Abstract

Modern-day search platforms generally have two layers of information presentation. The outer layer displays the collection of search results with attributes selected by platforms, and consumers click on a product to reveal all its attributes in the inner layer. The information revealed in the outer layer affects the search costs and the probability of finding a match. To address the managerial question of optimal information layout, we create an information complexity measure of the outer layer, and study the consumer search process for information at the expense of time and cognitive cost. Using a unique and rich panel tracking consumer search behaviors at a large online travel agency (OTA), we find cognitive cost is a major component of search cost, while loading time cost has a much smaller share. By varying the information revealed in the outer layer, we find price revelation shifts search behavior most dramatically compared to the other product attributes. We propose information layouts that Pareto-improve both revenue and consumer welfare for our OTA.

Keywords: online consumer search, cognitive modeling, information complexity, search intermediaries, platform design.

JEL Classifications: D83, L81, L86

*Scheller College of Business, Georgia Institute of Technology, chris.gu@scheller.gatech.edu. I am grateful to Pradeep Chintagunta, Jean-Pierre Dubé, Günter Hitsch, and Anita Rao. I thank Yike Wang, Elisabeth Honka, Michael Wichura, Coley Dale, Zhiying Hu, Ajay Coley, Adithya Pattabhiramaiah, Omer Cem Ozturk, Sumon Datta, Axel Stock, Bradley Shapiro, Rong Song, Mingjie Zhu, and Hua Zhao. I also thank the feedback from the audience of the Marketing Science Conference 2016, Georgia Research Symposium 2017, and the University of Chicago Marketing working group. Last but not least, I would like to thank the generous support of INFORMS Society for Marketing Science Doctoral Dissertation committee and Mary Kay Foundation Doctoral Dissertation committee. The usual disclaimer applies.
1 Introduction

The Internet has opened an expansive ocean of information and product choices to consumers. Search platforms are valuable compasses in this ocean helping consumers acquire information and make choices effectively and efficiently. Modern-day search platforms can generally be considered as having two layers of information presentation. The outer layer displays the collection of search results, and consumers may click on a search-result entry to learn about all the product details in the inner layer. To assist consumers during their search process, platforms usually summarize the products by selecting a subset of product attributes to display in the outer layer. If too much information is presented in the outer layer, the need to click beyond the outer layer is reduced; yet the search-result pages may become too complex and require too much cognitive effort for consumers to understand. If too little information is presented, the search-result pages may become trivial to understand, but consumers need to incur high clicking cost to find what they need.

We use the historical layout change in Google’s search-result pages as a motivating example. Figure 1 shows the layout of search-result pages for the keyword “minimalistic” in 2000 on the left and in 2016 on the right.[1] The left panel shows the early design, which only revealed the title of the search-result items. The right panel shows the modern design, under which Google reveals additional article information to consumers, including host names, URLs, and the tagline of the articles. Consumers gain more certainty about the link content by reading the additional information, and hence click fewer times.

Similar design decisions need to be made for shopping search platforms, such as Expedia, where the platform needs to decide which hotel attributes to display on the search-result pages, and how the search-result pages have to be revised for smaller screens, for example, mobile phones. Understanding how much information to reveal in the outer layer to help consumers search more effectively is in the interest of search platforms. This issue is a central platform design decision that balances simplicity and usefulness (Wong 2017).

Before we delve into further details, we lay out the key elements of the consumer search process and the terms used. A more thorough introduction will be provided in Section 3. During their search process on modern search platforms, consumers navigate the outer layer by visiting different

---

“refinements” that are collections of search results. Each refinement is defined by some filter criteria that narrow the search-result set, and by some sorting criteria that order the search results. If the consumers specify no sorting criteria, a default platform-suggested ranking is applied to order the search results. Once they observe the product availability and the partially revealed product information on the visited refinements, the consumers decide which products to click on for more information. While the consumers acquire product information in this process, they also incur various search costs, including time and cognitive cost. In particular, consumers need to wait for webpages to load, and to exert cognitive effort to understand and compare the revealed product information.

In this paper, we answer the research question of how consumers acquire information at the expense of time and cognitive costs. Answering this question helps us address the managerial question of the optimal amount of information to reveal in the outer layer. To answer these questions, we collect measures of time cost and innovate measures of cognitive costs. We then find model-free evidence that supports the impact of these costs on the consumer search process. We conceptualize the consumer search process described above by proposing a new structural model. Finally, given the above knowledge, we try to improve the platform design that balances consumer welfare and platform revenue.
We further expound on our research and illustrate its connection to the most related literature in terms of the research question, cost measures, data, and model.

In terms of the research question, the consumer search literature grew out of the economic theory of information search (Stigler 1961), in which consumers are no longer assumed to be all-knowing agents as in the traditional demand theory. Instead, consumers form their consideration sets for final purchase decisions by engaging in information search (Hauser & Wernerfelt 1990, Roberts & Lattin 1991). Search cost estimation is the focus of the empirical search literature because it lies at the heart of search models. Depending on the managerial questions being answered, past literature has looked at the overall search cost per consumer search, such as in (Honka 2014) for customer retention study, and at positional search cost for optimal default platform-suggested ranking, such as in (Ghose et al. 2012, 2014, Ursu 2016).

Both the refinement visiting and product clicking decisions are important for effective search. Scant research has studied refinement choice, because of the high demands on data. The closest papers to ours are (Chen & Yao 2016, De los Santos & Koulayev 2016), which highlight the importance of studying the consumer refinement choice in that they reflect consumer preferences. For example, (De los Santos & Koulayev 2016) assert that “consumers who sort by price in search are more price sensitive in choice.” Our insight is that different refinements have different “information complexity” (to be defined next) and hence require different cognitive effort to process. When consumers switch refinements, they not only change the potential information they will find and hence reflect consumer preference, but also alter the information complexity and hence reveal information on the cognitive search cost. We formalize this concept and study a new managerial question that is difficult to address otherwise. Furthermore, the study of optimal information layout complements the study of optimal ranking on search platforms, because if the default platform-suggested ranking ignores the cognitive cost component, the refinement pages that use this ranking may appear too complicated for consumers to understand. This difficulty will make consumers switch to some other user-specified ranking, or increase the probability of consumers leaving the search platform prematurely, and hence reduces the full potential of the ranking algorithm.

In terms of the cost measures, we propose using webpage loading time for the time cost, and we innovate “orderedness entropy” as an information complexity measure, which, together with consumer responsiveness to it, accounts for cognitive cost. The concept of entropy comes from...
information theory, first proposed by (Shannon 1948). The earliest appearance of an application of entropy in the marketing literature is (Herniter 1976), who builds a probabilistic model of market share based on an underlying consumer taste distribution, and estimates this distribution by maximizing the entropy of the probabilistic model. For a survey of how information theory is used in marketing as a unified modeling approach, see (Brockett et al. 1995). Our research also relates to the stream of recent economics research on rational inattention (Caplin 2016). Empirical literature in that stream has also applied the concept of entropy, but mostly in its original sense to quantify the randomness of and distances between distributions.

The essence of our measure lies in the insight that filtering and sorting not only bring the consumers’ favored products to display, but also reduce the consumers’ cognitive effort. As a result, an effective information complexity measure needs to capture, at its bottom line, the impact of the well-sortedness of a string of values on consumer perception. Once we have an orderedness measure, we use entropy to quantify the well-sortedness. Computing entropy for a string of values directly would miss the target, because permuting that string will generate the same entropy value.

In terms of the data, we have collected a rich data set that allows us to accomplish our task. As we mentioned, the study of refinement choice puts a high bar on the data for the following reasons. Each refinement is composed of many products, such as 25 items typical in the online hotel agency context, and each product has many attributes, some of which are dynamic. For a particular refinement on different days, the search platform may display different inventory and product attributes, such as price, due to their dynamic nature. Therefore, consumers may observe different information on different days even for the same refinement. To carefully model the consumer search process, we need to reconstruct the exact information consumers are likely to have observed. The search platform we investigate managed to store a large amount of refinements over a long period of time. We managed the big data challenge to retrieve the dynamic product attributes and rebuild the decision context during the time window in which consumers arrive. Finally, we organized the data amenable for efficient search model estimation in parallel. The panel data researchers have used in previous consumer search literature mainly contains the information for the products consumers have clicked on, but not the refinements consumers have visited, such as in (Chen & Yao 2016, Kim et al. 2016, Ursu 2016). The data that are closest to ours are used in (Ghose et al. 2014, De los Santos & Koulayev 2016).
In terms of the model, our central principle is to characterize the evolution of consumers’ information and cost structure during their search process, accounting for model tractability. Our model is rooted in the sequential search model proposed by (Weitzman 1979) and follows the same tradition as (Kim et al. 2010, 2016, Chen & Yao 2016, De los Santos & Koulayev 2016, Ursu 2016). The two papers that study refinement choices have the following defining characteristics. (Chen & Yao 2016) model each consumer search action as a combined refinement and click choice. The authors assume consumers are unaware of the attribute values of all unclicked products at any stage of the search process. As a result, after a consumer arrives at a particular refinement and clicks on a product, if he decides to click on another product on the same refinement, his information about and cost of the second product is assumed to be the same as if he had not previously arrived on the refinement, even though the refinement reveals some information about the second product. (De los Santos & Koulayev 2016) model refinement choice using sequential search model, and adopt a discrete choice model of product clicks in a reduced form fashion by assuming consumers click on at most one product on any refinement. The use of a discrete choice model also assumes the attributes and utility shocks of all products are known to consumers after they visit a refinement. We cannot adopt these models to address our research and managerial questions, because our model needs to have the flexibility to account for the following scenario. When more information is revealed on refinement pages, for example, our model should allow the consumers to utilize the additional information to help their product clicking decisions. They may click less because less uncertainty surrounds the products they have not sampled, even though the search cost for clicking has not changed. On the other hand, the additional revealed information should come at the expense of higher cognitive cost for refinements. As a result, our model needs to capture both the refinement visiting and product clicking decisions, to distinguish between what consumers know before and after these two types of search decisions, and to decompose search costs further into time and cognitive components. Our model uses sequential search framework to allow both refinement visiting and product clicking as search options, while adopting the same one-step-ahead simplification as in (De los Santos & Koulayev 2016) for model tractability. The model identification follows the same idea as (Kim et al. 2010, 2016, Chen & Yao 2016, De los Santos & Koulayev 2016, Ursu 2016): purchase decisions identify consumer preferences and cost measure variations identify consumer search costs.
In our settings, we find through model-free analysis that consumers tend to visit more refinements than to click on individual products within a search session (a session is the set of search activities within a start and end timestamp provided by the online travel agency (OTA), similar to the definition in the previous literature), which means consumers actively use search-result pages as an important source of information. The increase in the loading time of refinement and product-detail pages decreases the number of refinement visits and product clicks. Furthermore, we find orderedness entropy is a good information complexity measure for cognitive cost, and consumers are more likely to sort the information on a refinement when sorting induces a higher reduction in cognitive cost. Based on our model estimates, we find cognitive cost is a major component of all search costs. For a typical refinement visited, consumers incur a cost equivalent of $18.68 on average. Time cost is estimated to be $16.53 per hour on average, and its share in the search costs is much smaller compared to cognitive cost. We find the position cost to be $0.06, which is equivalent to a cost of $1.50 if consumers scroll to the end of a refinement page with 25 products on it. This position cost is similar to that of $0.05 estimated by Chen & Yao (2016) for an average consumer in their context.

We then vary the information revealed in the outer layer to characterize the various changes in the consumer search process. We find price is an important product attribute whose presence in the outer layer will dramatically change consumer search behavior. We also study the relationship between consumer welfare and platform revenue. On the Pareto frontier of this relationship, we find a general tradeoff in which consumer welfare increases and revenue decreases as more information is revealed. For our OTA under study, we find room for Pareto improvement in both revenue and consumer welfare by changing which product attributes are revealed. In the consumer welfare maximizing layout, price is revealed (among other changes in attribute revelation), and the platform may increase consumer welfare by 95% without losing any revenue. In the revenue maximizing layout, price is not revealed (among other changes in attribute revelation), and the platform may increase revenue by about 60% without hurting consumer welfare. We also look at win-win configurations in which both consumer welfare and platform revenue may be improved nontrivially.

To summarize our contributions in this paper, we extend the search literature to study cognitive and time costs that are crucial in the consumer online search process. We then make a novel first attempt at answering an important managerial question of how much information should be revealed
in the outer layer, offering implications for all modern search platforms. We innovate an information complexity measure for cognitive cost, and propose a new structural model to model the consumer endogenous choice of refinement visiting and product clicking decisions. We believe our modeling principle is in full accordance with the essence of search model literature, which is to capture what the decision makers do and do not know at each decision point, and to advance our understanding of the associated finer-level costs involved in the search process.

The rest of the paper is organized as follows. Section 2 describes the empirical context of our study. Section 3 illustrates the conceptualization of the consumer search process. Section 4 details our various search cost measures, in particular, cognitive cost. Section 5 summarizes the data. Section 6 presents model-free evidence aimed at exploring whether consumers respond to the cost measures. In section 7, we develop our structural model and discuss our estimation approach in Section 8. Section 9 presents and discusses our estimation results. Section 10 performs a managerial study of consumer welfare versus platform revenue by varying the amount of product information revealed in the outer layer. Section 11 concludes.

2 Empirical Context

Our empirical context is a major online travel agency (OTA) in China, which is the counterpart to Expedia in the United States. The site design can be seen in Figure 2. This search platform provides search functionalities in terms of filters and sorting variables. The filters allow consumers to reduce search-result sets by retaining results that fit the exact criteria specified, in terms of city, checkin and checkout dates, price, hotel star and so on. The sorting variables, including popularity, consumer rating, price, and hotel star, allow consumers to rearrange the order of display, and only one active sorting variable can be applied. Popularity is the default sorting variable; price and hotel star have both ascending and descending sorting capability, whereas the other two variables only have one sorting direction. A search query is defined in terms of both the filters and the sorting variable chosen by a consumer. When a consumer submits a search query, the search platform retrieves the search-result set that satisfies the filters, and generates a refinement page ordered by the sorting variable for the consumer to interact with further. A given search query may return different content at different times, because product availability and some product attributes are
dynamic, such as price and cancellation policy.

Beneath the search functionalities, the site displays the search-result set of up to 25 hotels on a single page, with more results to display in latter pages. These search-result pages are what we call “outer layers,” where the platform selects a subset of product attributes to display, including price, consumer rating, hotel star, breakfast, cancellation policy, and so on. If the consumer clicks on a hotel, he lands at the “inner layer” of the product-detail page, where he can learn about all the remaining attributes of the clicked hotel, including free newspaper, free guide service, free afternoon tea, and so on. All the products on the OTA under study have the same inner layer structure; they only differ in the value of the product attributes.
3 Conceptualization of Consumer Search Process

As we mentioned in the introduction, our central modeling principle is to characterize the evolution of consumers’ information and cost structure during their search process, accounting for model tractability. As a result, our model needs to capture both the refinement visiting and product clicking decisions, to distinguish between what consumers know before and after visiting refinements and clicking products, and to decompose search costs further, particularly into time and cognitive components.

We first recognize that refinement pages in the outer layer and product-detail pages in the inner layer are different information sources. Consumers learn about product availability and the partially revealed attributes of all the products on a given refinement, whereas they learn about the hidden product attributes of a single product they click on. The search costs associated with these two information sources are also different. Instead of having separate mean search costs for these two sources, we further recognize that information collection takes time and cognitive effort, and variation of the factors that influence these costs may influence the search process. For example, slow platform loading time or complex website design may induce consumers to search less and make do with less-than-ideal products they have found. In our conceptualization, we will parse out the time and cognitive components of the search costs.

Furthermore, in the platform design of the OTA we are studying, the inner layer of product-detail pages have the same page layout, except that the attribute values are different for different products. By contrast, the outer layer of refinement pages will, by consumer choice, have different compositions of products in the variation of product attributes and well-sortedness. For example, a search-result page that has a filter of 5-star hotels will have no variation in the hotel star attributes, enabling consumers to easily understand the star attribute. If the consumer further sorts the result set on price, the search-result page will be well sorted on price attributes, enabling consumers to easily compare products in the price dimension. This example calls for our information complexity measure, which is meant to capture the cost of comprehending and comparing different product attributes. When many products are available to choose from, comparing a randomly juxtaposed list of products is more costly than comparing a well-sorted list. Note that our information complexity measure will not capture the cognitive effort required to comprehend a single product, such as when
consumers are reading product-detail pages, and we leave that component in the residual search cost.

Finally, we assume consumers rely on the expected search costs to guide their search decisions, and they have rational expectations about time cost as well as the distributions of product attributes on refinements from which the expected cognitive cost is derived. In terms of loading time, because consumers can fairly easily get a sense of the platform response time after the first one or few pages they visit, we assume each consumer forms his or her own session-specific expectations about the loading time of refinement and product-detail pages. Because the computation required to construct different refinements on the server are the same, the expected loading time for different refinements within a session will be assumed to be the same. The same holds true for product-detail pages, and the expected loading time for different product-detail pages will be assumed to be the same. Furthermore, the consumers expect to incur the same loading time cost when they revisit a refinement due to the platform design that prohibits caching of refinement content.

In terms of the belief of the product attribute distributions on refinements, because each refinement typically contains a large amount of product attributes, and different refinements have different product attribute distributions, we assume it is hard for consumer to learn or update their beliefs of these distributions. Instead, their beliefs will reflect the long-term product attribute distributions for each refinement respectively. This assumption implies that expected cognitive cost and product attribute distribution of a particular refinement will be the same over search sessions for all consumers (the expected cognitive cost and product attribute distribution can be different for different refinements). We will lay out the rigorous definitions in Section 7.

In summary, we have the following types of search costs. To decide whether to visit a particular refinement, consumers will account for the session-specific expected loading time cost, the overall expected cognitive cost to understand the given refinement, and the residual search cost for refinement visits. To decide whether to click on a particular product, consumers will account for the session-specific expected loading time cost of the product-detail page, the product position cost of going down the search-result page to click on the link (as studied in the previous literature), and the residual search cost for product clicking.

Instead of describing the conceptualization of the consumer search process by writing down mathematical models at the first encounter, we illustrate the conceptualization by walking through a
hypothetical search scenario. During this hypothetical search, we lay out the information consumers acquire and the costs they incur at each step, using the diagrams in Figure 3. We will formalize the conceptualization as a model in Section 7. We also introduce some terminology we will use throughout the paper.

3.1 Hypothetical Search Scenario

Figure 3 shows 10 steps, each of which is depicted in a row. Suppose three refinements, Ref1, Ref2 and Ref3, are available denoted by the rectangles at the top of each row of a step. From the consumers’ perspective, a refinement is defined by the associated search queries, which include search filters and sorting variables (together with pagination for empirical work as we explain below). We use “search a query,” “visit a refinement,” and “visit a search-result page” interchangeably. Each refinement displays a certain number of search results for users to further click on, which, in our context, is 25 hotels. These hotels are denoted by the circles at the bottom and are linked to the refinement they belong to, using an edge. We use “click on a link” and “sample a product” interchangeably.

First the consumer arrives at the default refinement Ref1 in step (1). He learns the available products in Ref1 and observes the partially revealed product information $x_{i1}^I$, but does not know the unrevealed information $x_{i1}^N$ for product $i = 1, \ldots, 25$. The search costs involved in step (1) include the time cost of waiting for the search-result page to load, the cognitive cost of understanding the page, and any related residual cost. In step (2), the consumer can choose to search either one of the 25 partially revealed hotels on Ref1, or the unvisited refinements Ref2 and Ref3. These search options are denoted in yellow in Figure 3. Because he has not visited Ref2 and Ref3, he does not know the product availability of these refinements and cannot sample any product in those refinements directly.

In step (3), suppose the consumer decides to sample hotel 1, which reveals the hidden information $x_{11}^N$. The costs for step (3) include the time cost of waiting for the product-detail page to load, the position cost, and any related residual cost. The consumer ponders his search options again in step (4), which include the unsampled products on Ref1 and the unvisited refinements Ref2 and Ref3, all denoted in yellow in Figure 3. Suppose he decides to switch to Ref3 in step (5). At that point, he learns the product availability of Ref3 and observes the partially revealed product information.
He incurs the same cost of visiting a refinement in step (5) as in step (1). Now all the information contained in Ref3 is known to the consumer, and Ref3 is no longer a search option; instead, all its related products become search options, because they still have uncertainty to be resolved. After consideration in step (6), he samples hotel 25 of Ref3 in step (7) to learn the unrevealed product attribute, and incurs the usual clicking associated costs as in step (3). When he considers his search options in step (8), even though he is at Ref3, he still has unsampled products in Ref1 as his search options. He decides to switch back to Ref1 and samples hotel 2 in step (9). He learns the unrevealed product attribute and incurs the usual clicking cost as in step (3). Moreover, he incurs the time cost of loading Ref1, but he does not need to incur the other associated costs of visiting a refinement, including the cognitive and residual cost, because he already understands the refinement. Finally, he decides not to search anymore and to purchase hotel 25 of Ref3 among all the sampled products.

In addition to the conceptualization of the consumer search process above, we pay further attention to some details in our empirical work. The hotels displayed under different refinement are generally different, but some hotels could appear in multiple refinements if they satisfy the search query. For example, suppose hotel 1 appears in both Ref1 and Ref3 because it satisfies the criteria imposed by both search queries. Figure 4 shows such a scenario. Before visiting Ref3, the consumer does not know the product availability. In step (5), he visits Ref3 and learns its product availability. Then the product information for hotel 1 is carried over from Ref1, and he does not need to sample hotel 1 again.

Moreover, in our empirical work, we define refinements by search queries that include not only search filters and sorting variables, but also pagination. The rationale for this level of granularity is as follows. Typical search queries return many more results than the maximum number of 25 hotels that can be displayed on a search-result page. As a result, all the search results will be split into different pages, where the product attribute distributions are different on different pages due to sorting. For example, if a consumer is price sensitive, he may sort the search results by price in ascending order. The product attribute distribution on different pagination will be different, and this difference has behavioral implications whereby the consumer is more likely to look at earlier than later pages. By defining refinements with consideration of pagination as well, we can control for different product attribute distributions as pagination increases. Furthermore, our empirical model can account for refinements that have fewer results than the maximum number of 25 hotels.
that can be displayed on a search-result page. We just use 25 hotels in our illustration for clarity.

This search process is the basis of our search model and we will detail the model specification in Section 7.

4 Cost Measures

In Section 3 we described the different search costs consumers incur during the search process. Four basic types of search costs exist: time, cognitive, position, and residual cost. We can measure the time cost of waiting for the search-result and product-detail pages to load, using their respective loading times. We use the observed position of the products to measure their position costs. The residual costs capture all the remaining costs unobservable to the researchers. Because no quantification of the cognitive costs exists, we create an information complexity measure of a search-result page to help gauge the cognitive cost.

4.1 Information Complexity Measure

The information complexity measure is based on the entropy of orderedness of the product attributes on the refinement. Recall that information complexity measure, which, together with consumer responsiveness to it, accounts for cognitive cost, is meant to capture the cost of comprehending and comparing different product attributes.

Information entropy, first introduced by (Shannon 1948), measures the amount of uncertainty of a random variable and has fundamental importance in communication theory. The more structured and less random a variable is, the less entropy it has. An intuitive way to think about entropy is how many binary inquiries one needs to ask the person who knows the realization from a random variable to learn its value. Hence, entropy can be considered as the average minimum description length of a random variable, where the description length is the number of binary inquiries made. For example, a constant random variable has 0 entropy, because one knows everything about the value without making any inquiry. On the other hand, one needs to make some inquiries to learn about a non-degenerate distribution. The more uniform the distribution, the more inquiries one needs to infer its value. The entropy for a discrete random variable \( Y \) with probability mass function \( P(y) \) is defined as \( H(Y) = E[-\log_2(P(Y))] \) with units in bits. For continuous random variables,
Figure 3: Illustration of the consumer search process in the context of three available refinements, each with 25 products to sample. The consumer arrives at the default refinement Ref1 in step (1), and then decides the search option to proceed with among the 25 partially revealed products on Ref1 and the unvisited refinements Ref2 and Ref3 in step (2). Subsequently, he samples product 1 in step (3), considers again in step (4), switches to Ref3 in step (5), considers in step (6), samples product 25 of Ref3 in step (7), and considers again in step (8). At this point, even though he is at Ref3, he still has unsampled products in Ref1 as search options. He decides to switch back to Ref1 and samples product 2 in step (9). Finally, he decides to stop sampling and to purchase product 25 of Ref3.
Figure 4: Alternative scenario in which hotel 1 of Ref1 also appears as hotel 1 in Ref3. It could also appear as a hotel at some other position in Ref3 depending on the ranking. All the steps are the same as in Figure 3 except after step (5). Before visiting Ref3, the consumer does not know the product availability of Ref3. In step (5), the consumer visits Ref3 and learns its product availability. The product information for hotel 1 is carried over from Ref1, and hotel 1 in Ref3 is not a search option for the consumer for the rest of the search process, which differs from the previous scenario.
we first discretize them and then apply the same computation.

To create the information complexity measure of a search-result page, we cannot apply the concept of entropy directly to product attributes, because entropy ignores the order of the realizations of random variables. Instead, we argue that orderedness is precisely the quantity that matters for the comprehension and comparison of product attributes: the more ordered the value sequence of a given product attribute, the less cognitive costs necessary to understand the product attribute. We first define a new sequence, called an orderedness sequence, based on a given single product attribute. We then define information complexity for a single product attribute by combining the concept of entropy together with the orderedness sequence. Finally, we define the information complexity for a given refinement page using the information complexities of its constituent product attributes.

The orderedness sequence of the numeric sequence \( x \) is defined to be

\[
ord(x) = (x_{i+1} - x_i)_{i=1}^{m-1},
\]

where \( m \) is the length of \( x \). This definition can first capture whether a numeric sequence is ordered or not, that is, the qualitative aspect of orderedness. Suppose we want to understand the following numeric sequences:

\[4, 2, 5, 1, 3, 2, 3, 4, 1, 5\text{ vs }1, 1, 2, 2, 3, 3, 4, 4, 5, 5.\]

The second sequence is sorted but otherwise equivalent. Computing some summary statistics, such as the mean, and identifying and comparing two values in terms of their frequency in the second sequence is easier than doing so in the first. The orderedness sequences can be computed as

\[-2, 3, -4, 2, -1, 1, 1, -3, 4\text{ vs }0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0.\]

The first orderedness sequence has more variation than the second, which corresponds to the lack of orderedness of the first sequence.

The orderedness sequence can also capture how regular the successive changes are in the original sequence values, that is, the quantitative aspect of orderedness. For example, consider the following
numeric sequences:

\[1, 4, 5, 7, 11, 23, 27, 31, 32, 36 \text{ vs } 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.\]

Comprehending the second sequence is easier than comprehending the first, even though both are ordered. The rationale for this phenomenon is that a sequence of ordered and evenly spaced values is easier to understand than another ordered yet jumpy sequence. The derived orderedness sequences for this example is

\[3, 1, 2, 4, 12, 4, 4, 1, 4 \text{ vs } 1, 1, 1, 1, 1, 1, 1, 1, 1.\]

We again observe that the first orderedness sequence has more variation than the second.

Given these observations and the definition of the orderedness sequence, we complete the definition for the information complexity measure of a numeric sequence \(x\), or a product attribute, as the entropy of the derived orderedness sequence, namely, \(H(\text{ord}(x))\). The higher the orderedness entropy, the more variation exists in the orderedness sequence, and hence the harder it is to understand the original sequence. We may extend our definition to ordinal sequences by applying our definition to the associated ranking sequences. We may also extend our definition to binary sequences by setting one level to 1, and the other to 0, and applying our definition to the resulting numeric sequence. Our definition is invariant to which level is set to 1, because entropy is invariant to translation and scaling. Finally, we define the information complexity measure of a search-result page by summing over the information complexity of its constituent revealed product attributes.

In terms of our empirical context, when consumers rearrange search results by applying the sorting variable to, for example, the hotel star, they drastically reduce the complexity of the search-result page in that dimension. Similarly, applying filter variables will reduce the complexity in those dimensions. For example, filtering hotels such that only 4-star hotels remain in the search result will reduce the information complexity of that dimension to 0. Furthermore, due to the correlation among product attributes, sorting or filtering a subset of the product attributes will change the information complexity of the overall search-result page. We will show empirically in Section 6 that applying sorting and filtering variables, in general, reduces information complexity. More importantly, we show that people respond to information complexity by actively applying
Table 1: Descriptive Statistics. Search-result size refers to the number of matching hotels for a given search query.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search-Result Size</td>
<td>1</td>
<td>66</td>
<td>171</td>
<td>908</td>
<td>1550</td>
<td>4896</td>
<td>1271.5</td>
</tr>
<tr>
<td>Number of Purchases per Consumer</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.25</td>
<td>1</td>
<td>42</td>
<td>0.922</td>
</tr>
</tbody>
</table>

5 Data

In addition to the OTA description we provide in Section 2, we observe a detailed panel of consumer interactions with the site, including the search queries used to refine the search results, the content and loading time of the search-result and product-detail pages, the refinements visited, the clicks made, and the rooms purchased. The data set we use covers 29,065 randomly selected consumers who visited the site searching for hotels in six cities in China (Beijing, Shanghai, Guangzhou, Hangzhou, Chengdu, and Lijiang) from Jan 27, 2014, to May 29, 2014.

5.1 Data Description

Some key summary statistics of the data are presented in Table 1. The median number of hotels in the typical search-result set is 171, which need to be split into several pages to be displayed. We have removed search queries that provide no results. The number of purchases made by consumers in the observed window is slightly more than 1 on average, and the maximum number made by a single consumer is 42 times. A consumer may have several search sessions, any of which may result in no purchase. Our data do not have consumers who have not made any purchase in any search session.

In Table 2, we show the number of search-result pages visited and the number of product-detail pages clicked for each consumer session in different cities. We can see the distribution for the number of search-result pages visited and the number of product-detail pages clicked are quite similar across cities.

A noticeable and consistent phenomenon across different cities is that consumers tend to visit more search-result pages than to click on individual products. The implication is that consumers
Table 2: Distribution of the number of search-result pages visited for each consumer session in different cities.

actively use search-result pages as an important source of information, and they only click on a product if they are sufficiently interested in it to learn its hidden product attributes. This finding motivates us to treat refinement choice as an integral endogenous choice in the consumer search process.

5.1.1 Loading Time

Table 3 shows the loading time distribution of product-detail pages and search-result pages in seconds. We observe that the loading time of search-result pages is longer and has more variation than that of product-detail pages, because more substantial data extraction, computation, and rendering are performed at the platform side. For the same reason, the two sources of loading time are not highly correlated, with a correlation coefficient of 0.163. The variation in the loading time is also large, with some outlying sessions with very long loading times. We have removed sessions with loading times longer than 100 seconds, which usually corresponds to failed connections due to either consumer or platform side errors.
Table 3: Distribution of the loading time of product-detail pages and search-result pages.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product-Detail Page</td>
<td>0.056</td>
<td>0.931</td>
<td>1.364</td>
<td>2.273</td>
<td>2.182</td>
<td>99.500</td>
<td>4.004</td>
</tr>
<tr>
<td>Search-Result Page</td>
<td>0.032</td>
<td>1.547</td>
<td>2.249</td>
<td>3.369</td>
<td>3.504</td>
<td>99.070</td>
<td>4.571</td>
</tr>
</tbody>
</table>

Table 4: Distribution of orderedness entropies of price, consumer rating, and discrete attributes.

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>0.000</td>
<td>3.255</td>
<td>3.509</td>
<td>3.384</td>
<td>3.689</td>
<td>4.189</td>
<td>0.512</td>
</tr>
<tr>
<td>Consumer Rating</td>
<td>0.000</td>
<td>1.041</td>
<td>1.142</td>
<td>1.128</td>
<td>1.241</td>
<td>2.925</td>
<td>0.248</td>
</tr>
<tr>
<td>Discrete Attributes</td>
<td>0.000</td>
<td>4.338</td>
<td>4.913</td>
<td>4.857</td>
<td>5.462</td>
<td>11.330</td>
<td>1.020</td>
</tr>
</tbody>
</table>

5.1.2 Orderedness Entropy

In our empirical context, we have both numeric and binary product attributes. We convert binary product attributes to numeric values by setting one level to 1 and the other to 0, as suggested by the definition of orderedness entropy. All the product attributes are discrete, except that price and consumer rating of the hotels are continuous product attributes. We discretize price by rounding up in bins of 5 RMB and consumer rating in bins of 0.1 points. The chosen bin sizes are the smallest amount that people discern a difference of these variables in daily life.

For all the discrete attributes, we sum up their orderedness entropies, because they have the common natural unit, namely, bits. To be extra cautious in the discretization step, we keep the orderedness entropies of the continuous attributes separate, allowing their coefficients of the cognitive search cost to be different. The unit of orderedness entropy does not have an immediate economic equivalent unit. However, our structural model will transform bits into equivalent units of time or monetary value on par with other variables that enter consumer decisions.

The distributions of orderedness entropies of price, consumer rating, and discrete attributes are shown in Table 4. The minimum orderedness entropy being 0 comes from very small result sets that have no variation in the corresponding attributes. The interquartile range of the distributions is more reflective of the typical range of orderedness entropies.

To better understand orderedness entropy, we compare the consequences of adding search filters as well as applying sorting variables to any given search-result page. We show, based on the following analysis, that information complexity decreases on average as search filters are added or
sorting variables other than the default popularity ranking are applied. Furthermore, we show, in the next section, that consumers proactively reduce information complexity during their search process. These findings suggest the cognitive cost resulting from information complexity is an important component to study the refinement choice.

First, within each consumer search session, we identify all the pairs of search-result pages with the addition of search filters, while keeping sorting variables fixed. We then compare the difference of the orderedness entropy before and after the addition of any one filtering variable. We find the orderedness entropies of price, consumer rating, and discrete attributes combined decrease, on average, by 0.289, 0.029, and 0.363 bits, respectively, with p-value < 0.001. These values correspond to 8.5%, 2.6%, and 5.3% decrease, respectively, when compared to the mean levels.

Second, within each consumer search session, we identify all pairs of search-result pages that switch from sorting based on the default popularity to either sorting on price, consumer rating, or hotel star, while keeping the search filters fixed. We compare the differences in the orderedness entropy before and after the switching of sorting variables in Table 5. First, all the changes are significant with p-value < 0.001. Second, the largest percentage decreases in orderedness entropies are on the diagonal, which correspond to the variables being sorted. This finding shows that sorting is effective in simplifying the search-result page in the dimension being sorted. Third, due to the correlation structure in the product attributes, when some variables become sorted, some other variables may become unsorted. For example, when sorting search-result pages based on price, the consumer rating variable becomes more complex than sorting based on default ranking. Fourth, the overall orderedness entropies decrease when we apply any of the sorting variables on price, consumer rating, or hotel star. This finding shows that sorting is also effective in simplifying the overall search-result page.

6 Model-Free Evidence

In this section, we first empirically show that people respond to longer loading times by reducing their search and clicking. This finding establishes the basic empirical observation that clicking and searching is beneficial to the consumer as a way of information acquisition, and loading time is one of the relevant costs of search. Second, we confirm that orderedness entropy is a good measure of
<table>
<thead>
<tr>
<th></th>
<th>(bits)</th>
<th>Sort on Price</th>
<th>Sort on Consumer Rating</th>
<th>Sort on Hotel Star</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>-0.461</td>
<td>0.033 (1.0%)</td>
<td>-0.247 (-7.3%)</td>
<td></td>
</tr>
<tr>
<td>Consumer Rating</td>
<td>0.147</td>
<td>-0.331 (-29.3%)</td>
<td>0.143 (12.7%)</td>
<td></td>
</tr>
<tr>
<td>Discrete Attributes</td>
<td>-0.713</td>
<td>0.032 (0.5%)</td>
<td>-0.697 (-10.2%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Within each consumer search session, we identify all pairs of search-result pages that switch from sorting based on the default popularity to either sorting on price, consumer rating, or hotel star, while keeping the search filters fixed. This table shows the differences in the orderedness entropy before and after the switching of the sorting variables, all significant with probability < 0.001. The percentage change is compared to the mean levels of orderedness entropy of each attribute.

information complexity and a good proxy for the cognitive cost of search in our context, by showing consumers are more likely to sort when sorting induces a higher reduction in orderedness entropy.

6.1 The Effects of Loading Time on Number of Searches and Clicks

We run a Poisson regression of the number of result pages searched and the number of product-detail pages clicked on the loading time of the result and detail pages, respectively, controlling for individual, city, and time (day, morning, and afternoon) fixed effects. Table 6 shows significant consumer responses to search and click loading times. For search-result pages, a 1-second increase in the search-result page loading time leads to a 9.4% decrease in the number of searches conducted. At the mean number of 2.77 searches, that percentage amounts to a reduction of 0.26 searches. For product-detail pages, a 1-second increase in the product-detail page loading time leads to 4.7% decrease in the number of product-detail pages clicked, which amounts to a reduction of 0.05 clicks at the mean number of 1.02 clicks.

6.2 The Effects of Orderedness Entropy on Consumer Choice of Sorting Variable

In the previous section, we showed that adding search filters and changing the sorting variable changes the orderedness entropy. In this section, we further show that consumers respond to the cognitive cost of search.

To nail the impact of information complexity on consumer search behavior, we look at decision occasions in which consumers change the information complexity but not the content of the search-result pages. We can only look at whether consumers respond to information complexity by looking
Table 6: Poisson regression of the number of search-result pages searched and the number of product-detail pages clicked on the loading time of the search-result and product-detail pages, respectively, controlling for individual, city, and time (day, morning, and afternoon) fixed effects.

<table>
<thead>
<tr>
<th></th>
<th>Number of Searches</th>
<th>Number of Clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Search Loading Time</td>
<td>−0.094***</td>
<td>−0.047***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Click Loading Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Dest. City Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Day Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Morning Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Afternoon Fixed Effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>35,545</td>
<td>35,545</td>
</tr>
</tbody>
</table>

Standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

at their choices of sorting variables but not filtering variables, because filtering by definition changes the content of the search-result page. We also need to be careful with the choice of sorting variables, because sorting a search-result set larger than 25 will also change the content on a particular page. Therefore, we look at searches in which the result set can be fit on a single page. In this case, changing the sorting variable will not change the content on the refinement, and the only motivation for changing the sorting variable is the orderedness of the results.

In particular, we identify 19,597 consumer searches that use the default popularity ranking and result in fewer than 25 products. Consumers can choose to switch sorting variables to price, consumer rating, or hotel star, or they can choose not to switch. We observe switches to sorting on price 139 times (0.4%), sorting on consumer rating 226 times (0.7%), and sorting on hotel star 39 times (0.1%).

Given the above empirical setup, we study how much more likely consumers are to switch the sorting variable away from the default popularity ranking given how much the switching reduces information complexity. In particular, we use a multinomial logit regression to study the choice probability of sorting on $j \in SV = \{\text{Default Popularity, Price, Consumer Rating, Hotel Star}\}$ as a response to the change in the orderedness entropies, computed by $(\text{Entropy}_{\text{Default},v} - \text{Entropy}_{j,v})$
Table 7: Multinomial logit regression of the choice probability of sorting on default popularity, price, consumer rating, or hotel star as a response to the change in the orderedness entropies. The positive coefficients in front of the reduction of orderedness entropies show that the larger the reduction in the dimension of the sorting variable being applied, the more likely that sorting variable is chosen.

\[ P(\text{sort on } j) = \frac{\exp(\text{Intercept}_j + \sum_{v \in KV} \beta_v (\text{Entropy}_{\text{Default},v} - \text{Entropy}_{j,v}))}{\sum_{k \in SV} \exp(\text{Intercept}_k + \sum_{v \in KV} \beta_v (\text{Entropy}_{\text{Default},v} - \text{Entropy}_{k,v}))}. \]

Table 7 shows the coefficients \( \beta_v \) in front of the reduction of the orderedness entropy \( (\text{Entropy}_{\text{Default},v} - \text{Entropy}_{j,v}) \) are all positive for all \( v \in KV \). Thus when the entropy reduction along one dimension of the sorting option \( j \) increases, ceteris paribus, the consumer is more likely to choose that sorting option. The choice-specific intercepts are all negative, absorbing the loading time cost of switching away from the default popularity ranking.

7 Model

In this section, we formalize the conceptualization of the consumer search process specified in Section 3. We will first describe the components of the consumer product utility and search cost. Then we will lay out the consumer information and cost structure during the search process. Finally, we will specify the decision model for the search process.
7.1 Components of the Consumer Utility and Search Cost

The product attribute is composed of a revealed part displayed on the refinement $x^I \in \mathbb{R}^n$ and an unrevealed part $x^N \in \mathbb{R}^m$ displayed after clicking on the product-detail page. Because the attributes for a given product may change over a session due to their dynamic nature, we use $se$ to denote search session.

Consumer $h$’s utility for product $j$ at search session $se$ is given by

$$u_{hj,se} = V_{hj,se}^I + V_{hj,se}^N + \epsilon_{hj,se} = \beta_h^I x_{j,se}^I + \beta_h^N x_{j,se}^N + \epsilon_{hj,se},$$

where $\beta_h^I$ and $\beta_h^N$ are consumer-specific preference coefficients, and $\epsilon_{hj,se}$ is the consumer’s session-specific utility shock. The session-specific utility of the outside option $j = 0$ is assumed to be known by the consumer from the beginning. $\epsilon_{hj,se}$ is unknown to the econometrician.

Following our assumptions on consumers having rational expectations in Section 3, consumer $h$ has rational expectations about the current search session’s refinement loading time $SearchLoad_{h,se}$ and the product-detail page loading time $DetailLoad_{h,se}$, both computed as the sample average of the loading time of the visited refinement pages and product-detail pages during his search session $se$. For refinement $rf$, he also has rational expectations about the orderedness entropy of price $EntrPrice_{rf}$, of consumer rating $EntrRating_{rf}$, and of discrete attributes $EntrOther_{rf}$, which are the same over search sessions and consumers. Each of these terms is computed as the sample average of the respective orderedness entropy of all the realized page contents of refinement $rf$ over the observation window. The consumer also observes the product position $R_{rf,j,se}$ on the page. We adopt a convex cost structure for the orderedness entropy terms.

7.2 Consumer Information and Cost Structure

Our model will capture the consumer information and cost structure as follows:

- Before visiting a refinement $rf$, the consumer has a general belief about product attribute distribution of the refinement, but does not know the product availability on the refinement. As a result, the unvisited refinement is a search option for the consumer, but products
on this page are not. More specifically, the consumer knows the distribution of the utility \( u_{h,j,se} \) of a randomly drawn product on the refinement; however, he has not observed any information within the refinement, such as product availability or partially revealed product information, to form a more precise conditional belief. At this point, we need to make a modeling decision regarding the utility distribution of visiting a refinement. We follow the same distribution assumption adopted by (De los Santos & Koulayev 2016), and assume the utility distribution of visiting a refinement is the distribution of its first order statistic, that is, \( U_{h,rf} \sim \max_{j \in rf} \{ u_{h,j,se} \} \) (we drop the session \( se \) for \( U_{h,rf} \) because this distribution is common over all sessions). We will discuss our decision model simplification in more detail below.

In terms of the cost to search the unvisited refinement page, the consumer pays the expected refinement loading time cost, the expected cognitive cost, and the refinement residual cost:

\[
    c_{h,rf,se} = \gamma_{h,Search} + \gamma_{h1}SearchLoad_{h,se} + \gamma_{h2,1}EntrPrice_{rf} + \gamma_{h2,2}EntrRating_{rf} + \gamma_{h2,3}EntrOther_{rf} + \gamma_{h2,4}EntrPrice_{rf}^2 + \gamma_{h2,5}EntrRating_{rf}^2 + \gamma_{h2,6}EntrOther_{rf}^2.
\]

- After visiting a given refinement, the consumer knows the product availability on the refinement and the partially revealed products attributes on the refinement. As a result, the visited refinement page is no longer a search option for the consumer, but the unsampled products on this page are. Furthermore, the unsampled products on all previously visited refinement pages are also search options for the consumer. More specifically, the consumer observes, for product \( j \) in refinement \( rf \) at search session \( se \), the partially revealed product attributes \( \{ x^I_{j,se} \} \), the conditional distributions of the unrevealed product attributes \( \{ x^N_{j,se} | x^I_{j,se} \} \), and the distribution of idiosyncratic utility shock \( \{ \epsilon_{hj,se} \} \). In other words, the consumer knows \( \{ \beta^I_{h,j,se} \} \) and the distribution of \( \{ \beta^N_{h,x^I_{j,se},x^I_{j,se}} + \epsilon_{hj,se} \} \), that is, \( [u_{h,j,se}|x^I_{j,se}] \sim \beta^I_{h,j,se}x^I_{j,se} + \beta^N_{h,x^I_{j,se}}x^I_{j,se} + \epsilon_{hj,se} \). Here, the conditional product attribute distribution \( [x^N_{j,se}|x^I_{j,se}] \) also follows our assumptions in Section 3 that all consumers have the
same rational expectations about product attribute distribution over time. As a result, this
conditional distribution reflects a homogeneous prediction rule over search sessions and con-
sumers. In particular, we assume it is a normal distribution with its mean being the linearly
predicted product attributes and its variance being the residual variance, estimated over all
refinements and products. \( \epsilon_{hj,se} \) is normalized to follow \( N(0,1) \) for identification purpose (see
Section 8.2). Note the preference coefficient \( \beta^N_h \) affects the option value of clicking on an un-
revealed product through the variance of the conditional utility distribution, and will explain
the different clicking behaviors.

In terms of the cost to click the unsampled products,

- To click on an unsampled product on the current refinement page where consumers are,
the consumer pays the expected product-detail page loading time cost, the observed
position cost, and the product-detail page residual cost:

\[
c_{h,rf,j,se} = \gamma_{h,Detail} + \gamma_{h1} DetailLoad_{h,se} + \gamma_{h3} R_{rf,j,se}.
\]  

(5)

- To click on an unsampled product on a previously visited refinement page \( rf' \), the con-
sumer pays the additional expected refinement page loading time cost to switch refine-
ment:

\[
c_{h,rf',j,se} = \gamma_{h,Detail} + \gamma_{h1} (SearchLoad_{h,se} + DetailLoad_{h,se}) + \gamma_{h3} R_{rf',j,se}.
\]  

(6)

• For product \( j \) on a refinement page the consumer has visited and clicked, the consumer knows
all the product attributes \( x^I_{j,se} \) and \( x^N_{j,se} \) and their idiosyncratic utility shocks \( \epsilon_{hj,se} \); that
is, the consumer knows \( u_{hj,se} \). The sampled products are no longer search options, but are
candidates for purchase decision. The costs associated with the purchase decisions are the
observed prices of the sampled products.

7.3 Multiple Search Decision Model

We model the consumer search process using multiple search problems. Each search problem is
anchored to the current refinement the consumer is visiting. The consumer starts his search from
the default refinement on the search platform.
Within a given refinement or search problem, the consumer is assumed to engage in the traditional sequential search in which unsampled products and unvisited refinements are both search options. Furthermore, the model shares the same cost-benefit tradeoff as in [Weitzman 1979], where the consumer will evaluate each search option based on the expected gain of search and search cost. This evaluation can be equivalently characterized by the reservation utility of each search option. In particular, for each unsampled product $j$, the consumer computes its reservation utility $z_{h,rf,j,se}$ as the utility level that makes him indifferent between sampling product $j$ versus choosing the current best offer precisely at the utility level $z_{h,rf,j,se}$:

$$c_{h,rf,j,se} = \int_{z_{h,rf,j,se}}^{\infty} (u_{hj,se} - z_{h,rf,j,se}) dF\left(u_{hj,se}\mid x_{j,se}^f\right).$$  (7)

For each unvisited refinement $rf$, the consumer computes its reservation utility $z_{h,rf,se}$ similarly as

$$c_{h,rf,se} = \int_{z_{h,rf,se}}^{\infty} (U_{h,rf} - z_{h,rf,se}) dF\left(U_{h,rf}\right).$$  (8)

During the search process, the consumer will update his current best offer to be the maximum of the realized utility of the sampled products as

$$u_{h,se}^* = \max_{j \in S}\{u_{hj,se}\},$$

where $S$ is the set of sampled products for the consumer. The optimal strategy contains two steps, a selection rule and a stopping rule:

- **Selection rule:** If sampling a product or visiting a refinement, the consumer picks the search option with the highest reservation utility.

- **Stopping rule:** The consumer stops the search process if the current best offer exceeds the reservation utilities of all the unsampled products or unvisited refinements.

When the consumer decides to visit a different refinement from the one he is currently at, we assume the consumer enters a new search problem whereby he needs to evaluate the search options again, because he may learn a new set of products as search options with partially revealed product information, and the search costs of sampling products on previously visited refinements may change.
Hence, in our model, switching refinements means switching search problems. When the consumer switches refinements, he faces the new search problem in which he

- Carries over the current best offer;
- Updates his search costs for the unsampled products on the current and last-visited refinements, as described in Section 7.2;
- Carries over all the search options he has discovered so far, if sampling a product from a previously visited refinement;
- Merges all the search options he has discovered so far with a new set of partially revealed products, if searching an unvisited refinement.

Finally, when the consumer stops the search process, he chooses either a sampled product or the outside option, depending on which one has the highest realized utility.

7.3.1 Discussion

The multiple search decision model is a simplification of a unified sequential search model that contains all the search problems within. The difference is that, in the unified model, no simple search strategy for this general dynamic programming problem exists, and the consumer needs to mentally compute the payoff for all possible search paths. We can have a glimpse into this computation challenge by looking at some necessary computation steps needed. To evaluate the period utility of visiting a particular refinement along any search path, the consumer needs to account for the high-dimensional joint distribution of all the attributes for all products on that refinement. Estimate and integrating over the distribution is difficult. The consumer needs to repeat this computation for all refinements, along all paths. The curse of dimensionality in the number of search paths and the size of the probability space of the product attributes on refinement pages renders this modeling approach infeasible.

The simplification we adopt for our model is the same one-step-lookahead approach as in (De los Santos & Koulayev 2016), where the consumer does not mentally hypothesize what happens beyond reaching the new refinement, because that computation depends on too much information that is unobserved, and accounting for such uncertainty is too costly to perform. Instead, the consumer
uses a heuristic yet intuitive utility distribution to characterize the benefit of searching a refinement. When he arrives at a new refinement, the consumer observes new information and updates his search costs, based on which he re-evaluates the search options in that refinement. As a result, we split the unified sequential search problem into multiple related search problems, each anchored to the current refinement where the consumer is. These multiple search problems are connected by the information the consumer has acquired up to the point of entering a given search problem. When the consumer switches refinements, he switches search problems.

Within each problem, we retain the sequential search structure, because the curse of dimensionality of the form mentioned above does not exist: when the consumer samples products on the current refinement where he is, he has already observed product availability and the partially revealed product attributes. Furthermore, (Weitzman 1979) has characterized a simple search strategy that computes the reservation utility of each product separately using the conditional belief of the hidden product attributes for that product only. As a result, with the assumption that the consumer uses an intuitive utility distribution for searching a refinement, he treats unvisited refinements in the same way as unsampled products, and (Weitzman 1979) gives us a complete characterization of the consumer search process within each search problem and hence overall.

We adopt the same one-step-ahead approach as (De los Santos & Koulayev 2016). We also adopt the intuitive utility distribution to characterize the benefit of searching a refinement. We differ, however, in how the current best offer is updated and in the range of search options being considered during the search process. In terms of the update of the current best offer, their model assumes that, after switching refinements, the consumer decides whether to click on at most one product on the refinement, captured in a reduced form fashion by a discrete choice model. As a result, all product attributes and random utility shocks are assumed known to consumers after they visit a refinement. In our model, we assume the consumer still engages in sequential search, because he only observes the partially revealed product attributes and does not know the hidden attributes and utility shocks. The update of current best utility is realized through selective sampling of products, which informs us about the product search costs, separately from the refinement search costs. In terms of the refinement choice, both models assume the consumer computes and compares the reservation utilities of the unvisited refinements for search decision, but in our model, the consumer additionally considers the reservation utilities of the unsampled products, and the set of
the unsampled products may increase as the consumer uncovers new partially revealed products as he visits a new refinement.

In summary, we only simplify our model structure for the choices the consumer does not have much information about, while still maintaining the overall sequential search framework and model expressiveness. We believe this selective simplification is consistent with our central modeling principle to capture the consumer information and cost structure more realistically, while accounting for model tractability. Our model extends that of (De los Santos & Koulayev 2016). It grants us the flexibility to have distinct search costs for refinement visiting and product clicking, and allows the change in the platform design to affect refinement visiting through cognitive cost and product sampling through partially revealed information. This flexibility is crucial to answering our managerial question.

8 Estimation

Based on the model specification in Section 7 and the characterization of the optimal search strategy, we first derive the model likelihood and describe the estimation strategy. Then we discuss model identification.

8.1 Likelihood and Estimation Strategy

For a given consumer $h$ and search session $se$, suppose he engages $B_{h, se} \geq 1$ search problems and performs $K_{h, se, b} \geq 1$ searches within each $b$th search problem. For the $b$th problem and $k$th search, the model has five state variables (suppressing subscripts $h, se, b, k$): the set of unsampled products on the current refinement $\bar{S}$ ($S$ for the sampled counterpart), the set of unsampled products on previously visited refinements $\bar{T}$ ($T$ for the sampled counterpart), the set of unvisited refinements $\bar{R}$, the current refinement $rf$, and the current best utility $u^*$. Note that $u^*_{h, se, b, k} = \max_{j \in S \cup T} u_{h, se, j}$.

For the $b$th problem and $k$th search (suppressing subscripts $h, rep, b, k$), the search options include $\bar{S}$, $\bar{T}$, and $\bar{R}$. Let $J = |\bar{S} \cup \bar{T} \cup \bar{R}|$ denote the total number of search options, let $R(n)$ denote the index of the search options with the $n$th largest reservation utility, and recall that $z_{R(n)}$ is the reservation utility for the search option that ranks $n$.

We incorporate the observed consumer search process into the indicator function
that simultaneously satisfies all the following conditions, specified by Equations (9)-(14), for a given utility shock realization \( \epsilon_{h,se} = (\epsilon_{h,j,se})_j \). Then the likelihood for the observed consumer search process is

\[
L_{h,se}(\beta, \gamma; \text{data}) = \int 1_{h,se}(\beta, \gamma, \epsilon_{h,se}; \text{data}) \, dF(\epsilon_{h,se}) .
\]

We suppress subscripts \( h, se, b, k \) below, because each quantity pertains to the relevant index in its own context. The conditions for the indicator function \( 1_{h,se}(\beta, \gamma, \epsilon_{h,se}; \text{data}) \) are specified as follows:

- For the \( b \)th search problem before the last one in the search session, that is, \( 1 \leq b < B \), the consumer does not stop the search process. By definition of our search problem, the last search in the \( b \)th search problem must be an option among \( \bar{T} \cup \bar{R} \) in order for the consumer to switch search problem, and all the other searches (if \( K > 1 \)) must be among \( \bar{S} \). Therefore,
  - Each \( k \)th search action out of \( K \) searches removes the search option from the unsearched set and implies
    \[
    z_{R(k)} \geq \max_{n=k+1} z_{R(n)} \tag{9}
    \]
    \[
    z_{R(k)} \geq u^*. \tag{10}
    \]
- If the consumer searches an unvisited refinement or clicks on a product from a previously visited refinement, he switches refinements and starts a new search problem. In particular, the consumer
  - Carries over the current best offer;
  - Updates his search costs for the unsampled products on the current and last-visited refinements, as described in Section 7.2;
  - Carries over all the search options he has discovered so far, if sampling a product from a previously visited refinement;
  - Merges all the search options he has discovered so far with a new set of partially revealed products, if searching an unvisited refinement.
• For the last search problem, namely, $b = B$, the consumer searches in the first $K$ steps among $\bar{S}$, stops at the $(K + 1)$-th step, and makes the purchase decision:

- Each $k$th search action out of $K$ searches removes the search option from the unsearched set and implies

\[
Z_{R(k)} \geq \max_{n=k+1} Z_{R(n)}
\]
\[
Z_{R(k)} \geq u^*.
\]

- Stopping at the $(K + 1)$-th step implies

\[
u^* \geq \max_{m=K+1} Z_{R(m)}.
\]

- Purchasing the $j$th option (0 for choosing the outside option) implies

\[
u_j \geq u^*.
\]

The likelihood of the model, aggregated over all individuals and search sessions, is given by

\[
L(\beta, \gamma; data) = \prod_h \prod_{se} L_{h,se}(\beta, \gamma; data).
\]

We use simulated maximum likelihood to estimate the model, and smooth over the likelihood using a logit-smoothed AR simulator [Hajivassiliou & Ruud 1994]. To integrate over the random utility shocks, we simulate a large number (3,000) of draws from random utility shock $\epsilon_{hj,se}$ to ensure a smooth likelihood surface and the stability of the estimator.

8.2 Identification

The model identification follows the same idea as in [Kim et al. 2010, 2016, Chen & Yao 2016, De los Santos & Koulayev 2016, Ursu 2016]. The refinement and click baseline costs are not separately identifiable from the baseline utility level. For identification purposes, the outside option utility is normalized to be $u_{h0,se} = \epsilon_{h0,se}$, and $\epsilon_{hj,se}$ is normalized to follow $N(0,1)$. Furthermore, purchase
decisions identify consumer preference coefficients, and cost measure variations identify consumer search cost coefficients.

9 Results

The estimated preference coefficients can be found in Table 8. Almost all coefficients are significant. We see consumers care most about access to outdoor swimming pool, gym, wifi in public areas, and free domestic long-distance calls. Consumers do not value some product attributes, including free bottled water, free map, free newspaper, and free toiletries (less than six items).

The estimated search cost coefficients can be found in Table 9. We normalize the coefficient estimates to RMB by dividing them by the magnitude of the price coefficient. To make the cognitive search cost more interpretable, we convert the cognitive cost consumers encounter in a typical search-result page into a monetary value as follows:

\[
\text{Search Cost Coeff (util/bits) } \times \frac{\text{Entropy Search Cost (bits)}}{\text{Price Coeff (util/RMB)}}.
\]

For a typical search-result page, the orderedness entropy of price, consumer rating, and other variables are 3.23 bits/page, 1.18 bits/page, and 4.63 bits/page, respectively. Thus, consumers typically pay the cognitive cost of 41.68 RMB, 3.99 RMB, and 79.51 RMB, respectively, per refinement visit, or 125.18 RMB (18.68 USD) in total.

Furthermore, the share of loading time cost in the overall search cost is much smaller. A typical search-result page takes 2.25 sec to load, and consumers pay the time cost of 0.07 RMB (0.01 USD) per refinement visit. A typical product-detail page takes 1.36 sec to load, and consumers pay the time cost of 0.04 RMB (0.006 USD) per click. Time cost is estimated to be 110.77 RMB, or 16.53 USD, per hour on average.

Our estimated position effect is 0.42 RMB (0.06 USD), which is equivalent to the cost of 10.5 RMB (1.5 USD) if consumers scroll to the end of a refinement page with 25 products on it. This estimated position effect is close to that discovered by (Chen & Yao 2016), who find a multiplicative position effect that inflates the overall search cost from 1.01 for the first product to 1.28 for the 25th product. In their model, a typical consumer incurs the overall search cost of 4.85 USD at position
Table 8: Preference Estimates. The RMB ¥ column is the estimates divided by the absolute value of the price coefficient, and the USD $ column is further converted from RMB to USD by dividing by the currency exchange rate of 6.7. Bold fonts indicate estimates significant at the 95% level.

1, and 5.31 USD at position 10, totaling approximately 0.05 USD for each position.

### 10 Managerial Implications

In this section, we study our opening managerial question of how much information to reveal in the outer layer. We approach this question by varying the revealed product attributes in the outer layer and leaving the remaining product attributes to the inner layer. Each such configuration corresponds to one information layout, which is applied to all refinement pages. Then for each information layout, we simulate the various aspects of the consumer search process using our estimated model, including the number of clicks, the number of searches, purchase probability, purchase session utility, purchased product price, revenue, and consumer welfare. We define revenue to be
<table>
<thead>
<tr>
<th>Refinement</th>
<th>Coefficients</th>
<th>Est</th>
<th>SE</th>
<th>RMB ¥</th>
<th>USD $</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>0.0014</td>
<td>1.16E-04</td>
<td>2.75</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>Price Entropy</td>
<td>0.0004</td>
<td>4.00E-05</td>
<td>0.86</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Consumer Rating Entropy</td>
<td>0.0004</td>
<td>2.02E-04</td>
<td>0.81</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Other Entropy</td>
<td>0.0002</td>
<td>3.00E-05</td>
<td>0.44</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Price Entropy (Sq)</td>
<td>0.0019</td>
<td>1.60E-05</td>
<td>3.73</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>Consumer Rating Entropy (Sq)</td>
<td>0.0011</td>
<td>1.05E-04</td>
<td>2.19</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>Other Entropy (Sq)</td>
<td>0.0019</td>
<td>6.00E-06</td>
<td>3.61</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Time (Sec)</td>
<td>1.60E-05</td>
<td>6.00E-06</td>
<td>0.03</td>
<td>0.005</td>
</tr>
<tr>
<td>Click</td>
<td>Base</td>
<td>0.0408</td>
<td>1.01E-04</td>
<td>78.77</td>
<td>11.76</td>
</tr>
<tr>
<td>Position</td>
<td>0.0002</td>
<td>9.00E-06</td>
<td>0.42</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Search Cost Estimates. The RMB ¥ column is the estimates divided by the absolute value of the price coefficient, and the USD $ column is further converted from RMB to USD by dividing by the currency exchange rate of 6.7. Bold fonts indicate estimates significant at the 95% level.

the purchased product price accounting for purchase probability. We also define consumer welfare to be the purchase session utility, which is the utility of the purchased product or the outside option conditional on purchase decision, net of the total accumulated search cost. These behavioral aspects provide us another opportunity to further understand the properties of the model. Our final aim is to study the relationship between consumer welfare and platform revenue, which are often in conflict, and which platforms hope to balance. We will try to characterize the Pareto frontier of this relationship, which will guide us in answering the managerial question. The rest of the section follows this order of presentation.

10.1 Counterfactual Setup

Due to the large space of possible configurations (about 8.3 million), we randomly sample information layouts. We first generate layouts that cover the extreme cases, from revealing all attributes to revealing only one attribute. Then we randomly sample 12,000 information layouts. Once we have the general configuration space covered through randomization, we improve the resolution near the Pareto frontier between consumer welfare and platform revenue by adding and removing up to four attributes from the layouts on the sampled Pareto frontier. This procedure allows us to better construct the Pareto frontier.

Fixing a particular sampled information layout and using the estimated model, we simulate 500 runs of the consumer search process by generating random utility shocks. Due to the dynamic nature of product attributes, a given refinement will have different product availability and attribute
values at different times of visits. As a result, each simulation run will be based on the randomly sampled refinement realizations. We then average the behavioral aspects mentioned above over the 500 simulation runs as the typical consumer response to the given information layout.

10.2 Counterfactual Results

Figure 5 shows the various behavioral responses of the consumer search process as information layouts vary. Within any given plot, the horizontal coordinate of each point corresponds to the average orderedness entropy over all refinements on the search platform under a particular information layout, and the vertical coordinate corresponds to the average behavioral response simulated under that information layout.

A clear separation exists between two point clouds. The blue point cloud corresponds to the information layouts that do not reveal price, and the green point cloud corresponds to those that do reveal price. Because price is the most complex product attribute to understand, the information layouts in the blue point cloud have lower information complexity than those in the green point cloud, and hence are situated to the left in the horizontal coordinates. Putting the price attribute aside, the information layouts that have the same set of attributes (excluding price) revealed will have the same relative position within each of the two clouds. In other words, the leftmost points in each cloud correspond to the layouts that have few product attributes revealed, whereas the rightmost points correspond to those that have the most attributes revealed. We see that any two such identical information layouts (excluding price) differ in the level of the behavioral responses. The solid lines are the nonparametric regression fit to the point clouds.

The separation of the two point clouds shows price is an important product attribute whose presence in the outer layer will dramatically change the consumer search behavior. This phenomenon is not only because price is the most complex attribute to understand, but also because it has considerable uncertainty and the presence of this attribute in refinements will change the benefit of the product clicking behavior. Note that high uncertainty is not necessarily correlated with difficulty to understand, because a variable with high uncertainty but displayed in sorted order may be easier to understand than a moderately uncertain variable displayed in random order. We will further explain the impact of the presence of the price attribute when we discuss each behavioral response below.
To understand the product clicking behavior in the “Number of Product Clicks” figure, we note an inverse relationship exists between the information complexity and the conditional standard deviation of the unrevealed product utility: as the outer layer reveals more attributes, each refinement page has greater information complexity, and each product has less uncertainty for the consumers; furthermore, the conditional standard deviation of the unrevealed product utility becomes lower. Because the conditional standard deviation of the unrevealed utility captures the marginal benefit of clicking, this inverse relationship implies that as the information complexity increases, the benefit of clicking decreases, and hence the number of clicks decreases. The difference between the two point clouds comes from the fact that revealing price in the information layouts reduces the conditional standard deviation of the unrevealed utility, hence lowering the number of product clicks.

The “Number of Refinement Searches” figure is a direct illustration of how cognitive cost affects the consumer search process: as information complexity increases, the marginal cost of visiting additional refinement increases, and hence the number of refinement searches decreases. The two point clouds are different because revealing price adds to the cognitive cost and hence reduces the number of refinement searches in the green point cloud. This finding shows that, when considering platform design, search platforms need to take into account the cognitive cost embedded in information layouts.

The previous two results explain the pattern in the “Purchase Probability” figure: as the consumer searches less, he is less likely to find what he would like to purchase. As a result, we see purchase probability decreases as information complexity increases. Also, purchase probability is lower, at the same relative position, in the information layouts that reveal price versus those that do not reveal price.

The “Purchase Session Utility” figure shows how consumer utility at the end of the search process responds to information layouts in a purchase session. Here, consumer utility is defined to be the purchased product or the outside option utility, depending on the consumer purchase decision. Recall that this purchase session utility, like all the other behavioral statistics, is an average quantity over simulation runs under a given information layout, so it is the expected consumer utility as he goes through the search process. On average, consumer utility is higher when the consumer searches and clicks more. This finding is true both within and across the two point clouds.

The “Purchased Product Price” figure shows the response of the chosen product price conditional
on purchase with respect to the changes in information layouts. When consumers search within information layouts that reveal price, they may directly select on price. As information complexity increases, consumers search fewer refinements, and hence reduce their price selection ability. This leads to higher transaction prices as information complexity increases.

On the other hand, when consumers search within information layouts that do not reveal price, they need to form rational predictions about the product price given the revealed product attributes. This lack of price information limits consumers’ ability to select on price, and the positive search cost leads to general higher purchase prices when price is not revealed than when it is revealed. Furthermore, as more attributes are revealed, consumers may use that information to predict price more accurately, and hence select on expected price, leading to the downward sloping trend of the transaction price.

The “Revenue” figure shows the response of revenue to the change in information layouts. Recall that we define revenue to be the purchased product price accounting for purchase probability, so the observed pattern is a natural outcome of the respective figures we have seen above. In general, we see revenue decreases as more information is revealed, and not revealing price generates more revenue than revealing price does. Also, not revealing price makes revenue much more responsive to the change in information complexity than revealing price does.

Search platforms care not only about revenue, but also about consumer welfare, because arguably consumer welfare creates customer loyalty, which makes consumers more likely to return to the search platform even when their outside options increase. Recall that we define consumer welfare to be the purchase session utility (defined above) net of the total accumulated search costs. The “Consumer Welfare” figure shows the response of consumer welfare to the variation in information complexity. First, we observe that, in general, revealing price is more beneficial to the consumers than not revealing price. Second, providing consumers with more product information will improve consumer welfare, but too much information provision is not beneficial either. We see that very low information provision makes consumers incur a large amount of product clicks and refinement visits on the extensive margin, which corresponds to high total search cost and welfare loss. At the other extreme, too much information causes consumers to stop searching prematurely, leading to utility loss, and eventually welfare loss. Hence, the optimal information layouts for consumer welfare are in the intermediate range of information complexity.
Now we focus on the relationship between revenue and consumer welfare as information complexity varies in Figure 6. Each point corresponds to one information layout, and the color of the point corresponds to the information complexity of the layout. The red end corresponds to layouts with extremely low information provision, and the blue end to layouts with extremely high provision. Again, two point clouds exist: the top point cloud corresponds to the layouts when product price is not revealed, and the bottom corresponds to when price is revealed. When product price is not revealed, revenue decreases and consumer welfare increases as information complexity increases from Figure 5, and hence we see a downward-sloping trend in the top point cloud. Because revenue changes less when product price is revealed, the same pattern holds with a smaller slope in the bottom cloud. The grey line traces out the Pareto frontier of revenue versus consumer welfare. We find a general tradeoff whereby, as information complexity increases in the outer layer, revenue decreases and consumer welfare reaches the maximum at an intermediate level.

Finally, the current information layout of the search platform under study is marked by the circle in the figure, which shows room for improvement in both revenue and consumer welfare. This room for improvement motivates us to characterize the win-win situations that benefit both the consumer and the search platform in the next section.

10.3 Pareto Improving Information Layouts

To understand how the search platform may increase revenue and consumer welfare compared to the current information layout, we select several Pareto improving information layouts, and compare the associated information complexity, consumer welfare, and platform revenue in Table 10. In particular, we compare layouts that are minimalist style, welfare maximizing, and revenue maximizing, as well as two win-win scenario layouts to the current layout. We define minimalist style layout to be the one that reveals the fewest product attributes, given that it is a Pareto improving layout.

We see the minimalist style layout simplifies the current layout by removing three attributes from display. This simplification increases consumer welfare by 15.28% while keeping the current revenue level intact. In the consumer welfare maximizing layout, the platform may in addition display two more attributes that consumers care about to reduce search cost. This addition will increase consumer welfare by 95% without losing revenue. In the revenue maximizing layout, some important
Figure 5: This figure shows the various behavioral responses of the consumer search process as information layouts vary. Within any given plot, the horizontal coordinate of each point corresponds to the average orderedness entropy over all refinements on the search platform under a particular information layout, and the vertical coordinate corresponds to the average behavioral response simulated under that information layout.
Figure 6: This figure shows the relationship between revenue and consumer welfare as information complexity varies. Each point corresponds to one information layout, and the color of the point corresponds to the information complexity of the layout. The grey line traces out the Pareto frontier of revenue versus consumer welfare. The current information layout of the search platform under study is marked by the circle in the figure.
product attributes, including price, are replaced with less important ones. The main driver of the revenue hike is that price is not revealed, which will reduce consumers’ direct selection on price. The addition of the less important attributes will deter consumers from searching excessively, avoiding too much search cost. This change in layout results in about a 60% increase in revenue without hurting consumer welfare.

Many win-win information layouts exist in which both consumer welfare and platform revenue may be improved. We have selected two such layouts, one revealing product price and the other not. When product price is revealed, we may remove some attributes that consumers care about, so that consumers are more likely to choose the higher priced product, using price as predictors for the important unrevealed attributes. This change in layout results in a 46% increase in consumer welfare and a 25% increase in revenue. When price is not revealed, we may reveal some additional attributes to consumers to compensate for price and prevent consumers from engaging in excessive search. This change in layout leads to a 19% increase in consumer welfare and a 53% increase in platform revenue.

11 Conclusion

In this paper, we study the research question of how consumers acquire information at the expense of time and cognitive costs, in addition to the search costs we already understand from the search literature. Answering this question also helps us address a new managerial question of the optimal amount of information to reveal in the search-result pages of search platforms. To answer these questions, we collect a rich data set from a large OTA search platform. To quantify cognitive cost, we innovate a measure called orderedness entropy. We find through model-free analysis that (1) consumers actively use search-result pages as an important source of information, (2) loading time cost is a small yet non-negligible cost in the consumer search process, and (3) orderedness entropy is a good information complexity measure for cognitive cost, and consumers are more likely to sort when sorting induces a higher reduction in cognitive cost, leaving the content of the refinement pages unchanged.

We then conceptualize the consumer search process by proposing a new structural model to capture both the refinement visiting and product clicking decisions, to distinguish between what
<table>
<thead>
<tr>
<th></th>
<th>Current</th>
<th>Minimalist</th>
<th>Max Welfare</th>
<th>Max Revenue</th>
<th>Win-Win 1</th>
<th>Win-Win 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Complexity</td>
<td>0.0663</td>
<td>0.0484</td>
<td>0.0787</td>
<td>0.0629</td>
<td>0.0574</td>
<td>0.0566</td>
</tr>
<tr>
<td>Consumer Welfare</td>
<td>0.0563</td>
<td>0.0649</td>
<td>0.1097</td>
<td>0.0620</td>
<td>0.0821</td>
<td>0.0670</td>
</tr>
<tr>
<td>Revenue</td>
<td>67.11</td>
<td>68.56</td>
<td>67.48</td>
<td>107.12</td>
<td>83.73</td>
<td>102.47</td>
</tr>
<tr>
<td>Price</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Consumer Rating</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Hotel Star</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Breakfast</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Cancel Allowed</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Wifi in room</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Kingsize bed</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Wi-Fi in public areas</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Parking</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Meeting rooms</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Gym</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Restaurant</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Airport pickup service</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Outdoor swimming pool</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Free toiletries (less than six items)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Free toiletries (more than six items)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Free domestic long-distance calls</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Free bottled water</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Free map</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Free newspaper</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Pets Allowed</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Free Afternoon Tea</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Free Guide Service</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 10: We select several Pareto improving information layouts, and compare the associated information complexity, consumer welfare, and platform revenue in order to understand how the search platform may increase revenue and consumer welfare compared to the current information layout. An “x” in the layout column represents the revelation of the respective product attribute.
consumers know before and after visiting refinements and clicking products, and to decompose search costs further into time and cognitive components. Our model uses a sequential search framework to allow both refinement visiting and product clicking as search options, while adopting the same one-step-ahead simplification as in \cite{DeLosSantos2016} for model tractability. We find through model estimation that (1) cognitive cost is a major component of all search costs, (2) the share of the loading time cost in the search costs is much smaller compared to cognitive cost, and (3) our estimated position cost is very close to the position cost estimated by \cite{Chen2016}.

Finally, we vary the information revealed in the outer layer to characterize the various changes in the consumer search process, and study the relationship between consumer welfare and platform revenue. We find price is an important product attribute whose presence in the outer layer will dramatically change consumer search behavior. On the Pareto frontier of this relationship, we also find a general tradeoff whereby consumer welfare increases and revenue decreases as more information is revealed. For our OTA under study, we find room for Pareto improvement in both revenue and consumer welfare by changing which product attributes are revealed.

We provide a framework to study the optimal information layout of platform design, which could be used to improve the study of optimal ranking in a unified way. The study of optimal information layout complements the study of optimal ranking on search platforms, because if the default platform-suggested ranking ignores the cognitive cost component, the refinement pages that use this ranking may appear too complicated for consumers to understand. This difficulty will make consumers switch to some other user-specified ranking, and hence reduces the full potential of the ranking algorithm. The joint study of the optimal layout and ranking will be the subject of future research.
References


**URL:** [https://www.interaction-design.org/literature/article/simplicity-in-design-4-ways-to-achieve-simplicity-in-your-designs](https://www.interaction-design.org/literature/article/simplicity-in-design-4-ways-to-achieve-simplicity-in-your-designs)