 3	Select an alternative with the best value on a single attribute.	In addition to static features and overall rating, opinion attributes should be incorporated in sorting.
		Represent the ranges of attribute values across alternatives in the choice set.

Scenario	Functional Requirements	Data Requirements
Scenario 4	Give different weights to attributes to select an alternative by weighted additive value.	Both static features and opinion attributes are involved in multi-attribute sorting.
		The sentiment information toward each opinion attribute should be represented by both numerical and verbal values.

5.3 Interfaces in E-commerce Websites

If all entities of a category can be described by the same set of attributes, they are referred to as multidimensional discrete data. Each attribute defines a data dimension, and each entity can be thought of as a point in a multidimensional space. Taking hotels as an example, they are described by both static features (e.g., price and address) and opinion attributes (e.g., cleanliness, location and service).

At its most basic, to help users screen interesting alternatives within a large number of multidimensional discrete data, E-commerce websites employ the extremely common and powerful three-panel pattern: (1) the filter panel, (2) the sorting panel, and (3) the details panel. Figure 5.1 illustrates the interfaces of the two most popular E-commerce websites (Booking and Amazon), which are consistent with the three-panel pattern. In the following, we will separately explain the three panels in detail.

1. The filter panel

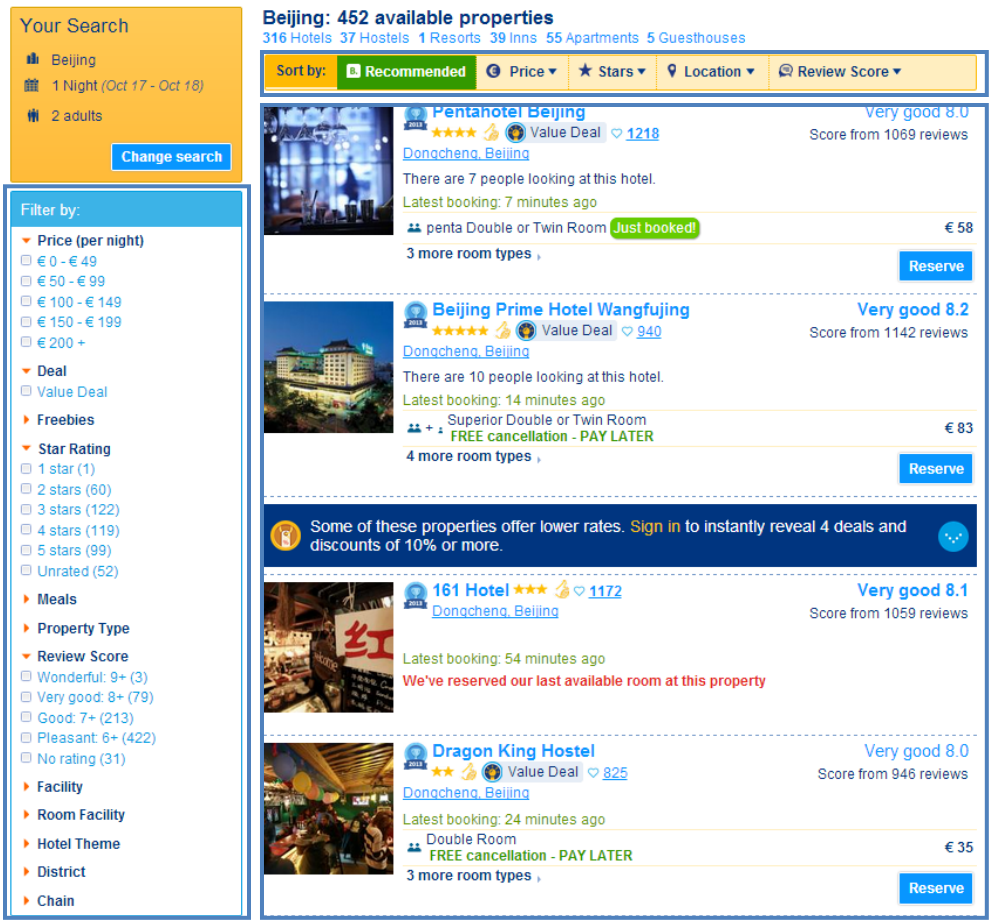
   The filter panel is composed of a set of controls, with which users interact to eliminate alternatives with values for an attribute below a cut-off to narrow down the option space displayed in the details panel. Each control restricts the range displayed on each of the data dimensions to allow users to eliminate unsatisfying options. Users can narrow down the set of alternatives from several dimensions by concocting a series of controls. Usually, it is either on the left or top of the screen to take advantage of the visual flow whereby most users read left-to-right and top-to-bottom [86].

2. The sorting panel

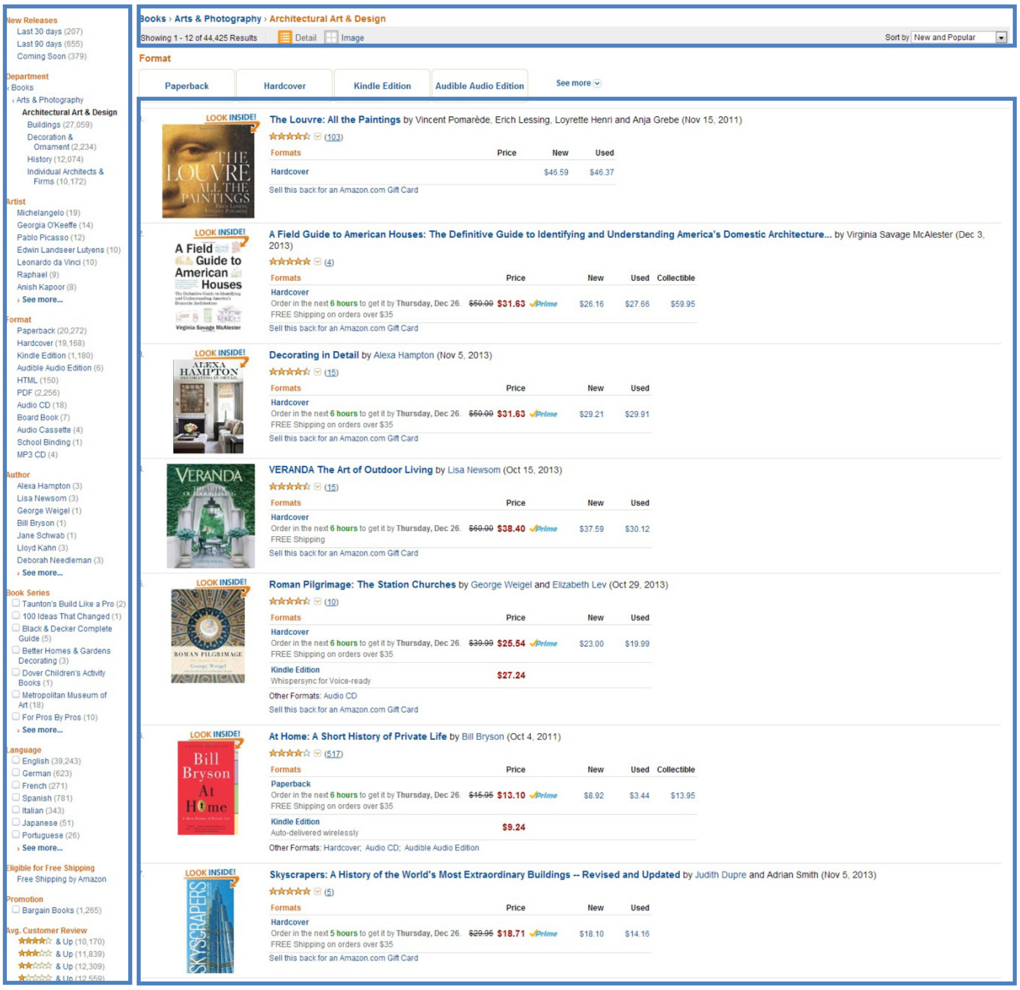
Generally located above the details panel, it enables users to sort alternatives by a specific attribute in descending/ascending order. The alternative with the best value on that attribute is represented at the beginning of a list, satisfying the match between the importance of information for the decision maker and the salience of the information display.

3. The details panel

Typically below or to the right of the filter panel, the details panel shows the details of the objects that match all constraints and facilitates side-by-side comparison based on attribute values. As soon as user adjusts the controls in the filter and sorting panels, there is real-time feedback on the details panel. In this way people can quickly shift their attention back and forth. The interactive hiding and revealing of data are referred to as dynamic queries [2].



(a) Booking



(b) Amazon

Figure 5.1 *The three-panel pattern of E-commerce websites*

Limitations: The first and the most important limitation for current E-commerce websites is that none of them incorporates opinion attributes (i.e., attributes extracted from customer reviews) in the interface for screening out interesting alternatives; rather, most of them merely include an overall user rating with respect to customer reviews (see Fig. 5.1). This obviously cannot satisfy users' requirements in terms of the results of the formative study, which revealed that people more frequently depend on opinion attributes compared with the overall user rating during product filtering, sorting and comparing.

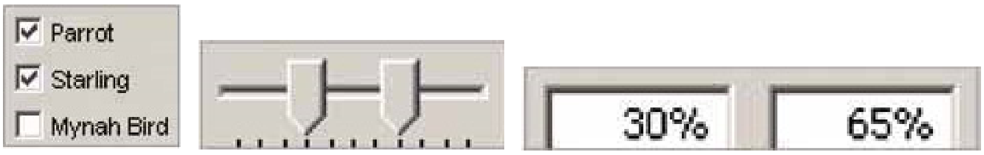
Because opinion attributes are not incorporated in existing interfaces at all, none of the requirements concerning how to represent opinion attributes is applied, e.g., how to represent the correlation between price and opinion attributes on a filter panel.

5.4 The Filter Panel

Driven by the extracted requirements, we begin to develop the opinion-attribute-embedded filter panel to satisfy users' decision making behaviors, namely eliminating alternatives with values for an attribute below a certain value.

5.4.1 The control widget for filter

For opinion attributes, people specify an arbitrary range of values for display, e.g., "*show data points that are greater than this number, but less than the other number*". Depending on their data type, i.e., numerical data with a continuous structure of selection, common easy-to-use controls include the following: (1) array of N checkboxes, (2) double sliders, and (3) Text fields to type in values (see Figure 5.2) [86].



(a) Checkbox (b) Slider (c) Fill-in-the-Blank

Figure 5.2 *Control widgets for filter*

Ahlberg *et al.* (1992) compared slider with fill-in form, and revealed that people performed statistically significantly faster using the interface with sliders than with text fill-ins, and text fields leave more room for errors than sliders and buttons [2]. Then, they (1994) developed an interface that provided the user with a slider to select on the basis of movie length, year of publication, actor and director name (see Figure 5.3) [3]. The checkbox is the traditional filter form utilized in most E-commerce websites, as shown in Figure 5.1. In the following sections, taking hotel domain as an example, the filter controls for opinion attributes are represented in the forms of checkboxes and sliders.

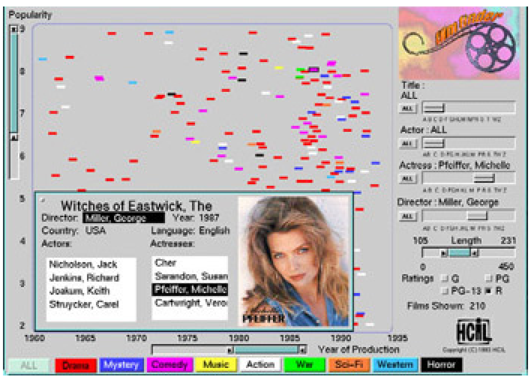


Figure 5.3 *The FilmFinder demonstration application*

5.4.2 Checkbox filter

Hotels are depicted by four opinion attributes, including cleanliness, location, value, and service. The filter for each attribute is composed of an array of N checkboxes representing different value ranges on the dimension. Users can select entities with values on one attribute within a certain range by clicking corresponding checkbox. For example, the checkboxes for cleanliness mean choosing hotels with values on cleanliness “above 4.5”, “above 4”, “above 3.5”, “above 3” or “all”. Filters are "rolled up" by default; that is, only their title bars are visible, whereas the checkboxes that display the contents are hidden.

Every checkbox is followed with the number of entities within the range (shown in the bracket), which is equivalent to the distribution of an attribute. More specifically, when the range of an attribute is adjusted (e.g., selecting hotels with scores of 4+ on cleanliness), the number of entities at each range of other attributes changes simultaneously according to the selected subset (see Figure 5.4). However, this method is not obvious enough for users to infer information about attribute correlation.

Figure 5.4 *Framework of checkbox filter*

5.4.3 Slider filter

In the frame of slider filter, users can drag a slider knob to narrow down the displayed subset of the data space on the basis of one dimension. By concocting these controls incrementally, users can find a smaller data set. There is a limitation on the number of sliders that can be meaningfully used; arbitrarily, the limit is put at 10 [90]. Figure 5.5 shows an interface design for online hotel booking, in which the controls for price and opinion attributes are represented in the form of a double slider.

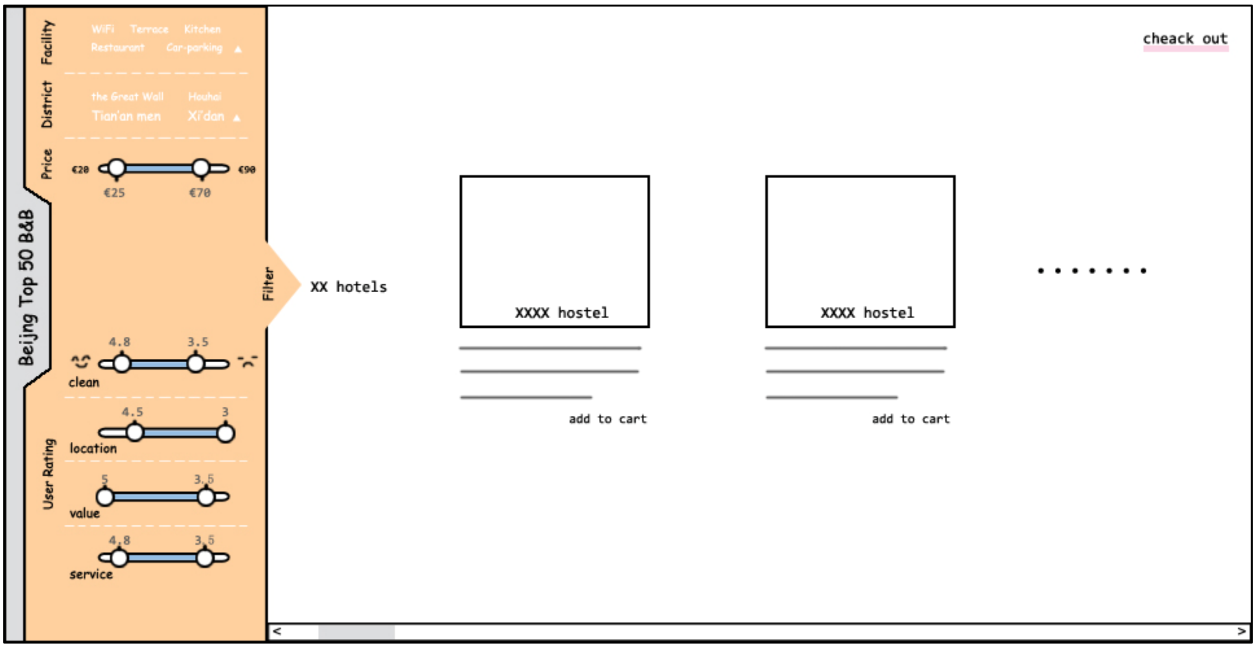


Figure 5.5 *Framework of slider filter (mouse-out)*

1. **Representing the distribution of each attribute via bars**

   Revealing the distribution of an attribute can help users avoid empty or invalid query results due to strict/loose restrictions. For a slider filter, the distribution of an attribute is represented in the form of bar chart (see Figure 5.6). The height of a bar is proportional to the number of hotels, which can be a good visualization of data because it may reduce learning time and potential misunderstanding [57]. And the number of hotels with values satisfying the specified range on an attribute is shown right above the bars on the basis of the Gestalt principles of proximity [90]. Because people more readily pick up information close to the fovea, less time and effort will be spent in neural processing and eye movements if related information is spatially grouped.

2. Displaying the correlation among attributes via simultaneous change

Real-time change in the range on sliders can relate values on different attributes [6]. More concretely, when the range on one of the sliders is adjusted, the ranges on other sliders simultaneously change. For instance, as the user drags the slider knob for maximum price to lower price direction, the maximum values for opinion attributes decrease corresponding to the new price range. While users increase the minimum values for opinion attributes, the minimum value for price increases accordingly. However, in light of the useful field of view (UFOV), people concentrate their attention around fovea and can quickly take in information only from 1 to 4 degrees of visual angles [91]. In other words, users tend to miss the changes.

With the purpose of clearly expressing attribute correlation between price and opinion attributes, we use lines to connect the corresponding knobs of different sliders, known as a node-link diagram, which is very powerful to express a relationship between them [90]. The likely explanation for why linking lines represent relationships so well is that there are deep metaphors based on our sensory perception of the world, that provide a scaffolding for even the most abstract concepts [47, 72]. In addition, in recent study, Holten and Van Wijk (2009) [38] showed that using tapered lines with the broadest end at the source node makes it easier to trace relationships compared to the use of conventional arrows. Hence, when the knob of one slider is dragged, the matching knobs of other sliders and their connecting lines synchronously move. Peterson and Dugas (1972) and Bartram & Ware (2002) suggested that motion works well in the periphery of vision. In other words, the UFOV function can be far larger for detection of moving targets than for detection of static targets [5, 71].

3. Hover query

We employ an easy, rewarding form of interactivity, which makes the critical information prominent and shows details "hidden behind" specific points only when hovering mouse over, referred to as “details on demand” [79]. More specifically, when users hover over or drag a slider knob, the value distribution of the attribute is represented in the form of bar chart; meanwhile, the relevant information linked with the knob on the graph is highlighted by connecting lines. In this way, we do not need to put all information on the surface, avoiding information overload, and users can easily look up what they need.

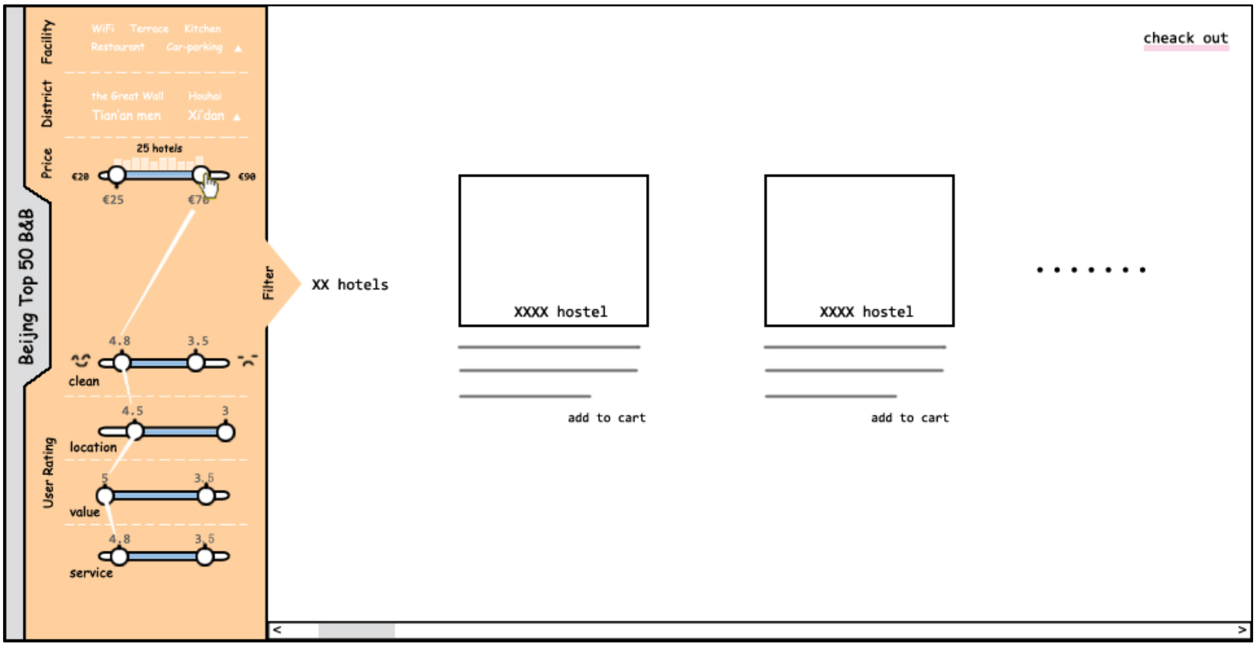


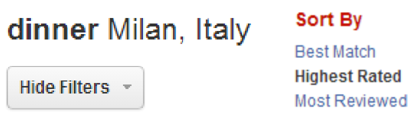
Figure 5.6 *Framework of slider filter (mouse-over)*

In comparison to pure sliders, two major modifications are made to accommodate presenting the distribution of each attribute and correlation among attributes when hovering over or moving a slider knob, to help users more effectively determine the cut-off value of an attribute by tweaking the knob.

We suggest that the modified slider filter can stimulate and facilitate users to consider the trade-off between attributes when determining the cut-off value for an attribute. Therefore, it can produce more effective adaptation of cut-offs by means of narrowing down alternatives based on a more compensatory process and consume less effort because an effortful assessment of choices is not required. We have verified this hypothesis by comparing the slider filter against the checkbox filter (which has been commonly adopted in E-commerce) (see Chapter 6 for more detail).

5.5 The Sorting Panel

The sorting algorithm (i.e., listing alternatives by a single attribute in ascending/descending order) in E-commerce facilitates users to choose an item with the best value on the most important attribute. The attributes with respect to customer reviews utilized in sorting contain an overall user rating and the number of reviews (see Figure 5.7). However, grounded in users' decision making behaviors, the existing sorting function cannot satisfy their requirements. In the following, we improve the design of the sorting panel in two aspects: (1) incorporating opinion attributes in sorting and (2) providing access to sort alternatives by weighted additive value.



(a) sorting panel in yelp.com

(b) sorting panel in HotelsCombined.com

Figure 5.7 *Sorting panel in E-commerce websites*

5.5.1 Attributes in sorting

The analysis of the results of the formative study revealed that both participants who selected the alternative with the best value based on a single attribute (i.e., LEX, EBA+LEX) and those who selected based on multiple attributes (i.e., EBA+ADDIF) highly depend on opinion attributes to make decisions, in comparison to an overall user rating. Therefore, we suggest that it is useful to incorporate opinion attributes in sorting, in addition to static features.

5.5.2 Multi-attribute utility theory

Decision problems often involve conflicts among attribute values; that is, no one option best meets all objects. Hence, instead of selecting an alternative in terms of merely a single attribute, some participants refer to multiple attributes and resolve their conflict by directly considering the extent to which one is willing to trade off more of one attribute (e.g., the cost of a car) for less of another attribute (e.g., the safety of the car), reflected as the relative importance or weight.

Grounded in the more compensatory decision making process, it is reasonable to assume that a decision maker's preference for each alternative can be predicted by multi-attribute utility theory (MAUT) [46], which allows normative trade-offs between attributes in terms of their importance. The alternative with the highest overall evaluation is predicted as the most satisfying one.

More specifically, the utility of an alternative A with multiple attributes $\{x_1 \dots x_n\}$ can be described by the multi-attribute utility function $U(x_1 \dots x_n)$. For the sake of simplicity, we think of all attributes as mutually independent. Under the additive independence assumption, the utility of an alternative is calculated by multiplying the weight by the single-attribute utility function $U_i(x_i)$ and summing these weighted values over all attributes. See Formula 5.1,

$$U(x_1 \dots x_n) = \sum_{i=1}^{n} \omega_i U(x_i) \tag{5.1}$$

where the constants ω_i are used to weight each single-attribute utility function $U_i(x_i)$ in terms of its importance, scaled from 0 to 1 and summed to 1 (see Formula 5.2).

$$0 < \omega_i < 1, \qquad \sum\nolimits_{i=1}^{n} \omega_i = 1 \tag{5.2}$$

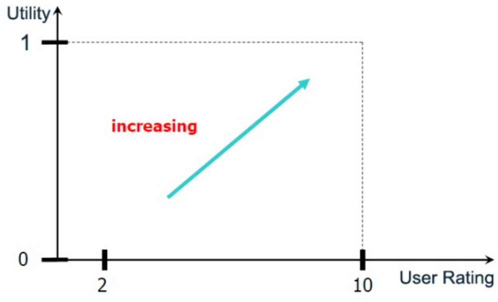
Assuming that an alternative has three attributes, x, y, and z, its additive utility function is shown in Formula 5.3.

$$U(x, y, z) = \omega_x U(x) + \omega_y U(y) + \omega_z U(z). \tag{5.3}$$

1. Single-attribute utility function

Utility is a subjective level of satisfaction that expands domain beyond a monetary value. It is measured in a conventional scale, usually from 0 to 1. The worst outcome of an attribute is associated with 0, whereas the best is associated with 1.

To estimate the utility function of each attribute, two main types of numerical attributes are considered: more-is-better and less-is-better. For example, users may prefer a product with the maximum rating. In other words, the opinion attribute is more-is-better. In contrast, price is less-is-better because users prefer to minimize it. Thus their utility functions can be respectively presented as a monotonic increasing/decreasing function (see Figure 5.8).



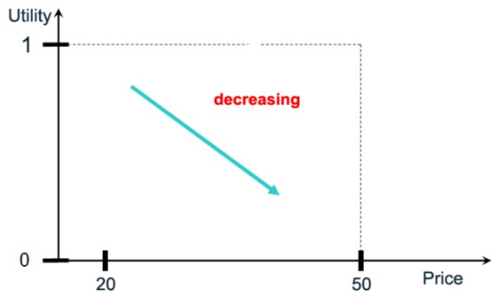


Figure 5.8 *Single-attribute utility function*

The single-attribute utility function is not necessarily straight; it can be a convex or concave line depending on users' subjective perceptions. To simplify the process of utility acquisition, we assume the utility function is linear, which most users will not change most of the time.

For the more-is-better attribute, we define its utility function as shown in Formula 5.4.

$$U(x) = \frac{x - \min(a)}{\max(a) - \min(a)} \tag{5.4}$$

For the less-is-better attribute, the utility function is defined as shown in Formula 5.5.

$$U(x) = \frac{\max(a) - x}{\max(a) - \min(a)} \tag{5.5}$$

min(a) and max(a) are the minimum and maximum values of attribute *a* in the choice set.

2. The relative importance of attributes

Weight is the relative importance of each attribute compared to the others. According to previous studies, there are three types of methods to acquire the weights of different attributes from users: direct assessment, indifference method and indirect measurement [46, 92]. In the following, we will explain the three methods in detail.

Direct assessment: direct-assessment methods ask users to directly assign numerical values to all concerned attributes, for example, ranging from 0 "least important" to 1 "most important", and then normalize them so that the sum is 1. According to Meyer and Eagle (1982) and Goldstein (1990), attribute weights vary as a function of the range of each attribute in the choice set. The greater the range, the greater the weight for the attribute should be. Grounded in this phenomenon, Fischer (1995) provided evidence that representing the range of outcomes associated with each attribute in direct assessment can effectively avoid range-insensitive bias [29].

Indifference method: indifference methods infer attributes' weights on the basis of an indifference point, which is obtained by modifying one of two sets of stimuli until subjects feel that there is no difference between the two [29]. For example, two jobs are identical in all aspects, but job A1 pays $40,000 per year & offers 5 days of vacation and job A2 pays $___per year & offers 20 days of vacation. Subjects are asked to specify how much salary they would give up in exchange for moving from the low end (5 days) to the high end (20 days) of the vacation day range. Assuming that the decision maker feels that $25,000 is the indifference point between the two jobs, the following equation 5.6 is built:

$$\omega * \mathrm{U}(\$40{,}000) + (1-\omega) * \mathrm{U}(5\text{ days}) = \omega * \mathrm{U}(\$25{,}000) + (1-\omega) * \mathrm{U}(20\text{ days}) \quad (5.6)$$

Then, the attributes' weights can be calculated as follows:

$$\frac{\omega}{1-\omega} = \frac{\mathrm{U}(20\text{ days}) - \mathrm{U}(5\text{ days})}{\mathrm{U}(\$40{,}000) - \mathrm{U}(\$25{,}000)}$$

Indirect measurement: Indirect-measurement techniques avoid directly asking people to rate the importance of attributes; instead, subjects simply state their preference for different alternatives [49]. Moreover, Huber *et al.* provided evidence that giving relative preference on pairs is better than rating an individual profile in predicting choice [44]. Hence, a pairwise preference comparison is taken as the form of indirect measurement in our study. For example, the decision maker can give a ratio rating for his/her preference on two Jobs: A1 pays $40,000 per year & offers 5 days of vacation and job A2 pays $20,000per year & offers 20 days of vacation (e.g., job A1 is twice as satisfying as job A2). The Formula 5.7 is established as follows, where α is the preference ratio:

$$\frac{\omega * \mathrm{U}(\$40{,}000) + (1-\omega) * \mathrm{U}(5\ \mathrm{days})}{\omega * \mathrm{U}(\$20{,}000) + (1-\omega) * \mathrm{U}(20\ \mathrm{days})} = \alpha \tag{5.7}$$

Accordingly, the attributes' weights can be inferred from the graded pairwise preference comparison.

5.5.3 Implementation

The three different methods for weight elicitation are implemented in the area of online hotel booking. Two static features (e.g., price and distance) and four attributes extracted from customer reviews (e.g., clean, location, value, and service) are included in the sorting panel.

Because users are required to specify trade-off judgments in all methods, we employ a dropdown list where options are interval-scaled measures, rather than typing in values as the input. In other words, users give their judgments by selecting from a list of options, which not only reduces the amount of work that users have to do but also offers an example of what type of answer is sought [19].

1. **Direct assessment:** each attribute utilized for sorting alternatives is followed by its value range in the choice set and a set of five interval-scaled options {0.2, 0.4, 0.6, 0.8, 1}, where users can choose the relative importance for each attribute. For example, when price and distance are selected, their value ranges and weights are represented as follows (see Figure 5.9):

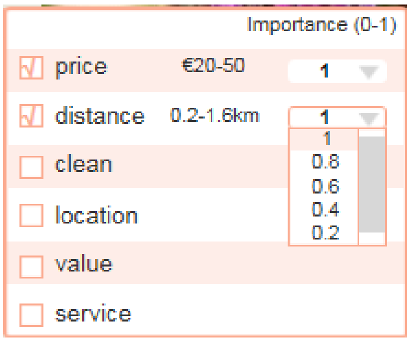


Figure 5.9 *Direct assessment*

2. **Indifference method:** when an attribute (except for price) is selected to sort alternatives, its weight can be elicited by requiring user to specify how much he/she would spend in exchange for moving from the worst to the best value on the attribute. Assuming that subjects sort hotels by price and distance to attraction, they are asked to determine the price of the nearest hotel (0.2 km away to attractions) to make it equal to the one far from attraction (1.5 km). Hotel price can be selected from a set of interval-scaled options (see Figure 5.9).

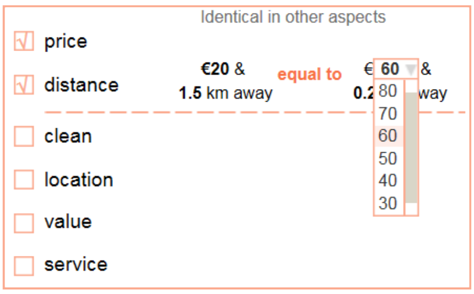


Figure 5.10 *Indifference method*

3. **Indirect measurement:** for each attribute, its weight can be inferred from a graded pairwise preference comparison of two options that are identical in all aspects but option 1 offers the lowest price & the worst value on the attribute and option 2 offers the highest price & the best value on the attribute. For example, when sorting alternatives by price and distance, users are required to state their preference on hotel A1=minimum price & maximum distance vs. hotel A2=maximum price & minimum distance, by selecting from {equal to, better than, worse than}. Moreover, when "better than" or "worse than" is selected, a secondary menu {A little, obviously, strongly,

extremely} will appear to assist users in defining a 4-level preference ratio (see Figure 5.10).

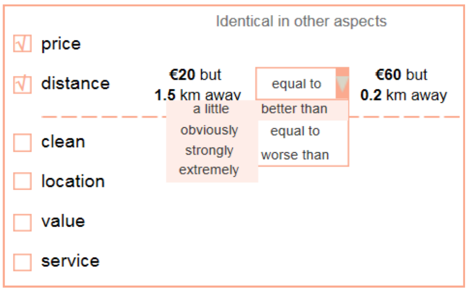


Figure 5.11 *Indirect measurement*

In addition, when users move mouse over one attribute, a tooltip window will be automatically shown to explain the attribute. Users are not forced to state their judgment but are provided with reasonable default values, i.e., the value that most users will not change most of the time [45]. In our interface design, the importance of all attributes being equal is set as the default. Even if the default is not what the user wants, a way for users to change the default value in terms of their interests is offered.

There has been some research evaluating the effectiveness of a particular method for measuring preferences [10, 44]. However, no clear basis exists for saying which of the methods have the highest validity in eliciting weights in an online environment. We will conduct a user study to compare the different methods, investigating which can evoke users to express more accurate preference (see Chapter 7 for more detail).

5.6 The Details Panel

Driven by the weighted additive process, in which users browse through options in an attribute-driven manner (e.g., how much extra has to be paid for a more convenient location or cleaner environment), to confirm an item with a relatively higher value, two design points are applied for the "List of objects" section.

1. **Incorporating opinion attributes in item description**

To facilitate attribute-driven comparisons, in addition to static features, each item should be described by opinion attributes instead of merely an overall user rating. As regard to the value of opinion attributes, users want to look for not only numerical data (i.e., the average rating and number of reviews) but also verbal data (i.e., adjective-noun word pairs) to learn the specific reason behind the numerical rating.

2. Horizontal list of items

The items can be represented by a horizontal list or vertical list. Horizontal movement is faster and easier for the eye than vertical movement (because the action of vertical rectus muscles is more complicated than that of the horizontal rectus muscles) [63], and the computer screen is longer left-to-right than top-to-bottom (meaning more items can be compared in a horizontal list). Thus, we adopt a horizontal list to visualize the set of items. In particular, each item occupies a column and each attribute dimension is represented by a row. The value difference between entities on each attribute is easier to measure.

Chapter 6

Experiment on

Customer-Review-Embedded Filter

In this chapter, we perform a user study to compare two alternative designs for a customer-review-embedded filter panel, called checkbox interface and slider interface. Online hotel booking is adopted as the test domain. Through analysis of objective and subjective measures, the effectiveness of incorporating attributes extracted from customer reviews in the filter panel has been verified. And the modified slider interface (visualizing the distribution of an attribute via bars and the correlation among attributes via simultaneous change) obtains higher scores regarding perceived decision accuracy, cognitive effort, satisfaction and intention to use, compared to conventional checkbox interface.

6.1 Experiment Interfaces

In this chapter, we compare two alternative designs that mainly differ in the filter panel, called checkbox interface and slider interface, through a user study in the context of online hotel booking. A dataset with 100 B&B (50 in Beijing and 50 in Rome) was used for the experiment, and hotels' specifications and reviews were crawled from Tripadvisor.com. Each hotel is described by three static features (i.e., facility, address, price) and four attributes extracted from customer reviews (i.e., cleanliness, location, service, value).

In the checkbox interface, the filter for each attribute is displayed in the form of an array of N checkboxes. Every checkbox represents a different value range for an attribute, followed by the number of entities within the range (shown in the bracket). For example, the filter for price is composed of four checkboxes that denote different price ranges, namely, "20-30 Euros", "30-40 Euros", "40-50 Euros", and "above 50 Euros" (see Figure 6.1).

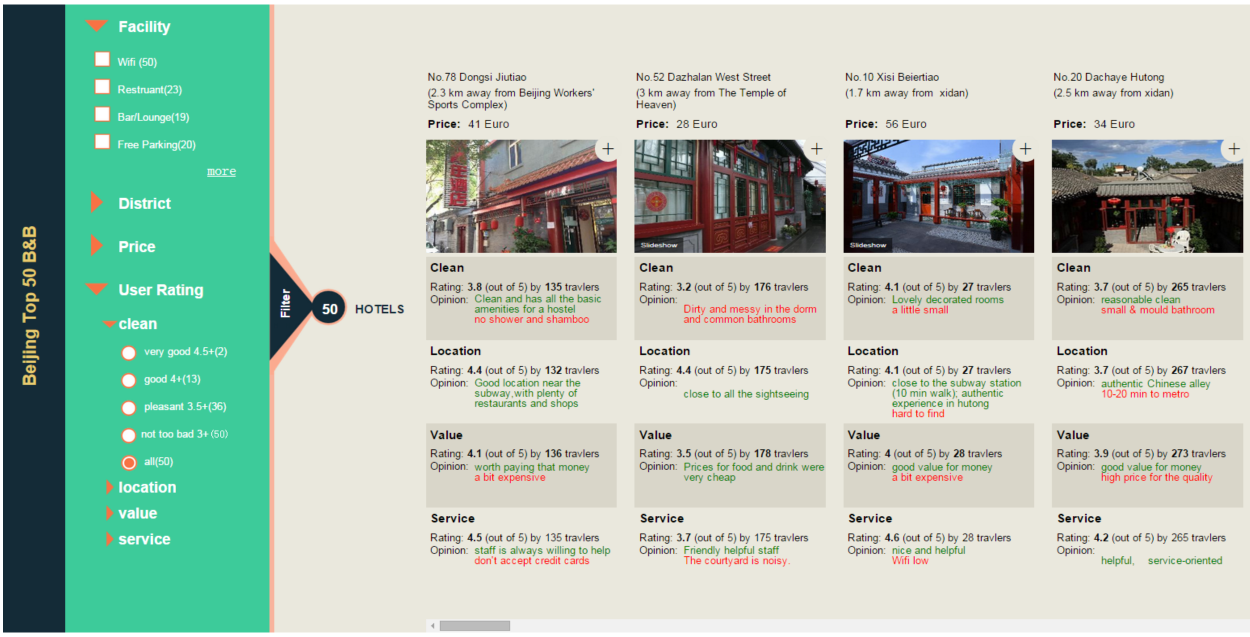
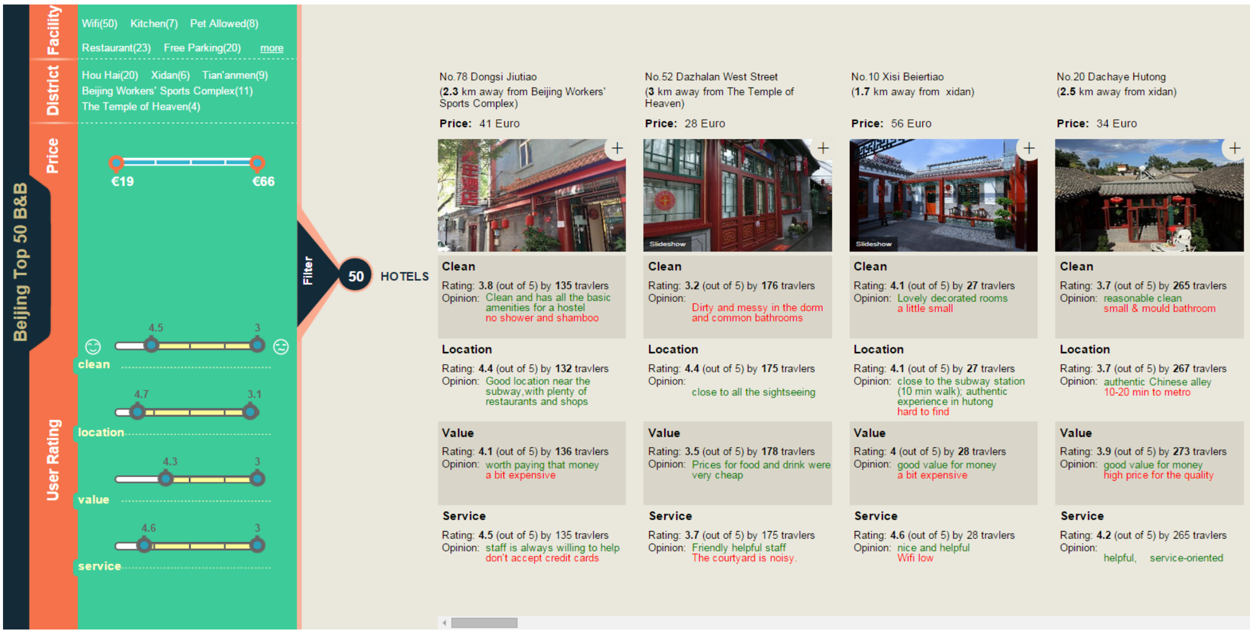


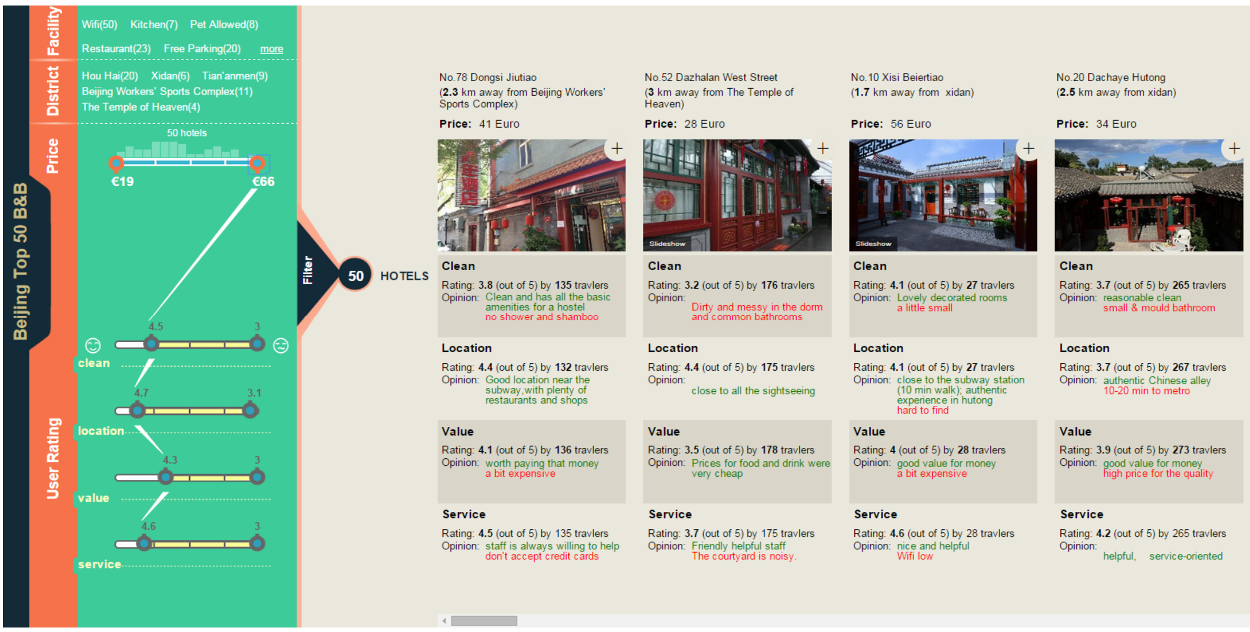
Figure 6.1 *Screenshot of the checkbox interface*

In the slider interface, nominal attributes (i.e., hotel facility and district) have discrete values. For example, the variable representing facility might have values such as wifi, kitchen, car parking, etc. In accordance with the manner of filtering, filters of nominal attributes are represented in the form of a "press-and stick" toggle button. Sliders are used for filters of numerical attributes (i.e. price and attributes extracted from customer reviews).

Each slider has two independently movable "thumbs" that let a user define a range to show data points that are greater than the smaller number but less than the larger number. For example, the price slider sets a range of 19 Euros to 66 Euros (see Figure 6.2 (a)). Moreover, as the user hovers over or moves either end of the price range, the price distribution of hotels is shown via bars and the ranges of other attributes simultaneously change in pace with the new price range through connecting lines (see Figure 6.2 (b)). The same is true for other sliders.



(a) Mouse-out



(b)Mouse-over

Figure 6.2 *Screenshot of the slider interface*

The details panel basically follows the standard multi-column tabular layout, where one product occupies one column. The static features appear at the top rows, while the opinion attributes are in line with the positions of corresponding attributes in the filter panel. Regarding each opinion attribute, both numerical values (i.e., the average rating and number of reviews) and verbal values (i.e., negative and positive opinion words) are presented.

6.2 Evaluation Framework

Given that the experiment's objective is to identify whether the interface with sliders outperforms checkboxes from the user's perspective, in other words, whether users are supported in making more effective and efficient product filters, the measurement is conducted from both objective and subjective perspectives.

6.2.1 Objective metrics

Specifically, to understand the behavior in the two interfaces, we consider two objective variables:

1. The attributes users adopted for narrowing down alternatives and

2. The time users consumed to finish each task

6.2.2. Subjective metrics

Except for above objective measures, the subjective measurement is mainly concerned with users' perceptions of the interface, i.e., the perceived usability of the interface. According to ISO (International Organization for Standardization), usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [94]. The definition consists of three aspects: Effectiveness – the accuracy and completeness with which users achieve certain goals, Efficiency – the resources expended in relation to the accuracy and completeness of goals achieved, and Satisfaction – the user's comfort with and acceptability towards use of the system.

Grounded in the framework, in this experiment we assessed users' subjective perceptions from three major aspects: decision accuracy, cognitive effort and satisfaction. To measure these subjective perceptions, a set of questions were pre-designed mostly from existing studies (see Table 6.1), where they were tested and found to have strong content validity and reliability.

1. **Decision Accuracy**

   The degree of accuracy users subjectively perceived while using a system is also called decision confidence [73]. The confidence can be assessed by requiring users to rate a statement, such as "I am confident that the subset of hotels I just got by means of filtering is the best choice for me" on a Likert Scale ranging from "strongly disagree" to "strongly agree".

2. **Cognitive Effort**

   Users' perceived decision effort refers to the psychological cost of gathering and processing information to make a decision. It represents the ease with which a subject can perform the task of obtaining and processing relevant information [16], so statements such as "I easily obtained and processed the relevant information to narrow down the hotels" are proposed.

3. **Satisfaction**

   Satisfaction can be directly measured by asking users whether the system is enjoyable to use [13]: "it is pleasant to use the filter function to narrow down alternatives". In addition, it can also be assessed indirectly by the extent to which a user is willing to return (i.e., user loyalty) [26, 58], such as "if possible, I would like to use it in the future".

Table 6.1 *Questions to measure users' subjective perceptions on two filter panel designs*

Measured aspects	Related questions
Accuracy	I am confident that the subset of hotels I just got by means of filtering is the best subset for me.
Cognitive effort	I easily obtained and processed relevant information to narrow down the hotels.

Measured aspects	Related questions
Satisfaction	It is pleasant to use the filter function to narrow down alternatives.
	If possible, I would like to use it in the future.

Each question is answered on a 7-point Likert Scale ranging from 1 "strongly disagree" to 7 "strongly agree".

6.3 Hypotheses

Hypothesis 1: On the basis of the results of preceding formative study that 68.1% of participants (31/47) eliminated options in terms of customer reviews and the proportion of users who adopted attributes extracted from customer reviews is significantly higher than that using overall review score (26/47 vs. 5/47), it can be inferred that attributes extracted from customer reviews play an important role in helping people narrow the range of options under consideration. We therefore assume that participants depend highly on opinion attributes to eliminate alternatives in both experimental interfaces.

Hypothesis 2: In light of design principles (see chapter 5.3 for more detail), the slider filter tends to stimulate and facilitate users to consider the attribute distribution and correlation when determining the cut-off value of an attribute. In other words, users likely perceive the slider filter to be more competent in helping them make an informed and accurate product filter in comparison to the checkbox format. Hence, we hypothesize that the slider interface outperforms the checkbox interface in terms of the perceived accuracy and cognitive effort for narrowing down alternatives. Consequently, users are more likely to be satisfied with it and intend to use it in the future.

6.4 Procedure

To collect users' actions and evaluations, we built an online experiment site, which included the task description, evaluated interfaces, and after-study questionnaire. All users' actions were automatically recorded in a log file.

The experiment was set up to compare two filter designs in a within-subject procedure. To avoid any carryover effects, we developed two experiment conditions with a varied order of showing the two interfaces. Participants were randomly assigned to one of the conditions. In addition, every interface was associated with one task scenario related to hotel filter. The two scenarios were randomly assigned to the two compared interfaces to avoid any biased effects on the final results.

Step 1: At the beginning of this experiment, each participant was required to fill in his/her personal background and E-commerce experiences.

Step 2: Next, participants were debriefed on the objective of the experiment and asked to go through two tasks.

Scenario one: imagine that you will have a trip to Beijing with your friends on New Year's holiday and need to book a hostel online. The top 50 Beijing Bed and Breakfast (B&B) are presented. Based on the value of each attribute (e.g., price and user rating), please narrow down the hotels (by means of filtering) to a smaller subset (no less than one hotel) that you are interested in for further consideration.

Scenario two: imagine that you will have a trip to Rome with your friends on Christmas, and need to book a hostel online. The top 50 Rome Bed and Breakfast (B&B) are presented. Based on the value of each attributes (e.g., price and user rating), please narrow down the hotels (by means of filtering) to a smaller subset (no less than one hotel) that you are interested in for further consideration.

Step 3: After finishing each task, users were asked to answer post-study questions about their subjective perceptions with the interface they just used, concerning whether it supported him/her in making an accurate and efficient decision (see Appendix 2).

Step 4: When the two tasks were completed, the participants were required to indicate which interface they preferred and the reasons.

6.5 Participants

We recruited 52 users. They are students pursuing Bachelor, Master or PhD degrees and staffs at Politecnico di Milano. Table 6.2 gives their demographic profile. In the pre-study questionnaire, they specified their frequency of Internet use (on average 4.94 "almost daily", S.D. =.42), e-shopping experiences (on average 3.63 "1-3 times a month", S.D. =.91), and online hotel booking experience (on average 2.60 "more than 3 times", S.D. =.50), which reflects that the test subjects are close to the population of target users.

Table 6.2 *Demographic profile of participants in the experiment for filter panel*

Gender	female (26); male (26)
Age	21-30 (42); >30 (10)
Major	Electronics, Engineering, Architecture, etc.
Internet usage	4.94 (daily/almost daily), S.D.=.42
E-commerce shopping experience	3.63 (1-3 times a month), S.D.=.91
Online hotel booking experience	2.60 (more than 3 times), S.D.=.50

The average scores for 'internet usage' and 'e-commerce shopping experience' are on a 5-point Likert Scale from 1 'least frequent' to 5 'very frequent', and 'online hotel booking experience' is on a 3-point Likert Scale.

6.6 Analysis of Results

After the study, we collected both objective and subjective measures. Because the same participant evaluated both interfaces of the experiment, with the purpose to examine whether there is a significant difference between the two interfaces, we analyzed data in terms of dependent t-test or Wilcoxon signed-rank test depending on whether the sampling distribution of these differences was normal [28].

6.6.1 Objective measures

For the time users expended on each interface, the results show that users on average took 105.65s (S.D. =100.17) with the checkbox interface to conduct narrowing down and 132.34s

(S.D. =95.72) with the slider interface. Through a test of normality, the sample distribution of the time difference is not significantly different from a normal distribution (P>.05). In terms of the dependent t-test analysis, there is no significant difference in the task time between the two interfaces, t (51) =-1.97, p>.05 (see Table 6.3). We can infer that although participants are unfamiliar with slider filter, they do not need to spend extra time on it to finish the task compared with traditional checkbox filter.

Table 6.3 *Comparison on task time with two filter panel designs*

Interface	Mean (S.D.) (second)	dependent t-test		
		t	*df*	*P*
Checkbox	105.65 (100.17)	-1.97	51	.054
Slider	132.34 (95.72)			

Moreover, we recorded the attributes users adopted for narrowing down options. Figure 6.3 lists the frequency of each attribute selection in the two interfaces. The high frequencies of opinion attributes indicate that it is important to incorporate them in the filter panel.

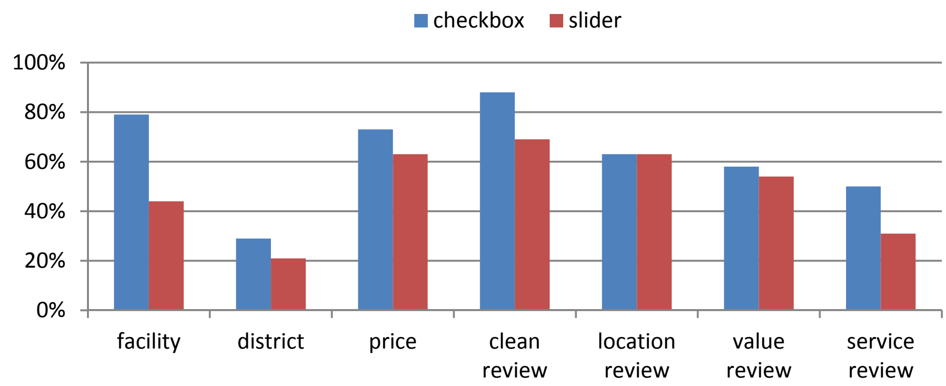


Figure 6.3 *Attributes used to filter options*

In the checkbox interface, the average number of attributes selected is 4.39 (S.D. =1.74), to which static features and opinion attributes contribute 1.81 (S.D. =.73) and 2.58 (S.D. =1.41), respectively. In the slider interface, an average of 3.6 (S.D. =1.70) attributes are used, composed of 1.35 (S.D. =.76) static features and 2.25 (S.D. =1.55) opinion attributes. More notably, significantly more opinion attributes were used to eliminate alternatives in both the checkbox interface (t (51) =-3.69, p<.05) and slider interface (t (51) =-3.51, p<.05) (see Table 6.4).

Table 6.4 *Comparison on the number attributes used to filter options*

Interface	attribute	Mean (S.D.)	dependent t-test		
			t	*df*	*P*
Checkbox	Static	1.81 (.73)	-3.69	51	.001
	Opinion	2.58 (1.41)			
Slider	Static	1.35 (.76)	-3.51	51	.001
	Opinion	2.25 (1.55)			

6.6.2 Subjective measures

We then analyzed users' answers to the post-study questionnaire. Table 6.5 lists users' average scores and the statistical analysis results.

We can observe that users gave more positive score on slider interface than on checkbox interface as to their perceived decision accuracy (mean=5.44, S.D. =1.13 vs. mean=5.15, S.D. =1.14), perceived cognitive effort (mean=5.52, S.D.=1.28 vs. mean=5.00, S.D.=1.34), pleasure (mean=5.69, S.D.=1.20 vs. mean=5.29, S.D.=1.58) and intention to return (mean=5.71, S.D.=1.35 vs. mean=5.33, S.D.=1.25). Because the distribution of these differences between scores is normal, with dependent t-test, the statistical analysis further shows that the slider interface obtains significantly higher rating in comparison with the checkbox interface regarding Q2, "perceived effort", with t (51) =-2.18, p<.05(1-tailed); Q3, "pleasant to use", with t (51) =-1.74, p<.05(1-tailed); and Q4, "intention to use it in future", with t (51) =-1.79, p<.05(1-tailed). Because we predicted that the slider interface outperforms the checkbox interface, in this case, we used one-tailed test.

Table 6.5 *Comparison on users' subjective perceptions on two filter panel designs*

Interface	Mean (S.D.)	dependent t-test		
		t	*df*	*P* (1-tailed)
Q1. I am confident that the subset of hotels I just got by means of filtering is the best choice for me. (perceived accuracy)				
checkbox	5.15 (1.14)	-1.65	51	.052
slider	5.46 (1.09)			
Q2. I easily obtained and processed relevant information to narrow down the hotels. (perceived effort)				
checkbox	5.00 (1.34)	-2.18	51	**.017**
slider	5.52 (1.28)			
Q3. It is pleasant to use the filtering function to narrow down alternatives.				
checkbox	5.19 (1.62)	-1.74	51	**.044**
slider	5.53 (1.16)			
Q4. If possible, I would like to use it in the future.				
checkbox	5.28 (1.16)	-1.79	51	**.040**
slider	5.69 (1.43)			

In the post-study questionnaire, users were asked to choose the interface they preferred. 65% of users (34 out of 52) favored the slider interface, whereas the other 35% of users (18 out of 52) liked the checkbox interface (see Figure 6.4). With Chi-square test, significantly more people liked the slider interface, $\chi^2(1) = 4.92$, $p<.05$. The results again confirm the superiority of the slider filter.

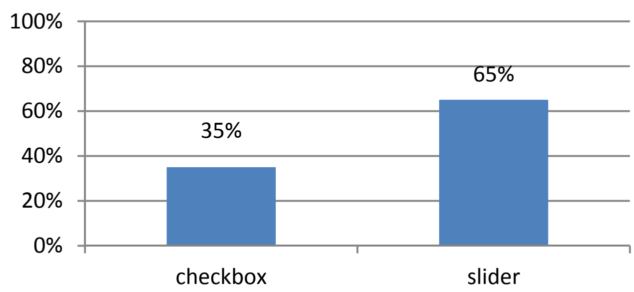


Figure 6.4 *Users' preferences on two filter panel designs*

6.6.3 Users' comments

33 users gave their comments on why they liked/disliked the two interfaces. We learned that there were 4 main reasons users preferred the slider filter. They were (1) more intuitive (6 users), e.g., "*it shows the information directly to me so that I can see the all details without doing anything, like clicking to open the details, whereas checkbox causes more actions to get the information*"; (2) more precise (4 users), e.g., "*I can choose the score more precisely compared with 'good' or 'very good' in checkbox*"; (3) easier to learn the distribution of an attribute (3 users), e.g., " *it is easy for me to pick up the 'main stream' zone*"; and (4) accessible to the attribute correlation (6 users), e.g., "*the available region of different attributes are related, showing how the change of one preference will affect the others, this eases the procedure of compromising between attributes*".

In contrast, users who expressed preference on the checkbox interface mentioned (1) "*the checkbox format is simpler to understand*" and (2) "*this way is more familiar to me*". The responses are not uncommon because the checkbox is the conventional form for filtering; some people do not like change but insist on traditional forms.

6.7 Discussion

As a summary, the user study mainly addressed two questions:

1. **Is it useful to embed opinion attributes in filters?**

   The results of the preceding formative study showed that 68.1% of participants (31/47) eliminated options in terms of customer reviews, and the proportion of users who adopted attributes extracted from customer reviews is significantly higher than that using an overall review score (26/47 vs. 5/47). Driven by these phenomena, we incorporated opinion attributes in our filter design. By recording the attributes selected in eliminating options, the results again empirically verified that in both interfaces users frequently utilized opinion attributes to narrow the range of options under consideration. The findings reinforced the important role of opinion attributes in filters.

2. **Is the modified slider presentation for filters more effective than the conventional checkbox form?**

   Our study empirically proved the superiority of the modified slider form compared with traditional checkbox filter by asking participants to perform several tasks requiring an interaction with the system. Through analysis of both objective and subjective measures, we found that the time users expended on each interface was not significantly different, but users gave higher score on the slider interface as to their perceived decision accuracy, perceived cognitive effort, pleasure and intention to return (most of the differences reached a significant level). The final interface preference again supports the hypothesis that slider filter performs better. Moreover, user comments revealed that participants are fully aware of the advantages of the modified slider filter in helping users more effectively and efficiently determining cut-off values.

Chapter 7

Experiment on

Customer-Review-Embedded Sorting

In this chapter, we compare three alternative designs that primarily differ in the way of eliciting relative importance (i.e., weight) of attributes, called direct assessment, indifference method, and indirect measurement. Online hotel booking is adopted as the test domain. Through analysis of objective and subjective measures, the multi-attribute sorting embedded with opinion attributes is verified to be beneficial to consumer online purchasing. Moreover, the direct way outperforms the indifference and indirect ways regarding perceived decision accuracy, cognitive effort, satisfaction and intention to use in an E-commerce environment.

7.1 Experiment Interfaces

The user study in chapter 6 demonstrated that the slider interface performs better in E-commerce. Hence, as the foundation of the interface design in this chapter, its filter panel is in accordance with the slider interface.

The sorting panel employs an N-item dropdown list, which supports users in sorting options by one or more attributes with low space consumption (see Figure 7.1). More concretely, users can freely select the attributes they would rely on by clicking their checkboxes. As the default, entities are sorted by price because the results of the preceding formative study revealed that customers most frequently compared alternatives on price to make their purchase decisions.

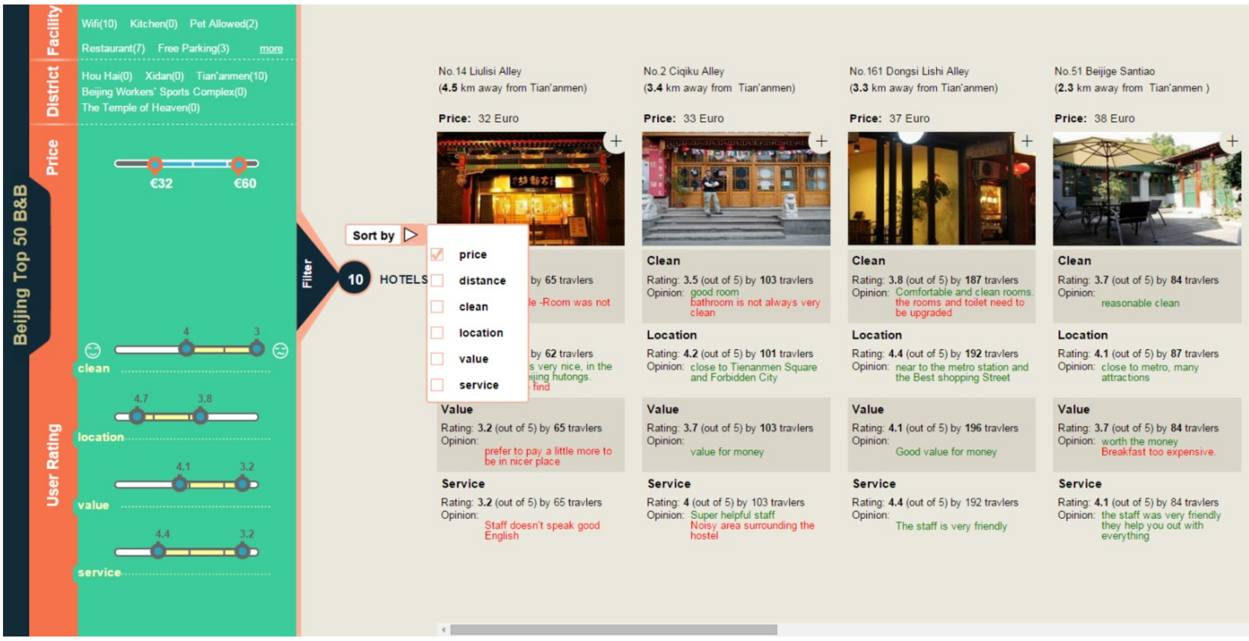
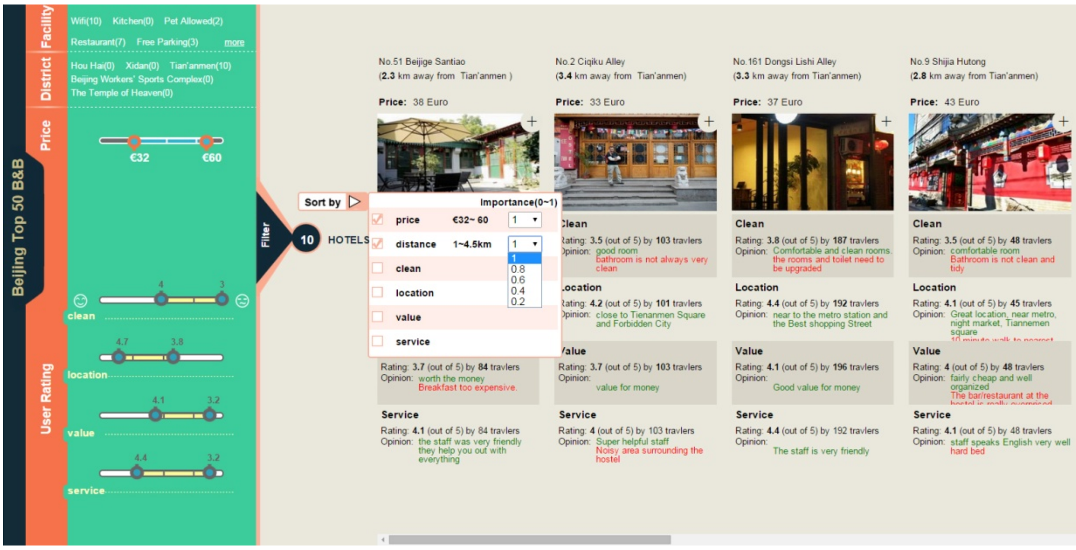
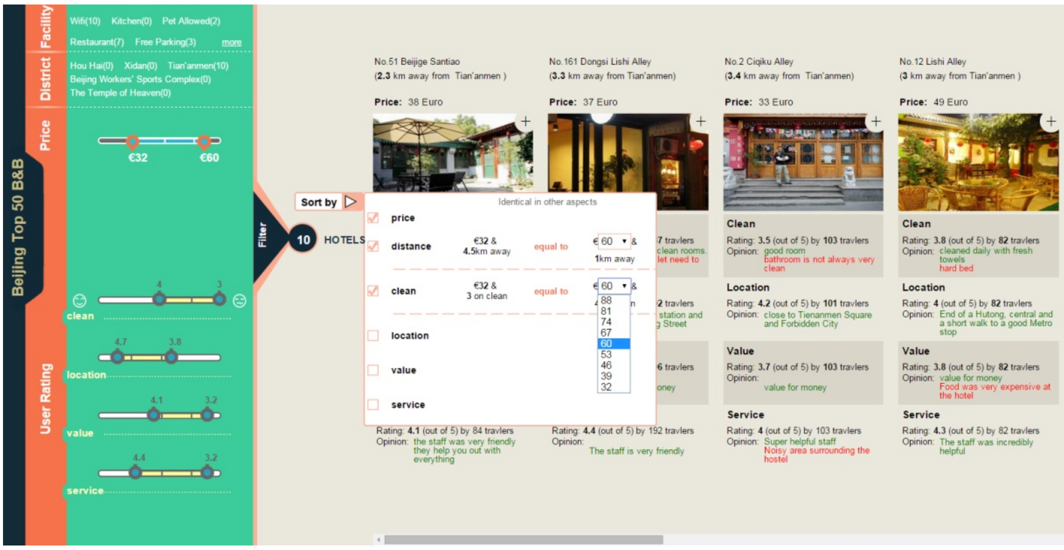


Figure 7.1 *Screenshot of the multi-attribute sorting*

When more than one attribute is selected, the questions for eliciting attribute weights are represented in three different ways: (1) direct assessment, asking users to directly assign weights to attributes (Figure 7.2 (a)); (2) indifference method, modifying one of two sets of stimuli until subjects feel that there is no difference between the two (Figure 7.2 (b)); and (3) indirect measurement, giving a preference ratio on a pair of alternatives (Figure 7.2 (c)). Users can specify their judgments by selecting from a list of interval-scaled options.



(a) Direct assessment



(b) Indifference method

(c) Indirect measurement

Figure 7.2 *Screenshot of the three ways to elicit attributes' weights*

7.2 Research Questions

We are interested in clarifying three issues through this experiment:

Although it has been known that many users compare alternatives from multiple dimensions in online shopping, existing E-commerce only supports users in sorting entities by a single attribute. This situation raises the following questions:

-Q1: Is it useful to provide multi-attribute sorting function in an E-commerce environment?

In general, the attributes with respect to customer reviews utilized in sorting are an overall user rating and the number of reviews. However, our formative study found that users depend more on opinion attributes (e.g., cleanliness) than on an overall user rating to make decisions.

-Q2: Is it beneficial to incorporate opinion attributes in multi-attribute sorting?

Concerning the three alternative designs for eliciting the relative importance of attributes, i.e., direct assessment, indifference method and indirect measurement, though it has been stated that in direct assessment the value ranges of attributes should be accompanied to avoid bias and pair-comparison outperforms individual rating for indirect measurement, previous experiments did not empirically prove which is more effective to elicit attributes' weights in an E-commerce environment.

-Q3: Which interface design is superior with respect to the effectiveness of improving consumer purchase decisions?

7.3 Measurement

Given that the objective of the experiment is to identify how users can be supported to perform informed and effective product sorting, similar to the user study in chapter 6, we assessed both objective metrics and subjective perceptions.

7.3.1 Objective metrics

More concretely, we recorded the following users' behaviors in the three interfaces:

1. Which attributes did users adopt to sort alternatives?
2. Which alternative did participants save in their shopping cart and is it the alternative located first after sorting? We calculate the percentage of times each method correctly predicted each individual's first choice.
3. How long did it take to finish each task?

7.3.2 Subjective perceptions

With the purpose of evaluating users' perceptions with the three ways of eliciting attribute weight, subjective measurement is collected from three aspects based on the definition of "usability": decision accuracy, cognitive effort and satisfaction. The associated questions are roughly the same as those in the preceding user study but with a slight modification to accommodate the different interfaces (see Table 7.1 for the details). Each question is answered on a 7-point Likert Scale ranging from 1, "strongly disagree", to 7, "strongly agree".

Table 7.1 *Questions to measure users' subjective perceptions on three sorting panel designs*

Measured aspects	Related questions
Decision accuracy	I am confident in my selection on the importance weight of different attributes/ the price to make the equation true/ the preference ratio between two hotels.
Cognitive effort	It is easy to answer the required questions to sort the hotels (e.g., selecting the importance weight/the price/the preference ration).
Satisfaction	It is pleasant to use the sorting function to rank alternatives.
	If possible, I would like to use it in the future.

7.4 Tasks and Procedure

To provide within-subject control in comparing the three designs, each participant was required to assign weights to attributes in three different ways. Considering counterbalancing, we developed six conditions showing the different interfaces in varied order (see Figure 7.3). Participants were assigned randomly to one condition.

Interface 1 –> Interface 2 -> Interface 3

Interface 1 –> Interface 3 –> Interface 2

Interface 2 –> Interface 1 –> Interface 3

Interface 2 –> Interface 3 –> Interface 1

Interface 3 –> Interface 1 –> Interface 2

Interface 3 –> Interface 2 –> Interface 1

Figure 7.3 *The order of showing three alternative interfaces*

7.4.1 Tasks

Every interface was associated with one task scenario. To avoid any bias, the three tasks were randomly assigned to the three compared interfaces. For each task, we manually retrieved 10 comparable hotels from the dataset. No hotel is dominated by another; that is, no one outperforms another in all attributes.

Task one: imagine that your parents are retired and recently planned to visit Beijing. They told you to help them book a hotel located around Tian’an men because the major attractions are nearby. Ten popular hotels near to Tian’an men are listed.

Task two: imagine that you and your wife/husband would travel to Rome for your wedding anniversary and need to book a hotel online. Considering the attractions to visit, ten hotels near the Colosseum and Trevi Fountain (wishing well) are presented.

Task three: imagine that you and your friend will travel to beijing to experience “hutong” culture. You know that the houhai area is famous for “hutong”. Thus, you plan to book a hostel near houhai. Ten popular hotels near to houhai are listed.

Please sort hotels by one or more attributes and add the most suitable hotel to the shopping cart for further consideration.

7.4.2 Procedure

Similar to the user study in chapter 6, to facilitate participants to take part in the experiment, an online procedure was created. Users’ behaviors and answers were automatically recorded in log files.

Step one: At the beginning, each participant was required to fill out his/her demographic profile and e-commerce experience (see Appendix 1 for detail).

Step two: Then, he/she was asked to imagine that he/she is in the scenario and use the assigned interface to select the most suitable one. After each task, he/she evaluated the interface that he/she just used by answering post-study questions as to if it supported him/her in making an accurate and efficient decision (see Appendix 3 for detail).

Step three: When all the tasks were completed, each participant was asked to indicate which interface he/she liked the most and the reason.

7.5 Participants

We recruited 45 new subjects to take part in the experiment. The demographic profile is basically similar to the user study in chapter six regarding internet usage, E-commerce and online hotel booking experience, without significant differences by independent t-test ($p>0.1$) (detailed demographic information is shown in table 7.2).

Table 7.2 *Demographic profile of participants in the experiment of sorting panel*

Gender	female (20); male (25)
Average age	<=20 (2); 21-30 (39); >30(4)
Major	Mechanical engineering, design, computer, etc
Internet use	4.98 (daily/almost daily), S.D.=.15
E-commerce shopping experience	3.6 (1-3 times a month), S.D.=.96
Online hotel booking experience	2.29 (1- 3 times), S.D.=.76

The average scores for 'internet usage' and' e-commerce shopping experience' are on a 5-point Likert scale from 1, 'least frequent', to 5, 'very frequent', and 'online hotel booking experience' is on a 3-point Likert scale.

7.6 Analysis of Results

Because the same participants took part in all three tasks of the experiment, we analyze whether there are significant differences between the three interfaces by repeated-measures ANOVA or Friedman's test depending on whether the sampling distributions within groups are normal [28].

7.6.1 Objective measures

With respect to the time users on average expended on each interface, the results show that direct assessment took on average 110.23s (S.D.= 155.04) for users to sort alternatives until making a choice, while indifference too k 92.47s (S.D.= 94.96) and indirect-measure took 108.01s (S.D.= 85.62). Mauchly's test indicated that the assumption of sphericity (i.e., both the variances across conditions and the covariances between pairs of conditions are equal) had been violated, $\chi^2(2)=9.51$, $p<.05$, therefore multivariate tests are reported ($\varepsilon=.83$). The results show that there is no significant difference between them, V=.04, $F(2, 43)=.98$, $p>.05$, as shown in Table 7.3.

Table 7.3 *Comparison on task time with three sorting panel designs*

Interface	Mean(S.D.) (second)	Repeated-measures ANOVA				
		Value	F	Hypothesis df	Error df	*P*
direct	110.23 (155.04)					
indifference	92.47 (94.96)	.04	.98	2	43	**.382**
indirect	108.01 (85.62)					

To further understand how users actually behaved in the three interfaces, we measured the attributes they used to sort options. On the whole, 73.3%, 73.3% and 80% of participants sorted options by more than one attribute in the direct, indifference and indirect interfaces, respectively. Users on average sorted by 3.33 attributes (1.53 static features &. 1.80 opinion attributes) in the direct-assessment; 3.27 attributes (1.58 static features & 1.69 opinion attributes) in the indifference-method; and 3.09 attributes (1.49 static features & 1.60 opinion attributes) in the indirect-measurement. The number of attributes used in sorting is not significantly affected by the interfaces, V=.03, F (2, 43) =.68, p>.05 (see Table 7.4). Hence, it can be inferred that the multi-attribute sorting function assists and stimulates users in considering trade-offs among attributes.

Table 7.4 *Comparison on the number of attributes used to sort options*

Interface	Mean(S.D.)	Repeated-measures ANOVA				
		Value	F	Hypothesis df	Error df	*P*
direct	3.33 (2.00)					
indifference	3.27 (1.89)	.03	.68	2	43	.514
indirect	3.09 (1.62)					

Figure 7.4 represents the percentage of people who sorted by static features, opinion attributes or both in the three interfaces. Overall, 72%, 62% and 67% of users referred to opinion attributes in the three interfaces. The high reference frequencies indicate the benefit of incorporating opinion attributes in sorting.

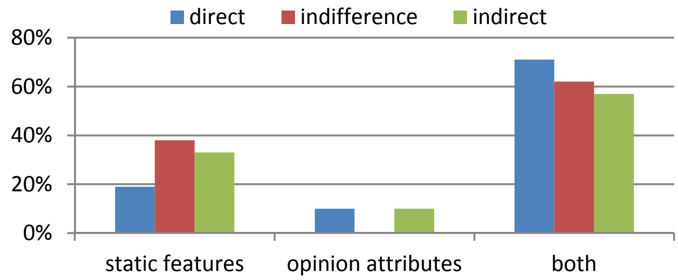


Figure 7.4 *Attributes used to sort options*

The proportions of users who actually did choose the predicted option as the most suitable choice were 62%, 38% and 52% in the three interfaces (see Figure 7.5). We can see that the direct interface most correctly predicted each individual's choice, followed by the indirect interface; the indifference interface led to the worst prediction.

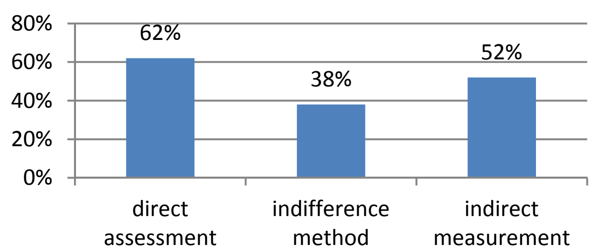


Figure 7.5 The frequency of correct prediction

7.6.2 Subjective measures

We then analyzed users' answers to the post-study questionnaire. As described above, each question was answered on a 7-point Likert scale (from 1"strongly disagree" to 7 "strongly agree"). Table 7.5 lists users' average scores and the significance analysis results.

We can observe that users gave more positive score to the direct assessment than to both indifference method and indirect measurement for all four questions, including their perceived decision accuracy (Q1), cognitive effort (Q2), pleasantness to use (Q3) and intention to return (Q4). The statistical analysis shows that the type of interface significantly affects the perceived accuracy (Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2)$ =10.30, $p<.05$, therefore multivariate tests are reported (ε =.82), F (2, 43) =4.26, $p<.05$), cognitive effort (F (2, 88) =6.00, $p<.05$), pleasantness to use (F (2, 88) =9.82, $p<.05$) and satisfaction (F (2, 88) =8.21, $p<.05$).

The pairwise comparisons further show that direct interface obtained significantly higher score than indirect interface regarding Q1 "perceived accuracy" and significantly higher scores than both indirect measurement and indifference method regarding Q2 "cognitive effort", Q3 "pleasantness to use" and Q4 "intention to return".

Table 7.5 *Comparison on users' subjective perceptions on three sorting panel designs*

Interface		**Mean (S.D.)**	**Repeated-measures ANOVA**	
			F	*P*
Q1. I am confident in my selection on the importance weight of different attributes/ the price to make the equation true/ the preference ratio between two hotels. (perceived accuracy)				
Direct		**5.64 (1.25)**		
Indifference		5.31 (1.35)	F (2, 43) =4.26	**.020**
Indirect		5.22 (1.28)		
(I) interface	(J) interface	Mean difference (I-J)	Std. Error	*P*
Direct	Indifference	.33	.21	.363
	Indirect	.42	.14	**.016**
Q2. It is easy to answer the required questions (i.e., selecting the importance weight/the price/the preference ration) to sort the hotels. (perceived effort)				
Direct		**5.69 (1.28)**		
Indifference		5.11 (1.37)	F (2, 88) =6.00	**.004**
Indirect		5.24 (1.45)		

Interface		Mean (S.D.)	Repeated-measures ANOVA	
			F	*P*
(I) interface	(J) interface	Mean difference (I-J)	Std. Error	*P*
Direct	Indifference	.58	.18	**.007**
	Indirect	.44	.17	**.036**
Q3. It is pleasant to use the sorting function to rank alternatives.				
Direct		**5.87 (1.27)**		
Indifference		4.98 (1.55)	F (2, 88) =9.82	**.000**
Indirect		5.22 (1.44)		
(I) interface	(J) interface	Mean difference (I-J)	Std. Error	*P*
Direct	Indifference	.89	.21	**.000**
	Indirect	.64	.22	**.014**
Q4. If possible, I would like to use it in the future.				
Direct		**5.91 (1.40)**		
Indifference		5.31 (1.47)	F (2, 88) =8.21	**.001**
Indirect		5.36 (1.45)		
(I) interface	(J) interface	Mean difference (I-J)	Std. Error	*P*
Direct	Indifference	.60	.15	**.001**
	Indirect	.56	.18	**.010**

In the post-study questionnaire, users were asked to choose the interface they preferred. 68.9% of users (31 out of 45) favored direct assessment, followed by 20% (9 out of 45) who liked indirect measurement and 11.1% (5 out of 45) who preferred indifference method (see Figure 7.6). With Chi-square test, the difference is significant ($\chi^2(2)$ =26.13, p<.05). The findings again confirm the superiority of direct interface.

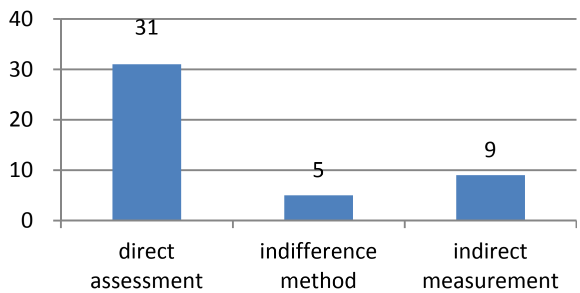


Figure 7.6 *Users' preferences on three sorting panel designs*

7.6.3 Users' comments

37 users gave their comments on the reasons why they liked/disliked the three interfaces. We learned that the reasons associated with direct assessment are mainly that "*it is clear to understand how to operate*", "*it is easy to make a judgment*", and "*it is simple and convenient*". The reasons some users liked indirect measurement are that "*it is good to learn preference through hotel comparison*" and "*it caters to the desire for novelty*". In comparison, most users expressed weak points related to indifference method, including "*having no confidence in the price*", "*Difficulty in giving the answers. I was not able to answer in a meaningful way how much I do pay for 'location=3.8' compared with '€ 54and location 4.8'*".

7.7 Discussion

Here, we answer the questions that we proposed at the beginning of the study.

1. **Is it useful to provide multi-attribute sorting in E-commerce?**

   The conventional sorting panel in current E-commerce allows users to sort alternatives by a single attribute that is thought of as the most important. Though it was recognized that some people explicitly consider trade-offs among values instead of using a non-compensatory heuristic, few works have been devoted to facilitating this process. In our study, it was found that more than 80% of users sorted options by multiple attributes rather than a single attribute when they were provided with such support.

2. **Is it beneficial to incorporate opinion attributes in sorting?**

   The majority of users, i.e., more than 60%, sorted options by both static features and opinion attributes. In addition, the average number of opinion attributes utilized in sorting was more than that of static features.

3. **Which way performs better to elicit the relative importance of attributes in online purchasing?**

   Though there are different ways to elicit the weights of attributes, few works have empirically identified which interface is superior in the environment of online shopping.

Through comparison of three interfaces, we found that the direct assessment is more effective, whereas the indifference method is the least informative. The indirect measurement turns out to be in the middle. Objectively, direct interface correctly predicted individual's first choice more times compared to indifference and indirect interfaces. Subjectively, users gave higher score on direct interface in terms of perceived decision accuracy, cognitive effort, pleasantness and intention to use. The final interface preference again verifies the advantage of direct assessment in an online environment.

Chapter 8

Conclusion

Motivated by the results of a series of user studies, we have derived a set of guidelines on how to design interfaces for user decision improvement in E-commerce. Then, we point out the limitations in our research and propose several relevant research topics that we can work on in the future.

8.1 Design Guidelines

In the following, the design guidelines will be elaborated in terms of three primary interfaces: the interface for screening out interesting alternatives, the interface for evaluating alternatives in detail and the interface for comparing candidates.

8.1.1 The interface for screening out interesting alternatives

The results of the formative study show that in the process of screening out interesting alternatives, 94% of participants began by narrowing down the set of options in terms of Eliminate-by-aspect (EBA) strategy. The majority of them (55.3%) eliminated alternatives by both static features and customer reviews. Moreover, significantly more users employed opinion attributes (e.g., cleanliness and location) to execute an attribute-driven action, compared to those using an overall user rating.

In the experiment of comparing two alternative designs for an opinion-attribute-embedded filter panel, we recorded the attributes users adopted to eliminate options. The results show that participants depended highly on opinion attributes as well as static features in both interfaces, which again demonstrated the benefit of incorporating opinion attributes in filters. Therefore, we propose the following design guideline:

Guideline 1: *Include both static features and opinion attributes in filter panel*

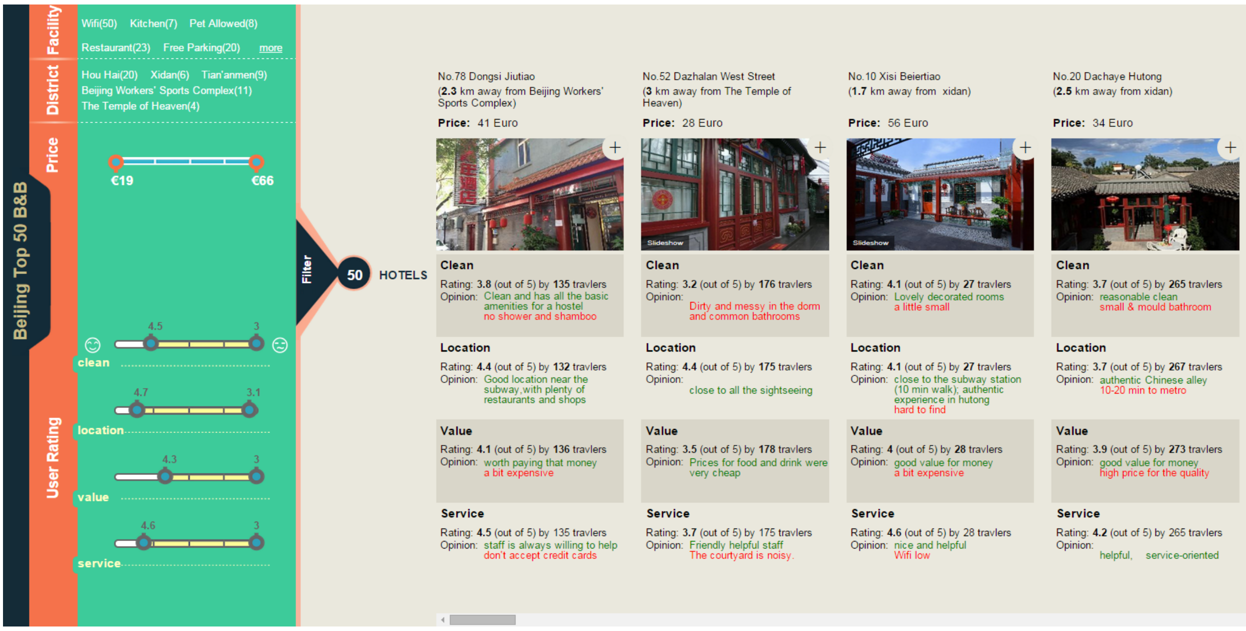
The generation of cut-off values does not always follow stable preference, but is contingent on the characteristics of tasks. Grounded in the results of the formative study, we notice that the cut-off value of an attribute is influenced by its value distribution and correlation with other attributes, in addition to user’s stable favor. More concretely, when determining the cut-off value of an attribute, users often browse its value distribution to avoid invalid filter, such as no/too many options available due to strict/loose restraints. Meanwhile, users who explicitly consider trade-offs among attributes (i.e. EBA+ADDIF) more frequently refer to the correlation among attributes for determining cut-off values. For example, someone explores a hotel with a price above his/her original price limit just to see how much better it is. If it greatly exceeds expectation the cut-off may be shifted, whereas, the cutoff is reinforced. We hence conclude the guideline:

Guideline 2: *To help users effectively determine cut-off values, provide an easy access to the value distribution of each attribute and the correlation among attributes.*

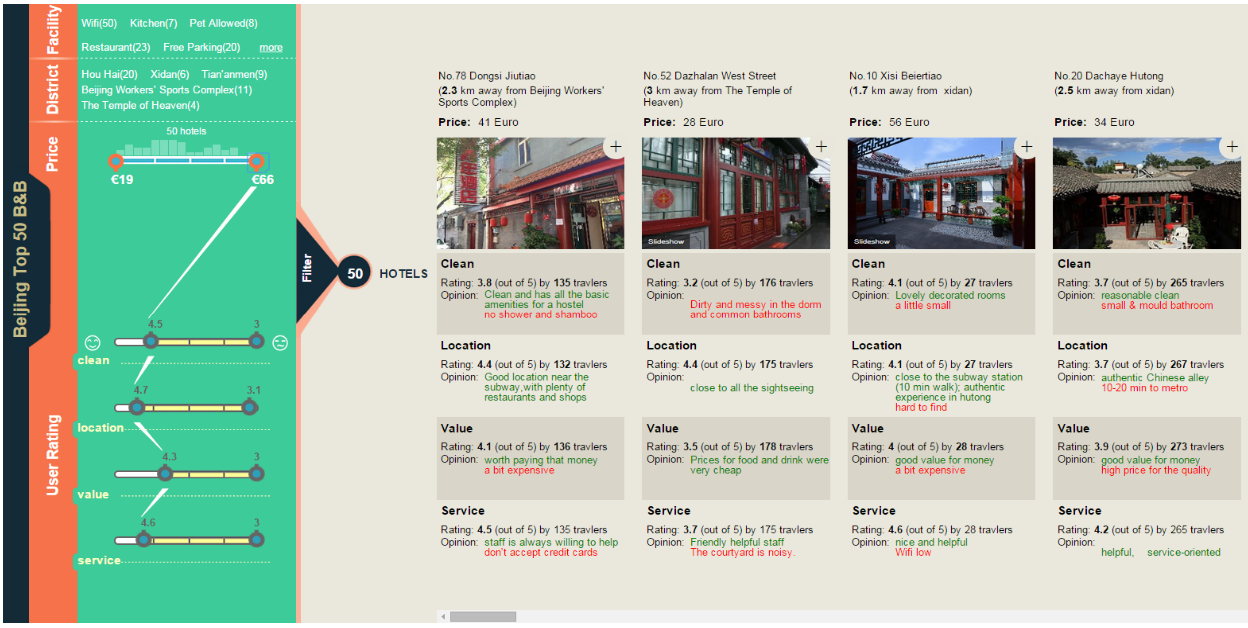
Based on the findings of the formative study, we developed two alternative designs for an opinion-attribute-embedded filter panel, respectively called checkbox interface and slider interface. In checkbox interface, the filter for each attribute is represented in the form of an array of N checkboxes, which is utilized in most E-commerce websites. Each checkbox standing for a certain value range on one dimension is followed by the number of entities with values within the range. In slider interface, users can narrow down the displayed subset by dragging the knobs of each slider. Moreover, when hovering over or moving the knob of a slider, the value distribution of the attribute is visualized via bars and the correlation among attributes is represented via simultaneous change (i.e., when the range of one attribute is changed, the corresponding knobs of other attributes and connecting lines will simultaneously move). Then, we conducted a user study comparing the two interfaces by asking users to perform several tasks requiring an interaction with the interfaces. The results show that the slider interface can stimulate and facilitate users to consider trade-offs among attributes and achieves significantly higher user assessments in terms of perceived decision accuracy, perceived ease of use, satisfaction and intention to use. In terms of these results, we propose that:

Guideline 3*: The filter for a numerical attribute can be represented in the form of modified slider, which visualizes the value distribution via bar charts and correlation among attributes via simultaneous change*

The above three design guidelines with respect to filter panel are applied in online hotel booking as an example (see Figure 8.1).



(a) Mouse-out



(b) Mouse-over

Figure 8.1 *Example of filter panel design*

In the formative study, we notice that 24% (12/50) of participants selected alternatives with the best value on a single attribute (i.e. Lexicographic); whereas, 40% of participants (20/50) turned to more compensatory strategy (i.e. Weighted Additive Difference) after narrowing down alternatives to a smaller set. In other words, they compared the remaining alternatives on multiple attributes and then selected the alternative with the best weighted additive value. More notably, both participants who selected alternative by a single attribute and those who made their decisions by multiple attributes depended highly on opinion attributes. Here we suggest the following two design guidelines:

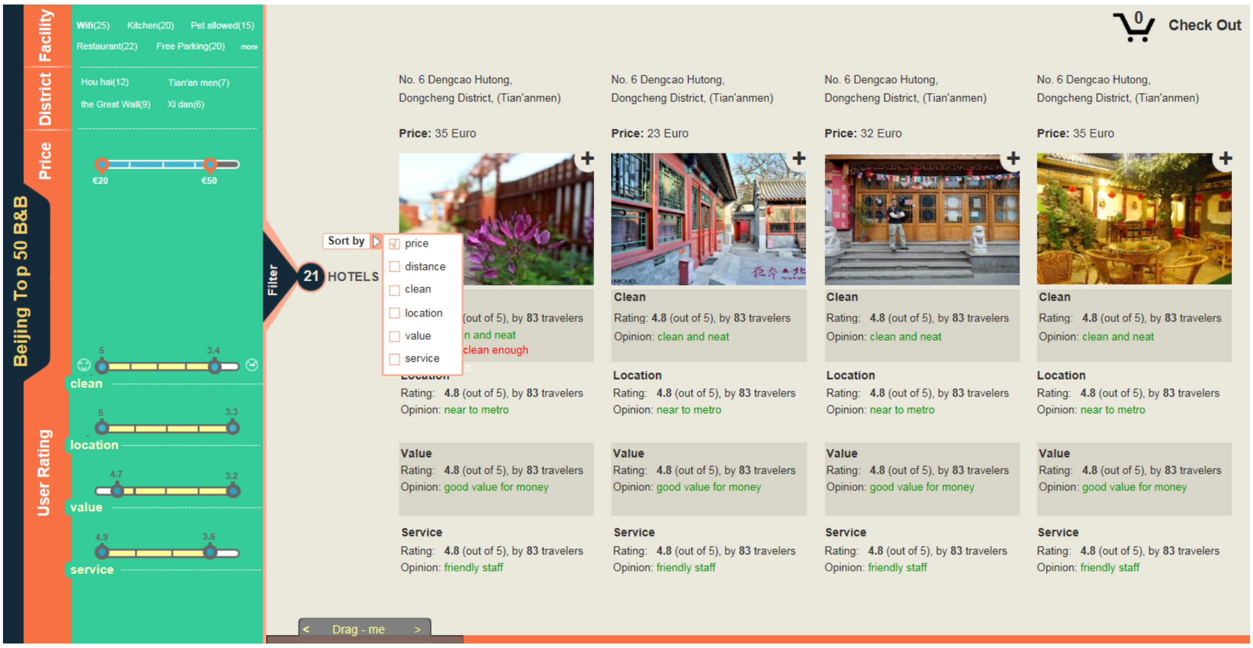
Guideline 4: *Enable users to sort alternatives by multiple attributes and provide the access to give different weights to attributes.*

Guideline 5: *In addition to static features, integrate opinion attributes into sorting panel.*

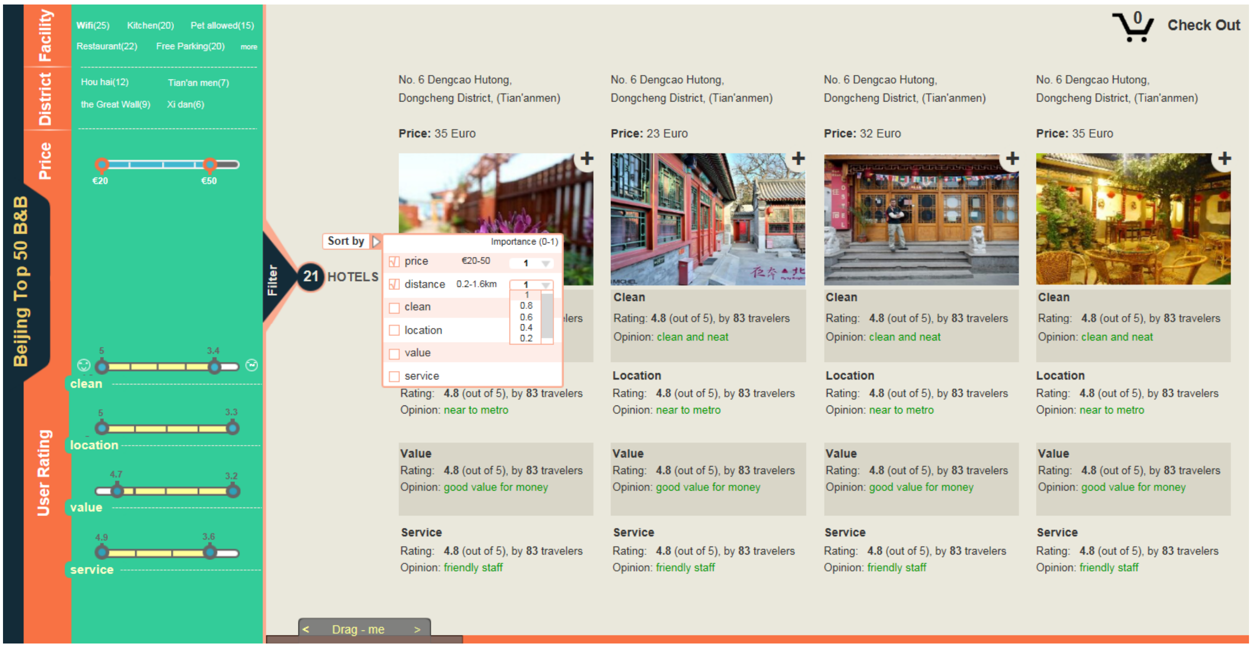
Grounded in these user decision making behaviors, we developed multi-attribute sorting panel embedded with opinion attributes. Furthermore, when more than one attribute is selected, the multi-attribute sorting panel expands to three alternative designs that mainly differ in the way of eliciting relative importance of attributes, (1) direct assessment: asking users to directly assign weights to attributes, (2) indifference method: modifying one of two sets of stimuli until subjects feel that there is no difference between the two, and (3) indirect measurement: giving relative preference on pairs of alternatives. In the user study, participants were required to sort options by one or more attributes and select the most suitable alternative with the assistance of three designs. Through analysis of objective and subjective measurements, the effectiveness of multi-attribute sorting function was verified, whereby, on average, more than 80% of participants sorted options by more than one attribute in three interfaces. Users gave significantly higher rating to the direct assessment than to indifference method and indirect measurement regarding perceived decision accuracy, cognitive effort, satisfaction and intent to use.

Guideline 6: *For different ways of eliciting relative importance of attributes, directly asking users to assign weights performs better, regarding improving decision accuracy, effort and satisfaction, in E-commerce environment.*

As shown in Figure 8.2, we apply the three design guidelines concerning sorting panel in online hotel booking.



(a) selecting a single attribute to sort options



(b) Selecting multiple attributes to sort options

Figure 8.2 *Example of sorting panel design*

8.1.2 The interface for evaluating alternatives in detail

In the second task of the formative study, each participant was required to decide whether to save an entity as a purchase candidate by evaluating its detailed information. The analysis of the results show that participants were mainly concerned with three types of information: static features, item sample and customer reviews. More concretely, the number of participants reading reviews in a feature-driven manner is significantly larger than the number of participants doing so in a holistic manner, $\chi^2(1)=23.27$, $p<.05$. We come up with the following design suggestion:

Guideline 7: *The information in detail page should include static features, item sample and customer reviews. Customer reviews should be categorized according to the features of the item being rated.*

For each opinion attribute, users firstly inspected its numerical values to gain a quick understanding on how much it is liked or disliked. In the formative study, 24% and 39% of participants evaluated the average rating and average rating plus the number of reviews, respectively. Whereas, due to that opinion often changes over time, the other 37% also read the time distribution of all and the 5-point customer reviews to examine whether there is a download trend for customer reviews.

Guideline 8: *In addition to the average rating and number of reviews for each opinion attributes, represent the time distribution of customer reviews to support the analysis of temporal evolution.*

In addition to numerical values of each opinion attribute, participants inspect its verbal values to get information about why that rating was given. In the formative study, 7%, 56.5% and 36.5% of participants referred to adjective-noun word pairs, raw reviews and both, respectively. Overall, 93% of participants read raw reviews (i.e. the textual comments) to assist in the context understanding behind positive/negative reviews. Due to the large quantity of raw reviews, participants performed two types of behavior: seeking the latest and/or the most negative reviews.

Guideline 9: *Couple numerical values with verbal values (i.e., adjective-noun word pairs and raw reviews) for each opinion attribute. And facilitate users to inspect the latest and most negative reviews.*

8.1.3 The interface for comparing candidates

At the third stage of online purchasing, users compare the selected candidates to confirm the final choice. There are three types of information that participants would compare at this stage: (1) static attributes, (2) item sample and (3) customer reviews. The formative study show that participants who adopted a compensatory strategy at stage one, i.e. EBA+ADDIF, focused significantly more on customer reviews (χ^2 (1) =16.59, p<.001) and less on photos (χ^2 (1) =7.34, p<.01) compared to participants who adopted non-compensatory strategies, i.e. EBA, LEX and EBA+LEX. More notably, significantly more participants used {opinion attribute, sentiment} pairs extracted from customer reviews to perform feature-by-feature and side-by-side comparison between products, compared to those merely referring to an overall review score, χ^2 (1) =10.7, p<.001.

Guideline 10: *In comparison interface, present static feature, item sample and customer reviews for each entity. Moreover, customer reviews should be summarized in the form of {opinion attribute, sentiment} to facilitate attribute-driven comparison.*

With respect to opinion attributes, the majority of participants made their decisions based on numerical values (e.g. the average rating and number of reviews), followed by a combination of numerical and verbal values. The smallest proportion relied on only verbal values (i.e., adjective-noun word pairs).

Guideline 11: *Represent both numerical values and verbal values toward each opinion attribute.*

To confirm the final choice, consumers calculate value differences on multiple attributes across candidates and sum over the differences to select the alternative with the best overall evaluation. Because time and self-reported effort of executing a choice strategy can be estimated in terms of the time and effort associated with each elementary information process, so decision effort can be reduced by facilitating the calculation and add of value differences.

Guideline 12: *Enable users to easily calculate and sum value differences on each attribute across alternative.*

8.2 Limitations

1. Ecological validity

In research, the ecological validity of a study means that the methods, materials and setting of the study must approximate the real-world that is being examined [96]. In our formative study that aims to formally detect users' behaviors, all relevant information is hidden in boxes until the subject moves mouse to click, which is less similar to real-world process. However, the computerized process tracing technique is not likely to have substantial influence on the research questions we investigated.

2. Generalization

In our research, design solutions are applied in online hotel booking. However, depending on the different importance and irreversibility of decision task, the relative weight placed in various goals (maximizing accuracy and minimizing effort) is different, so decision making behavior can vary greatly. For high-risk task, such as purchasing a house or choosing a spouse, decision maker will give more weight on accuracy, so that referring to extensive amount of information and applying more compensatory decision strategies in order to make a decision as best as possible. On the contrary, for choice of a book or movie involving a negligible amount of risk, decision maker tends to spend less effort on it, with more selective and non-compensatory decision strategies. Further experiment is required to ascertain the boundary to generalize our results beyond the sample domain by including more product domains.

For each experiment, we recruited around 50 participants. Because from the perspective of statistics, the sampling distribution should be normal if the samples contain more than about 50 scores. However, in the field of E-commerce, the number of participants is still limited. Another limitation of participants is the diversity. Most of the participants are university students, although we tried to recruit them from as many different departments as possible. It is ideal to recruit users with diverse ages, educations and professions.

8.3 Future Work

1. Concrete interface design solutions

In this thesis, we put forward concrete design solutions to the interface for screening out interesting alternatives in the context of online hotel booking. Although some design guidelines have been generated grounded in user decision-making behavior, we did not translate them into specific design solutions with respect to the interfaces for evaluating alternatives in detail and comparing candidates. Chen *et al.* (2014) designed three alternative sentiment-embedded comparison interfaces based on popular techniques. The results of a user study show that opinion bar chart, that visualizes numerical sentiment via bars and verbal sentiment via tool tip, achieves significantly higher user assessments compared to opinion table and opinion cloud. [18]

2. Interface design for review writing

While 90 percent of shoppers use online reviews to make purchase decisions, only 6 percent actually write reviews. The propensity to post online reviews is low [95]. Beenen *et al.* (2004) found that most consumers are social loafers rather than contributor [7]. Therefore, how to proactively induce customers to “spread the word” and providing high-quality reviews is emerging as a question.

With the purpose to reduce review entry effort and motivate user to write high-quality reviews, Dong *et al.* (2011) and Bridge *et al.* (2011) developed systems that provide authors real-time suggestions on what they may wish to write in terms of previous reviews [24, 11]. The results of user trial show that users feel they get support while writing reviews and would like to incorporate the suggestions into reviews.

However, little attention has been devoted to the antecedents of online product review, i.e., the forces that motivate consumers to write online reviews. According to prior research [23,84,36], motives of word-of-mouth (WOM) communication consist of four main categories: (1) product involvement, consumers feel urge to talk about for very good or very bad products that trigger strong feelings; (2) self-enhancement, people engage in review writing to show connoisseurship and gain attention; (3) concern for others, consumers feel genuine need to help others make better decision; (4) economic incentives offered by website for posting online reviews. Dellarocas and Narayan (2011) paid special attention on how properties of online medium affect reviewers' willingness to review a product online [21]. They found that there is a positive relationship between the propensity to review a movie and the level of disagreement among professional reviews, which demonstrated the validity of self-enhancement motive and concern for others.

Appendix 1: Pre-study questionnaire

1. Name: ___________________

2. Age

   (A) <= 20 (B) 21 – 30 (C) 31 – 40 (D) >= 41

3. Gender

   (A) Male (B) Female

4. Major: ___________________

5. How often do you use the Internet?

   (A) Daily/almost daily

   (B) At least 1 time per week

   (C) At least 1 time per month

   (D) Just a few times overall

   (E) Never

6. How often is your online shopping experience?

   (A) Very often (1-3 times a week)

   (B) Often (1-3 times a month)

   (C) Occasionally (a few times every 3 months)

   (D) Rarely (just a few times overall)

   (E) Never

7. How many times have you booked hotel/hostel online?

   (A) More than 3 times

   (B) 1-3 times

   (C) Never

Appendix 2: Post-study questionnaire for user study one

1. I am confident that the subset of hotels I just got by means of filtering is the best choice for me.

   strongly disagree **1** **2** **3** **4** **5** **6** **7** strongly agree

2. I easily obtained and processed relevant information to narrow down the hotels.

   strongly disagree **1** **2** **3** **4** **5** **6** **7** strongly agree

3. It is pleasant to use the filtering function to narrow down alternatives.

   strongly disagree **1** **2** **3** **4** **5** **6** **7** strongly agree

4. If possible, I would like to use it in the future.

   strongly disagree **1** **2** **3** **4** **5** **6** **7** strongly agree

Appendix 3: Post-study questionnaire for user study two

1. I am confident in my selection on the importance weight of different attributes/ the price to make the equation true/ the preference ratio between two hotels.

   strongly disagree **1** **2** **3** **4** **5** **6** **7** strongly agree

2. It is easy to answer the required questions (i.e. selecting the importance weight/the price/the preference ration) to sort the hotels.

   strongly disagree **1** **2** **3** **4** **5** **6** **7** strongly agree

3. It is pleasant to use the sorting function to rank alternatives.

   strongly disagree **1** **2** **3** **4** **5** **6** **7** strongly agree

4. If possible, I would like to use it in the future.

   strongly disagree **1** **2** **3** **4** **5** **6** **7** strongly agree

REFERENCES

1. Abelson R. P. and Levi, A. Decision making and decision theory. In G. Lindsay & E. Aronson (Eds.), *Handbook of social psychology (3rd Ed.)*, pp. 231–309, Random House, New York, 1985.

2. Ahlberg C., Williamson C., and Shneiderman B. Dynamic queries for information exploration: an implementation and evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '92,* pp. 619–626, 1992. ACM.

3. Ahlberg C. and Shneiderman B. N. Visual information seeking using the FilmFinder. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '94,* pp. 433-434, 1994. ACM.

4. Bakeman, R. Behavioral observation and coding. In H. T. Reis & C. M. Judge (Eds.), *Handbook of research methods in social and personality psychology,* pp. 138-159, Cambridge University Press, 2000.

5. Bartram L. and Ware C. Filtering and brushing with motion. *Information Visualization,* 1 (1): 66–79, 2002.

6. Becker R. A. and Cleaveland W.S. Brushing scatterplots. *Technometrics* 29 (2): 127–142, 1987.

7. Beenen G., Ling K., Wang X. Chang K., Frankowski D., Resnick P. and Kraut R.E. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4): 00, 2005.

8. Bettman J. R., Johnson E. J., and Payne J. W. A Componential Analysis of Cognitive Effort in Choice. *Organizational Behavior and Human Decision Processes,* 45 (1): 111-139, 1990.

9. Bettman J. R., Luce M. F., and Payne J. W. Constructive Consumer Choice Processes. *Journal of Consumer Research,* 25 (3): 187-217, 1998.

10. Birnbaum M. H., Coffey G., Mellers B. A. and Weiss, R. Utility Measurement: Configural-Weight Theory and the Judge's Point of View. *Journal of Experimental Psychology: Human Perception and Performance,* 18: 331–346, 1992.

11. Bridge D. and Healy P. The ghostwriter-2.0 case-based reasoning system for making content suggestions to the authors of product reviews. *Journal of Knowledge-Based Systems,* 29: 93-103, 2011.

12. Card S. K., Moran T. P., and Newell A. *The psychology of human–computer interaction.* Lawrence Erlbaum, Hillsdale, February 1986.

13.Carenini G., Mittal, V., and Moore, J. Generating patient-specific interactive natural language explanations. *Proc Annu Symp Comput Appl Med Care,* pp. 5–9, 1994.

14.Carenini G., Ng R.T., and Pauls A. Interactive multimedia summaries of evaluative Interfaces text. In *Proceedings of the 11th International Conference on Intelligent User interfaces (IUI '06)*, pp. 124–131, 2006. ACM.

15.Carenini G. and Rizoli L. A multimedia interface for facilitating comparisons of opinions. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI '09)*, pp. 325–334, 2009. ACM.

16.Chen L. PhD thesis: *User decision improvement and trust building in product recommender systems*. EPFL, Lausanne, Switzerland, 2008.

17.Chen L. Towards Three-Stage Recommender Support for Online Consumers: Implications from a User Study. *Web Information Systems Engineering – WISE 2010,* 6488: 365-375, 2010.

18.Chen L., Wang F., Qi L. and Liang F. Experiment on sentiment embedded comparison interface. *Journal of Knowledge-Bases Systems*, 64: 44-58, 2014.

19.Cooper A., Reimann R. and D. Cronin. *About Face 3: the essentials of interaction design (*3rd Ed.). Wiley publishing, Indiana, May 2007.

20.Coupey E. Restructuring: Constructive processing of information displays in consumer choice. *Journal of Consumer Research,* 21(1): 83-99, Jun. 1994.

21.Dellarocas C. and Narayan R. What motivates consumers to review a product online? A study of the product-specific antecedents of online movie reviews. In *WISE*, pp: 1-6, 2006.

22.Dewey M. E. Coefficients of agreement. *British Journal of Psychiatry,* 143: 487-489, 1983.

23.Dichter E. How word-of-mouth advertising works. *Harvard Business Review*, 44 (Nov-Dec): 147-166, 1966.

24.Dong R., McCarthy K., O'Mahony M., Schaal M. and Smyth B.Towards an intelligent reviewer's assistant: Recommending topics to help users to write better product reviews. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces IUI'12*, pp. 159-168, 2011.

25.Feldman J. M. and Lynch J. G. Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intentions, and Behavior. *Journal of Applied Psychology,* 73 (3): 421–435, 1988.

26.Felfernig, A., and Gula, B. Consumer behavior in the interaction with knowledge-based recommender applications. In *ECAI 2006Workshop on Recommender Systems*, pp. 37–41, 2006.

27.Felferig A., Friedrich G., Gula B., Hitz M., Kruggel T., Melcher R., Riepan D., Strauss S., Teppan E., and Vitouch O. Persuasive recommendation: serial position effects in knowledge-based recommender systems. In *Proceedings of the 2nd International Conference of Persuasive Technology* (Persuasive ’07), pp. 283–294, 2007.

28.Field A. *Discovering Statistics using IBM SPSS Statics (4th Ed.)*. SAGE Publications Ltd, London, January 2013.

29.Fischer G. W. Range Sensitivity of Attribute Weights in Multiattribute Value Models. *Organizational Behavior and Human Decision Processes*, 62 (3): 252-266, 1995.

30.Ford J. K., Schmitt N., Schechtman S. L., Hults B. M., and Doherty M. L. Processing tracing methods: Contributions, problems and neglected research questions. *Organizational Behavior and Human Performance,* 43 (1): 75–117, 1989.

31.Ghose A. and Ipeirotis P. G. Designing Novel Review Ranking Systems: Predicting Usefulness and Impact of Reviews, In *Proceedings of the Ninth International Conference on Electronic Commerce ICEC'07*, pp. 303-309, 2007. ACM.

32.Goldstein W. M. Judgments of relative importance in decision making: Global vs. local interpretation of subjective weight. *Organizational Behavior and Human Decision Processes,* 47 (2): 313-336, 1990.

33.Goodwin K. *Designing for the Digital Age: How to Create Human-Centered Products and Services*. Wiley, March 2009.

34.Gray W.D., Sims C.R., Fu W.T. and Schoelles M.J. The soft constraints hypothesis: a rational analysis approach to resource allocation for interactive behavior. *Psychological reviews*, 113(3): 461, 2006.

35.Haubl G. and Trifts V. Consumer decision making in online shopping environments: the effects of interactive decision aids. *Marketing Science,* 19 (1): 4–21, 2000.

36.Hennig-Thurau T., Gwinner K.P., Walsh G. and Gremler D.D. Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?. *Journal of interactive marketing*, 18(1): 38-52, 2004.

37.Hogarth R.M. *Judgment and choice (2nd Ed.).* Wiley, New York, 1987.

38.Holten D., and van Wijk J.J. A user study on visualizing directed edges in graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI’ 09* pp. 2299–2308, 2009.

39.Hu N., Pavlou P. A., and Zhang J. Can online reviews reveal a product’s true quality? Empirical findings and analytical modeling of online word-of-mouth communication. In *proceedings of the 7th ACM conference on electronic commerce (EC’06),* pp. 324-330, 2006.

40.Hu M., Yang H., Zhou M., Gou L., Li Y., and Haber E. Opinion Blocks: a crowd-powered, self-improving interactive visual analytic system for understanding opinion text. In *Proceedings of INTERACT '2013*, pp. 116—134, 2013.

41.Huang J., Etzioni O., Zettlemoyer L., Clark K., and Lee C. RevMiner: an extractive interface for navigating reviews on a Smartphone. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pp. 3-12, New York, NY, USA, 2012.

42.Huang S.-W., Tu P.-F., Fu W.-T., and Amamzadeh M. Leveraging the crowd to improve feature-sentiment analysis of user reviews. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces (IUI '13)*, pp. 3–14, 2013. ACM.

43.Huber J. and N. M. Klein. Adapting cutoffs to the choice environment: the effects of attribute correlation and reliability. *Journal of Consumer Research*, 18: 346-357, 1991.

44.Huber J., Wittin D. R., Fiedler J. A. and Miller, R. The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice. *Journal of Marketing Research*, 30: 105–114, 1993.

45.J. Johnson. *Web Bloopers: 60 Common Web Design Mistakes and How to Avoid Them (*1st Ed.). Morgan Kaufmann, San Francesco, April 2003.

46.Keeney, R. L., and Raiffa, H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, Cambridge, July 1993.

47.Lakoff G., and Johnson M. *Metaphors we live by.* University of Chicago Press, 1980.

48.Lee Y. E. and Benbasat I. Interaction design for mobile product recommendation agents: Supporting users' decisions in retail stores. *ACM Transactions on Computer-Human Interaction,* 17 (4): 1-32, Dec. 2010.

49.Lehto R. M. and Nah F. Decision-making models and decision support. In Salvendy G. (Ed.), *Handbook of Human Factors and Ergonomics (3rd Ed.),* pp. 191-242, Wiley, 2006.

50.Lelis S. and Howes A. Informing decisions: how people use online rating information to make choices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '11*, pp. 2285-2294. 2011. ACM.

51.Levin I., Schneider S., and Gaeth G. All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76 (2): 90–98, 1998.

52.Liu B., Hu M., and Cheng J. Opinion observer: analyzing and comparing opinions on the Web. In *Proceedings of the 14th International Conference on World Wide Web*, pp. 342–351, 2005. ACM.

53.Lohse G. L. and Johnson E. J. A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, 68 (1): 28-43, 1996.

54.Lombard M., Snyder-Duch J. and Bracken C. C. Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects, http://matthewlombard.com/reliability/#How should content analysis researchers properly assess and report intercoder reliability

55. MacGregor D. and Slovic P. Graphical representation of judgmental information. *Journal of Human-Computer Interaction,* 2 (3): 179-200, 1986.

56. Mandel N, Johnson E. *Constructing preferences online: can web pages change what you want?* Unpublished paper. The Wharton School, University of Pennsylvania (Feb, 1999), 1999.

57. McLachlan P., Munzner T., Koutsofios E., and North S. LiveRAC: interactive visual exploration of system management time-series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '08*, pp. 1483–1492, 2008. ACM.

58. McNee S.M., Lam S.K., Konstan J.A., and Riedl, J. Interfaces for eliciting new user preferences in recommender systems. *User Modeling,* 2702: 178–187, 2003.

59. Meyer R. J. and Eagle T. C. Context-induced parameter instability in a disaggregate-stochastic model of store choice. *Journal of Consumer Research,* 19 (1): 62-71, 1982.

60. Miao Q., Li Q., and Dai R. AMAZING: A sentiment mining and retrieval system. *Expert Systems with Applications,* 36 (3): 7192-7198, 2009.

61. Miles, G. E., Howes, A., and Davies, A. A framework for understanding human factors in web-based electronic commerce. *International Journal of Human-Computer Studies*, 52 (1): 131-163, 2000.

62. Neuendorf, K. A. *The content analysis guidebook.* Sage Publications, Inc, Thousand Oaks, Dec. 2001.

63. Noorden V., and Gunter K. *Binocular vision and ocular motility: theory and management of strabismus (4th Ed.)*. Mosby, Sydney, 1990.

64. Oelke D., Hao M., Rohrdantz C., Keim D. A., Dayal U., Haug L.-E., and Janetzko H. Visual opinion analysis of customer feedback data. In *IEEE Symposium on Visual Analytics Science and Technology,* pp. 187-194, 2009.

65. Ozakca M. and Lim Y.-K. A study of reviews and ratings on the internet. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems,* pp. 22-27, Montreal, Quebec, Canada, April 2006.

66. Payne J. W. Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organizational Behavior and Human Performance,* 16 (2): 366-387, 1976.

67. Payne J. W. and Braunstein M. L. Risky choice: An examination of information acquisition behavior. *Memory and Cognition,* 6 (5): 554–561, 1978.

68. Payne J. W. Contingent Decision Behavior. *Psychological Bulletin,* 92 (2): 382-402, 1982.

69. Payne J. W. and Bettman J. R. Measuring Constructed Preferences: Towards a Building Code. *Journal of Risk and Uncertainty,* 19 (1-3): 243-270, 1999.

70. Payne J. W., Bettman J. R., and Johnson E. J. *The Adaptive Decision Maker*. Cambridge University Press, Cambridge, UK, 1993.

71. Peterson H.E. and Dugas D.J. The relative importance of contrast and motion in visual detection. *Human Factors,* 14 (3): 207–216, 1972.

72. Pinker S. *The stuff of thought: Language as a window into human nature.* Viking Press, New York, 2007.

73. P. Pu and P. Kumar. Evaluating example-based search tools. In *Proceedings of the 5th ACM Conference on Electronic Commerce (EC'04*), pp. 208- 217, New York, NY, USA, 2004. ACM.

74. Russo J. E. The value of unit price information. *Journal of Marketing Research,* 14: 193-201, 1977.

75. Russo, J. E. Adaptation of cognitive processes to eye movement systems. In Senders J. W., Fisher D. F., & Monty R. A. (Eds.), *Eye Movements and the Higher Psychological Functions,* pp. 89–109, Lawrence Erlbaum, Hillsdale, November 1978.

76. Russo, J. E. Eye fixations can save the world: A critical evaluation and a comparison between eye fixations and other information processing methodologies. *Advances in consumer research,* 5: 561–570, 1978.

77. Russo J. E. and Dosher B. A. Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9: 676-696, 1983.

78. Russo J. E. Characteristics of successful product information programs. *Journal of Social Issues,* 47 (1): 73-92, 1991.

79. Shneiderman B. *Designing the user interface (3rd Ed.)*. Addison Wesley, 1997.

80. Simon, H. A. *The sciences of the artificial (3nd Ed.).* MIT Press, Cambridge, UK, 1996.

81. Simonson I. and Tversky A. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research,* 29 (3): 281–295, 1992.

82. Sinha R., and Swearingen K. The Role of Transparency in Recommender Systems. In *CHI 2002 Extended Abstracts on Human Factors in Computing Systems*, pp. 830-831, 2002. ACM.

83. Stone D. N. and Schkade D. A. Numerical and linguistic information representation in multiattribute choice. *Organizational Behavior and Human Decision Processes*, 49: 42-59, 1991.

84. Sundaram D.S., Mitra K. and Webster C. Word-of-mouth communications: A motivational analysis. *Journal of Advances in consumer research*, 25(1): 527-531, 1998.

85. Swearingen K. and Sinha R. *Interaction design for recommender systems. In Designing Interactive Systems 2002*, ACM, 2002.

86. Tidwell J. *Designing Interfaces (2nd Ed.)*. O'Reilly Media, January 2011.

87. Todd P. and Benbasat I. The influence of decision aids on choice strategies: An experimental analysis of the role of cognitive effort. *Organizational Behavior and Human Decision Processes,* 60 (1): 36-74, October 1994.

88. Tversky A. Elimination by Aspects: A Theory of Choice. *Psychological Review*, 79 (4): 281-299, Jul. 1972.

89. Tversky A. and Kahneman D. Choices, values, and frames. *American Psychologist,* 39 (4): 341–350, 1984.

90. Ware C. *Information Visualization: perception for design (3rd Ed.)*. Morgan Kaufmann, June 2012.

91. Wickens C. D. *Engineering psychology and human performance (2nd Ed.)*. HarperCollins, New York, 1992.

92. Winterfeldt, D. V., and Edwards, W. *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge, 1986.

93. Yatani K., Novati M., Trusty A., and Truong K.N. Review spotlight: a user interface for summarizing user-generated reviews using adjective-noun word pairs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '11,* pp. 1541–1550, 2011. ACM.

94. ISO usability definition, http://en.wikipedia.org/wiki/Usability#Definition.

95. Customer feedback matters, https://moderncomment.com/customer-feedback-stats

96. The definition of Ecological validity, http://en.wikipedia.org/wiki/Ecological_validity