Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue

Anindya Ghose, Panagiotis G. Ipeirotis, Beibei Li

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–23
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2013.1828

Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue

Anindya Ghose
Department of Information, Operations and Management Sciences, Department of Marketing, Stern School of Business, New York University, New York, New York 10012, aghose@stern.nyu.edu

Panagiotis G. Ipeirotis
Department of Information, Operations and Management Sciences, Stern School of Business, New York University, New York, New York 10012, panos@stern.nyu.edu

Beibei Li
Information Systems and Management, Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, beibeili@andrew.cmu.edu

In this paper, we study the effects of three different kinds of search engine rankings on consumer behavior and search engine revenues: direct ranking effect, interaction effect between ranking and product ratings, and personalized ranking effect. We combine a hierarchical Bayesian model estimated on approximately one million online sessions from Travelocity, together with randomized experiments using a real-world hotel search engine application. Our archival data analysis and randomized experiments are consistent in demonstrating the following: (1) A consumer-utility-based ranking mechanism can lead to a significant increase in overall search engine revenue. (2) Significant interplay occurs between search engine ranking and product ratings. An inferior position on the search engine affects "higher-class" hotels more adversely. On the other hand, hotels with a lower customer rating are more likely to benefit from being placed on the top of the screen. These findings illustrate that product search engines could benefit from directly incorporating signals from social media into their ranking algorithms. (3) Our randomized experiments also reveal that an "active" personalized ranking system (wherein users can interact with and customize the ranking algorithm) leads to higher clicks but lower purchase propensities and lower search engine revenue compared with a "passive" personalized ranking system (wherein users cannot interact with the ranking algorithm). This result suggests that providing more information during the decision-making process may lead to fewer consumer purchases because of information overload. Therefore, product search engines should not adopt personalized ranking systems by default. Overall, our study unravels the economic impact of ranking and its interaction with social media on product search engines.

Keywords: travel search engine; randomized experiments; hierarchical Bayesian methods; information systems; IT policy and management; electronic commerce
History: Received May 28, 2012; accepted May 12, 2013, by Lorin Hitt, information systems. Published online in *Articles in Advance*.

1. Introduction

Over the last few decades, search engines have emerged as a significant channel for promoting and selling products. In information search engines (e.g., Google) the ranking of the search results is an immediate signal of the relevance of the result to the query. However, in *product* search engines, the ranking of the displayed products is often based on criteria such as price, product rating, etc. In such a setting, we may often have multiple, potentially conflicting signals given to the customer about the products' rankings. For example, if we rank by price, then the cheapest products sometimes have low product ratings, or products appearing on top of the list may be too expensive for the customer. Effectively consumers have to observe multiple, competing ranking signals and come up with their own ranking in their minds; in some settings, the product search engine will also generate personalized results, trying to rank the products according to the preferences of the consumer. In such an environment, we want to understand which factors influence the decision-making process of the customers and the magnitude of that influence. Are consumers influenced by the display ranking order, by the product rating, by price, and to what degree? How does this interplay affect the revenue that a search engine can generate?

1.1. Related Work

In the last 10 years, the literature in e-commerce has shown the existence of a strong primacy effect in environments wherein consumers make choices among offers displayed in *information search engines* such as Google, Yahoo, or Bing. Specifically, we have learned that an online position effect exists and that rank order has a significant impact on the click-through rates and conversion rates (e.g., Ghose and Yang 2009, Yang and Ghose 2010, Agarwal et al. 2011, Jerath et al. 2011, Rutz and Trusov 2011, Narayanan and Kalyanam 2011, Animesh et al. 2011, Baye et al. 2012, Jeziorski and Segal 2012, Rutz et al. 2012, Abhishek et al. 2013, Ghose et al. 2013a). These papers focused primarily on evaluating the effect of screen position on user behavior, controlling for the quality of the advertisement. However, in *product search engines*, the observed demand patterns can be influenced by the joint variation in product ratings (either professional rating or user rating) and online screen position. The first goal of our study is to examine the position effect in product search engines, conditional on its interaction with product ratings.

Search engines are beginning to adopt signals from social media sites directly into their ranking mechanism design (e.g., Bing Social Search, TripAdvisor). Recent work has found that a utility-based ranking mechanism on product search engines that incorporates multidimensional consumer preferences and social media signals can lead to significant surplus gain for consumers (Ghose et al. 2012). However, given that price was not the top priority considered in the ranking recommendation, whether such a mechanism can actually benefit product search engines is unclear because their revenues are normally commission based. Therefore, the second goal of our study is to examine the effect of different ranking mechanisms on product search engine revenue.

Outside of search, one of the most important ways for shoppers to discover products has been through recommendation engines (Chittor 2010). However, although some online retailers use recommendation systems, many product-specific search engines (e.g., travel search engines) still do not provide personalized ranking results in response to consumer queries, presumably because these product search engine companies are unsure whether providing extra information to consumers will lead to an increase in profit. Existing research holds two different opinions on the effects of personalization. One stream of work is supportive of personalization (e.g., Rossi et al. 1996, Ansari and Mela 2003, Arora and Henderson 2007, Yao and Mela 2011), whereas another stream of work is a bit more skeptical (e.g., Zhang and Wedel 2009, Aral and Walker 2011, Goldfarb and Tucker 2011, Lambrecht and Tucker 2013), suggesting that although personalization can lead to higher customer satisfaction and profits, it will not work as well universally.[1] However, none of these papers have examined the effect of information availability and personalization in a search engine context. Koulayev (2014) examines consumers' costly search behavior on travel search engines through the formation of consideration sets. Ghose et al. (2013b) build a structural econometric model to predict individual consumers' online footprints on product search engines to improve user experiences under the context of social media overload. Chen and Yao (2012) use secondary data to examine how the sorting and filtering tools on travel search engines influence consumer hotel search. They find these tools result in a significant increase in total search activities, but they also lead to lower overall welfare because of the disproportional engagement induced by the refinement tools. With these findings in mind, our third goal is to examine how different kinds of personalized ranking mechanisms in product search engines affect consumer behavior and search engine revenues. Specifically, does allowing users to interact with the ranking algorithm to proactively personalize their search results lead to more or fewer purchases?

1.2. Contributions and Results

We situate our study in a travel search engine context, looking specifically at consumer selection of a hotel. We first apply archival data analysis to gain insights into the product-rating effects and ranking effects on consumers' click and purchase behaviors. Using a panel data set from November 2008 to January 2009 containing approximately one million online user search sessions—including detailed information on consumer searches, clicks, and transactions obtained from Travelocity—we propose a hierarchical Bayesian framework in which we build a simultaneous equation model to jointly examine the interrelationship between consumers' click and purchase behavior, search engine ranking decisions, and customers' ratings.

Toward the first goal, we examine the variation in the ratings of different hotels (both hotel "class" rating and customer rating) at the same rank on the travel search engine over time. In addition, our data setting has variation in the rank of the same hotel over time because the same hotel appears at different positions at different points in time. Controlling for room prices, such variation allows us to model the interaction effect of hotel class and customer ratings with rank and to measure its effect on demand.

Toward the second goal, we examine how different ranking mechanisms affect the search engine revenue. We achieve this goal by conducting a set of

[1] For a good review of the stream of work on personalization, refer to Arora and Henderson (2007).

policy experiments. We consider six different ranking designs: utility, conversion rate (CR), clickthrough rate (CTR), price, customer rating, and the Travelocity default algorithms. Then we estimate our model and predict future search engine revenues under each ranking mechanism.

Toward our third goal, we examine how different levels of personalized ranking mechanisms affect consumer behavior and search engine revenue. Particularly, we compare two types of personalization mechanisms used to drive the ranking of results in response to a query: *active personalized ranking* and *passive personalized ranking*. In our context, a ranking system that allows consumers to proactively interact with the recommendation algorithm prior to the display of results from a search query is classified as "active." By contrast, a ranking system that does not allow customers to interact with the recommendation algorithm is classified as "passive."

As of today, no hotel search engine has explicitly adopted a personalization-based approach to hotel ranking because they are still grappling with the issue of whether such an approach is useful.[2] Hence, to our knowledge, no archival data in any product search engine have information on the effect of personalized ranking on user behavior. Therefore, we designed randomized experiments using a hotel search engine application that we built. Our randomized experimental results are based on a total of 900 unique user responses over a two-week period via the Amazon Mechanical Turk (AMT) crowdsourcing platform. We use a customized behavior-tracking system to observe the detailed information of consumers' search, evaluation, and purchase decision-making process. By manipulating the default ranking method and by enabling or disabling a variety of personalization features on the hotel search engine website, we are able to study the effect of personalized ranking on consumer behavior.

Our archival data analysis and randomized experiments are consistent in demonstrating the following: (1) A utility-based ranking mechanism can lead to a significant increase in the overall search engine revenue. (2) Significant interplay occurs between search engine ranking and product ratings. An inferior rank affects "higher-class" hotels more adversely. On the other hand, hotels with a lower customer rating are more likely to benefit from being placed on the top of the screen. These findings illustrate that product search engines could benefit from directly incorporating signals from online social media into the ranking algorithms. (3) Our randomized experiments also reveal that an active personalized ranking mechanism that enables consumers to specify both search context and individual preferences leads to more clicks but lower purchase propensities and lower search engine revenue, compared with passive personalized ranking mechanisms. A plausible explanation is related to theories of consumer cognitive cost. Prior theoretical work has shown that information overload and nonnegligible search costs can discourage decision makers from evaluating choices, leading to a scenario where they make no choices at all (Kuksov and Villas-Boas 2010). Our empirical finding dovetails with the theoretical conclusion by Kuksov and Villas-Boas that providing more information can actually lead to fewer purchases. It is also consistent with Dzyabura (2014), who shows that consumers who do not have well-formed preferences at the start of their search may be better off with uncertainty about product attribute levels rather than perfect knowledge of the attributes of all available products. Therefore, although an active personalized ranking recommendation may help consumers discover what they want to buy, product search engines should not ubiquitously adopt it.

Two recent studies that are closely related to the current paper are Ghose et al. (2012) and Ghose and Yang (2009). However, this current paper distinguishes itself from the two previous studies in the following ways: (1) Ghose et al. (2012) do not focus on how rankings can benefit the search engine companies (in addition to the customers)—i.e., is it profitable for a search engine company to implement a utility-based ranking mechanism? In particular, we focus on examining whether the utility-based ranking leads to a significant improvement in the CTR, CR, and the total revenue for search engines. (2) Ghose et al. (2012) examine the direct WOM (word-of-mouth) effect on demand, without considering the impact of rank on the search engine. However, in our paper, we examine the WOM effect conditional on the ranking position of the product on search engines. We focus on examining the interaction effect between ranking and product ratings (both professional hotel class rating and online customer rating). Our findings illustrate that product search engines could benefit from directly incorporating signals from online social media into the ranking algorithms. (3) In Ghose et al. (2012), the authors did not focus on personalized rankings. Our randomized experiments reveal that an active personalized ranking mechanism that enables consumers to specify both search context and individual preferences leads to more clicks but lower purchase propensities and lower search engine revenues, compared with a passive personalized ranking mechanism. Therefore, although active personalized ranking recommendation may help consumers discover

[2] This finding is based on our personal communication with Travelocity.

what they want to buy, it should not be adopted ubiquitously. (4) Ghose and Yang (2009) study the effect of keyword ranking on CTR and CR in the sponsored search context. However, our current paper looks at a different research context of ranking in product search engines. Compared with Ghose and Yang (2009), we make two method-based improvements in this paper. First, in addition to CTR, CR, and ranking, we model customer rating as a fourth dependent variable in the simultaneous model framework. Second, we allow for unobserved heterogeneity in all time-varying covariates. Our model fitness comparison results show that the model with full heterogeneity on all time-varying variables provides the overall best performance.

2. Data

Our data set consists of detailed information on a total of 969,033 online sessions from Travelocity.com, including consumer searches, clicks, and conversions that occurred within these sessions between November 2008 and January 2009. In addition, we have hotel-related information, such as hotel class, brand, online reviewer rating, and number of reviews. We collected customer reviews from Travelocity.com. We collected the online reviews and reviewers' information on a daily basis up to January 31, 2009 (the last date of transactions in our database). This process provides us with a final data set containing 29,222 weekly observations for 2,117 hotels in the United States.[3]

We define an "online session" to capture a set of activities by an online user, identified by a unique cookie. In our data, a starting indicator and an ending indicator with a corresponding time stamp (provided by the company) can characterize each unique online session. More specifically, a typical online session involves the initialization of the session, the search query, the results (in a particular rank order) returned from that search query, the sorting method, the click(s) on hotel(s) if any exist, the login and actual transaction(s) if any conversion occurs, and the termination of the session. The ending indicator marks the termination of a session.

[3] We aggregate our data to a weekly level mainly to make them computationally tractable. For a robustness check, we have also tried using data from a daily level directly. Because of the size of the data set (approximately one million user sessions with more than 14 million individual events [impressions, clicks, or conversions]), we randomly select 10% of the observations from our original data set. We then conduct the estimation on the random selected sample at a daily level. We find the estimated coefficients are qualitatively consistent with the ones from a weekly level. We have also selected 15%, 20%, and 25% of the observations to form different random samples. We find the results are similar. We provide the estimation results from the 10% sample at a daily level in Online Appendix C (online appendices available at http://www.andrew.cmu.edu/user/beibeili/HotelExperiments-app.pdf).

We count a "display" for a hotel if that hotel appears visible to a consumer on the webpage in an online search session. Meanwhile, we count a "click" if a consumer selects the hotel and a "conversion" if a consumer has completed the payment in that online session. We only consider sessions with at least one display.[4] A display can lead to a click, but it may not lead to a purchase. Each hotel that counts for a display is associated with a page number and a screen position, which capture the corresponding page order and (within-page) rank order of that hotel in the search results. Note that Travelocity only shows 25 hotels per page when it displays the hotel search results on a webpage.[5] This design restricts the rank order for each hotel within the range from 1 to 25. Meanwhile, to facilitate consumer search, Travelocity provides a sorting criterion called "Travelocity Pick" by default. It also provides multiple alternative sorting criteria: price, hotel class, hotel name, and customer review rating. To capture consumers' particular sorting preferences that may potentially influence the position effect, we include a set of control variables in our study to indicate how frequently a hotel appears in a result list under different sorting criteria. In particular, we use a vector (*SpecialSort*) that contains six control variables to capture the frequency of six sorting criteria that consumers use during their searches: default (*DFT*), price ascending (*PRA*), class descending (*CLD*), class ascending (*CLA*), city name ascending (*CNA*), and hotel name ascending (*HNA*).

In summary, each observation in our data set contains the hotel ID, week ID, number of competing hotels, number of displays, number of clicks, number of conversions, average screen position (i.e., rank on the result page), average page number, and corresponding hotel characteristics in that week. For a better understanding of the variables in our setting, we present the definitions and the summary statistics of our data variables in Table 1.

3. Empirical Model

In this section, we discuss how we develop our simultaneous model in a hierarchical Bayesian framework. Then we describe how we apply the Markov chain Monte Carlo (MCMC) methods (Rossi and Allenby 2003) to empirically identify the effects of product quality and ranking position on consumer search and purchase behavior. More specifically, our model is

[4] In some cases, users may initiate a session and look for general travel information, such as the area of the city, rather than search for any hotels; thus, no hotels will be displayed on any webpage. We exclude such sessions from our analysis.

[5] Recently, Travelocity upgraded its webpage design by showing 10 hotels per page. However, during our examination time period, the number was still 25.

Table 1 Definitions and Summary Statistics of Variables

Variable	Definition	Mean	Std. dev.	Min	Max
Price	Transaction price per room per night	120.45	73.25	25.77	978
Display	Number of displays	213.65	382.28	1	4,849
Click	Number of clicks	2.99	3.55	0	56
Conversion	Number of conversions	1.26	0.66	0	9
Page	Page number of the hotel	20.86	13.44	1	192
Rank	Screen position of the hotel within a page	12.09	4.32	1	25
Class	Hotel class	3.36	1.37	1	5
ReviewCnt	Total number of reviews	21.06	29.28	1	202
Rating	Overall reviewer rating	3.84	0.85	1	5
SpecialSort	Vector of six control variables indicating the frequency of using different sorting methods				
DFT	Default sorting	188.50	369.58	0	4,711
PRA	Price ascending	13.99	23.34	0	338
CLD	Class descending	1.49	3.42	0	37
CLA	Class ascending	0.16	0.65	0	11
CNA	City name ascending	0.13	0.54	0	9
HNA	Hotel name ascending	0.35	0.95	0	15
H	Total number of hotels in a city	24.03	56.48	1	922
Brand	Dummies for nine hotel brands: Accor, Best Western, Cendant, Choice, Hilton, Hyatt, Intercontinental, Marriott, and Starwood	—	—	0	1
	Number of observations (weekly level): 29,222	Time period: 11/1/2008–1/31/2009			

motivated by the work of Ghose and Yang (2009). The general idea is as follows: We propose to build a simultaneous equations model of clickthrough, conversion, rank, and customer rating. We model the clickthrough and conversion behavior as a function of hotel brand, price, rank, page, sorting criteria, customer rating, and hotel characteristics. The rank of a hotel is modeled as a function of hotel brand, price, sorting criteria, customer rating, hotel characteristics, and performance metrics such as previous conversion rate. The customer rating of a hotel is modeled as a function of hotel brand, price, rank, page, sorting criteria, and hotel characteristics. Each function contains an unobserved error that is normally distributed with mean zero. To capture the unobserved covariation among clickthroughs, conversions, rank, and customer rating, we assume the four error terms are correlated and follow the multivariate normal distribution with mean zero. We describe our model next.

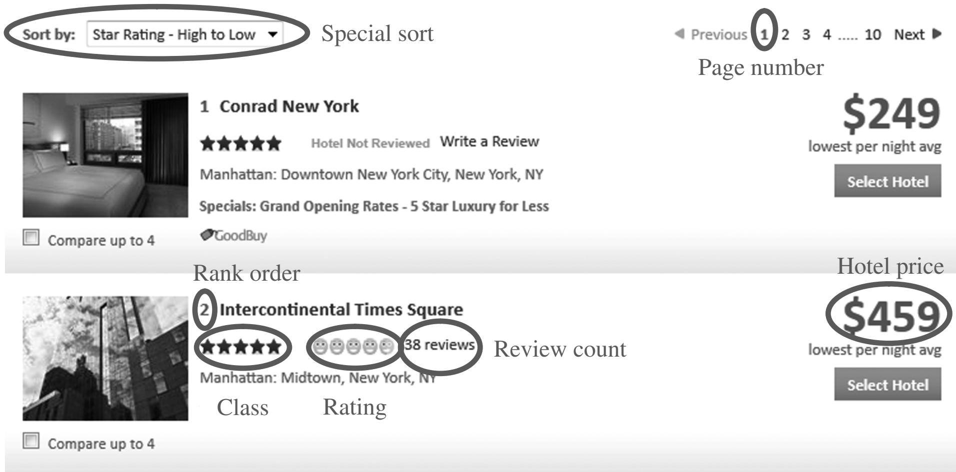
3.1. A Simultaneous Equation Model of Clickthrough, Conversion, Rank, and Rating

First, we define our unit of observation to be a "hotel-week." Thus, for hotel j in week t, we use n_{jt} to denote the clickthroughs among N_{jt} displays ($n_{jt} \leq N_{jt}$ and $N_{jt} > 0$). We also denote with m_{jt} the conversions among the n_{jt} clickthroughs ($m_{jt} \leq n_{jt}$). We further denote with p_{jt} the probability of having a clickthrough and with q_{jt} the probability of having a conversion, conditional on a clickthrough.

We model the clickthrough, conversion, rank, and customer rating simultaneously in a hierarchical Bayesian framework. In particular, we divide our model into four interactive components.

3.1.1. Clickthrough Rate Model. First, a consumer's decision to click on a hotel is based on the information available on the Travelocity search results page. Figure 1 provides a screenshot of a sample webpage of hotel search results on Travelocity. As denoted in Figure 1, information that enters the consumer decision-making process includes *hotel price*, *hotel class*, *reviewer rating*, *review count*, *rank order*, and *page number*. Prior literature has shown that *rank order* and *page number* are significant determinant of clicks on the results of a search engine query (e.g., Rutz et al. 2012, Ghose and Yang 2009, Jerath et al. 2011, Rutz and Trusov 2011, Ghose et al. 2013a). In addition, previous studies have found that *rank* has a significant and nonlinear effect in the context of keyword advertising (e.g., Ghose and Yang 2009, Agarwal et al. 2011). To account for the potential nonlinear ranking effect in hotel search, we consider an additional quadratic term of rank in the model. Recent theoretical work has argued that product price affects consumer actions, such as clicks and conversions, and search engine decisions (Dellarocas 2012). De los Santos and Koulayev (2013) and Yao and Mela (2011) have shown that user ratings affect clickthrough rates on search engines. Hence, we incorporate the volume and valence of reviews. Recent studies have shown that online search refinement tools such as the sorting selection menu can affect consumers' searches and intentions to purchase (Chen and Yao 2012). Therefore, to capture the effect associated with the search refinement tools and to control for consumers' particular sorting preferences, we include a vector $SpecialSort_{jt}$ that contains six control

Figure 1 Screenshot of the Search Result Page on Travelocity.com



variables to capture the frequency of six sorting criteria that consumers use during the search process for hotel j in week t. Moreover, previous research has shown that a product brand can influence consumers' perceptions of quality and willingness to buy (e.g., Dodds et al. 1991, Nevo 2001). Thus, we include hotel *brand* dummies to control for the unobserved hotel characteristics. Finally, prior literature has demonstrated that the number of competitors in the local market can affect consumers' clicks for a product online (e.g., Baye et al. 2009). Therefore, to control for the competition in the local market, we include the *total number of hotels* in j's city, H_j, as a control variable. This setting gives us the following equation:

$$p_{jt} = \frac{\exp(U^p_{jt})}{1+\exp(U^p_{jt})} \tag{1}$$

where $U^p_{jt} = \beta_{j0} + \beta_{j1} Rank_{jt} + \beta_{j2} Rank^2_{jt} + \beta_{j3} Page_{jt} + \beta_{j4} Price_{jt} + \beta_{j5} Rating_{jt} + \beta_{j6} ReviewCnt_{jt} + \alpha_1 Class_j + \alpha_2 H_j + \alpha_3 Brand_j + \alpha_4 SpecialSort_{jt} + \varepsilon_{jt}$. To capture the unobserved heterogeneity, we model β, the intercept and the coefficients for the time-varying variables, to be random coefficients:[6]

$$\beta = \begin{bmatrix} \bar{\beta}_{j0} \\ \cdots \\ \bar{\beta}_{j6} \end{bmatrix} + \Pi^\beta D_j + \begin{bmatrix} \sigma^\beta_{j0} \\ \cdots \\ \sigma^\beta_{j6} \end{bmatrix}, \tag{2}$$

where we assume each random coefficient to vary along its population mean and the hotel-specific characteristics. More specifically, D_j is a $d \times 1$ vector of observed hotel-specific characteristics. In our model, we consider three time-invariant variables that capture the hotel quality: *hotel class*, *average hotel price*, and *average reviewer rating* (i.e., $d = 3$). We consider Π^β to be a $Z \times d$ matrix of coefficients that measures how hotel utility varies with observed hotel characteristics (i.e., $Z = 7$ is the dimension of vector β). Moreover, we model the unobserved error terms to be correlated and to follow a multivariate normal distribution with mean zero in the following way:

$$[\sigma^\beta_{j0}, \ldots, \sigma^\beta_{j6}]' \sim MVN(0, \Sigma^\beta),$$
$$\text{where } \Sigma^\beta \text{ is a } 7 \times 7 \text{ covariance matrix.} \tag{3}$$

3.1.2. Conversion Rate Model. Second, we note that the set of features denoted in Figure 1 is the key determinant for a consumer's purchase decision making as well. Moreover, prior work has shown that price and quality as well as the volume and valence of online reviews will affect product sales (e.g., Chevalier and Mayzlin 2006, Ghose et al. 2012). Meanwhile, several studies have shown how screen position and page number are important factors that influence consumer demand on search engines (e.g., Rutz et al. 2012, Ghose and Yang 2009, Jerath et al. 2011, Rutz and Trusov 2011, Agarwal et al. 2011). Thus, we model the probability of a consumer's conversion as a function of the set of hotel price-, quality-, review- and screen-position-related factors: *hotel price*, *hotel class*, *reviewer rating*, *review count*, *rank order*, and *page number*. To account for the nonlinear effect of ranking effect, we include the quadratic term of rank order. Based on the previous findings that market competition (e.g., Baye et al. 2009), product brand (e.g., Dodds et al. 1991, Nevo 2001), and online consumer search refinement tools (Chen and Yao 2012) are key determinants of the elasticities of demand, we include the *total number of hotels, brand, and special sort*

[6] As a robustness check, we have tried an alternative model setting with partial heterogeneity by allowing only the intercept and the rank variable to be associated with random coefficients. We have considered a similar setting for the clickthrough model, conversion model, ranking model, and rating model. We find the estimation results are qualitatively consistent with our main model estimation results. We provide the results from the alternative model with partial heterogeneity in Online Appendix A.

as additional control variables. The conversion equation is written as follows:

$$q_{jt} = \frac{\exp(U^q_{jt})}{1+\exp(U^q_{jt})}, \tag{4}$$

where $U^q_{jt} = \gamma_{j0} + \gamma_{j1}Rank_{jt} + \gamma_{j2}Rank^2_{jt} + \gamma_{j3}Page_{jt} + \gamma_{j4}Price_{jt} + \gamma_{j5}Rating_{jt} + \gamma_{j6}ReviewCnt_{jt} + \theta_1 Class_j + \theta_2 H_j + \theta_3 Brand_j + \theta_4 SpecialSort_{jt} + \eta_{jt}$.

Similar to (3), we model γ as random coefficients with the following properties:

$$\gamma = \begin{bmatrix} \bar{\gamma}_{j0} \\ \cdots \\ \bar{\gamma}_{j6} \end{bmatrix} + \Pi^{\gamma} D_j + \begin{bmatrix} \sigma^{\gamma}_{j0} \\ \cdots \\ \sigma^{\gamma}_{j6} \end{bmatrix}. \tag{5}$$

In Equation (5), D_j also contains *hotel class*, *average hotel price*, and *average reviewer rating*. Moreover, we model the unobserved error terms in (5) to be correlated in the following way:

$$[\sigma^{\gamma}_{j0}, \ldots, \sigma^{\gamma}_{j6}]' \sim MVN(0, \Sigma^{\gamma}),$$
$$\text{where } \Sigma^{\gamma} \text{ is a } 7\times 7 \text{ covariance matrix.} \tag{6}$$

3.1.3. Ranking Model. Equations (1)–(6) model consumers' behavior of clickthrough and conversion. Meanwhile, we can model search engines' ranking decision. Prior research in keyword search advertising has found that both the bid price and the quality of the keyword affect ranking (e.g., Ghose and Yang 2009). Building on the previous findings along with our further interaction with Travelocity, we model the rank order of hotel j in week t as being dependent on the set of hotel price and quality characteristics. In particular, we use the *previous conversion rate*, $CR_{j,t-1}$, as a quality performance metric.[7] We consider the same set of control variables used in the previous consumer behavior models. The model is written as[8]

$$\begin{aligned} \ln(Rank_{jt}) = {} & \omega_{j0} + \omega_{j1}CR_{j,t-1} + \omega_{j2}Price_{jt} + \omega_{j3}Rating_{jt} \\ & + \omega_{j4}ReviewCnt_{jt} + \kappa_1 Class_j + \kappa_2 H_j \\ & + \kappa_3 Brand_j + \kappa_4 SpecialSort_{jt} + \upsilon_{jt}. \end{aligned} \tag{7}$$

Similarly, we model ω as random coefficients to vary along the population mean and the hotel-specific characteristics D_j, which contain *hotel class*, *average hotel price*, and *average reviewer rating*:

$$\omega = \begin{bmatrix} \bar{\omega}_{j0} \\ \cdots \\ \bar{\omega}_{j4} \end{bmatrix} + \Pi^{\omega} D_j + \begin{bmatrix} \sigma^{\omega}_{j0} \\ \cdots \\ \sigma^{\omega}_{j4} \end{bmatrix}. \tag{8}$$

Meanwhile, we model the unobserved error terms in (8) to be correlated in the following way:

$$[\sigma^{\omega}_{j0}, \ldots, \sigma^{\omega}_{j4}]' \sim MVN(0, \Sigma^{\omega}),$$
$$\text{where } \Sigma^{\omega} \text{ is a } 5\times 5 \text{ covariance matrix.} \tag{9}$$

3.1.4. Rating Model. Note that customer ratings on product search engines can be endogenous and often determined by many hotel-specific characteristics, such as price, class, brand, and so on. To account for the endogeneity of rating, we model it as the fourth dependent variable in the simultaneous framework. Prior work has shown that product price and product quality affect customer ratings (Li and Hitt 2010). Therefore, we model the customer rating of hotel j in week t as being dependent on the set of hotel price and quality-related characteristics. Meanwhile, we include the screen position and sorting method of the hotel in the last period to control for the visibility of the hotel. We also control for hotel brand and the total number of hotels in the local market:

$$\begin{aligned} Rating_{jt} = {} & \rho_{j0} + \rho_{j1}Rank_{j,t-1} + \rho_{j2}Rank^2_{j,t-1} \\ & + \rho_{j3}Page_{j,t-1} + \rho_{j4}Price_{jt} + \rho_{j5}ReviewCnt_{jt} \\ & + \chi_1 Class_j + \chi_2 H_j + \chi_3 Brand_j \\ & + \chi_4 SpecialSort_{j,t-1} + \psi_{jt}. \end{aligned} \tag{10}$$

We model ρ as random coefficients to vary along the population mean and the hotel-specific characteristics D_j. In the rating model, we consider D_j to contain *hotel class* and *average hotel price*:

$$\rho = \begin{bmatrix} \bar{\rho}_{j0} \\ \cdots \\ \bar{\rho}_{j5} \end{bmatrix} + \Pi^{\rho} D_j + \begin{bmatrix} \sigma^{\rho}_{j0} \\ \cdots \\ \sigma^{\rho}_{j5} \end{bmatrix}. \tag{11}$$

We model the unobserved error terms in (11) to be correlated in a similar fashion:

$$[\sigma^{\rho}_{j0}, \ldots, \sigma^{\rho}_{j5}]' \sim MVN(0, \Sigma^{\rho}),$$
$$\text{where } \Sigma^{\rho} \text{ is a } 6\times 6 \text{ covariance matrix.} \tag{12}$$

Finally, to capture the unobserved covariation and the potential endogenous relationship among clickthrough, conversion, rank, and rating, we assume the four error terms in Equations (1), (4), (7), and (10) to be correlated as follows:

$$[\varepsilon_{jt}, \eta_{jt}, \upsilon_{jt}, \psi_{jt}]' \sim MVN(0, \Omega_{jt}),$$
$$\text{where } \Omega_{jt} \text{ is a } 4\times 4 \text{ covariance matrix.} \tag{13}$$

[7] Using the prior conversion rate as a proxy for quality is similar to using the prior clickthrough rate (e.g., Ghose and Yang 2009). In addition, based on our communication with Travelocity, their default ranking is a function of commission based on previous revenue. Therefore, we tried alternative performance metrics such as revenue in the previous week, monthly averaged conversion rate, and monthly averaged revenue. The results are consistent across all these specifications.

[8] As a robustness check, we considered an alternative model using an ordered probit for the ranking model. We found the estimation results remain qualitatively consistent with the main model.

3.2. Likelihood Function

The consumer decision process involves two steps. In the first step, the consumer sees a hotel displayed on the search result webpage and decides whether or not to click on it. In the second step, if the consumer clicks on the hotel, he decides whether or not to purchase it. Accordingly, we would expect to observe three types of events:

(1) A consumer sees a hotel but does not click or purchase. The probability of such an event is $1-p_{jt}$.

(2) A consumer sees a hotel, clicks through, but does not purchase. The probability of such an event is $p_{jt}(1-q_{jt})$.

(3) A consumer sees a hotel, clicks through, and makes a purchase. The probability of such an event is $p_{jt}q_{jt}$.

Therefore, we can derive the probability of observing the joint occurrence of n_{jt} clickthroughs and m_{jt} conversions among N_{jt} displays, (n_{jt}, m_{jt}), to be the following:[9]

$$\begin{aligned}\Pr(n_{jt}, m_{jt} \mid p_{jt}, q_{jt}) &= C_{N_{jt}}^{n_{jt}} \cdot (p_{jt})^{n_{jt}} \cdot (1-p_{jt})^{N_{jt}-n_{jt}} \cdot C_{n_{jt}}^{m_{jt}} \cdot (q_{jt})^{m_{jt}} \cdot (1-q_{jt})^{n_{jt}-m_{jt}} \\ &= \frac{N_{jt}!}{m_{jt}!(n_{jt}-m_{jt})!(N_{jt}-n_{jt})!} \cdot (p_{jt}q_{jt})^{m_{jt}} \\ &\quad \cdot [p_{jt}(1-q_{jt})]^{n_{jt}-m_{jt}} \cdot (1-p_{jt})^{N_{jt}-n_{jt}}. \end{aligned} \tag{14}$$

Note that our simultaneous equations resemble a triangular system (Lahiri and Schmidt 1978, Hausman 1975, Greene 1999). Such a triangular system allows us to identify our parameters in a similar fashion as in other recent studies in sponsored search (e.g., Ghose and Yang 2009, Agarwal et al. 2011). In addition, to ensure that our parameter estimates are accurate, we have simulated the clicks, conversions, rank, and customer rating for each hotel according to the model and the actual independent variables observed in our data. By repeating the estimation with this simulated data set, we were able to recover our parameter estimates. This step provides empirical support that our parameters are fully identified.

4. Empirical Analyses and Results

To estimate our model, we applied the MCMC methods using a Metropolis–Hastings algorithm with a random walk chain (Chib and Greenberg 1995). In particular, we ran the MCMC chain for 80,000 iterations and used the last 40,000 iterations to compute the mean and standard deviation of the posterior distribution of the model parameters. We provide more details on the MCMC estimation algorithm in Online Appendix D.

4.1. Clickthrough Rate Model

First we present the results of the clickthrough model in panel (a) of Table 2. All coefficients are statistically significant. The coefficients of both *Rank* and *Page* are negative and statistically significant, confirming a position effect does exist. A hotel that appears on an earlier page in the search results or on a higher position on the screen will receive more clicks than a hotel that appears on a latter page or on a lower position. A one-position increase in rank leads to a 10.07% increase in clickthroughs on average. Moreover, we found a positive coefficient on the quadratic term of rank, suggesting the negative effect of rank on CTR increases at a decreasing rate. Consistent with theory and existing empirical findings (e.g., Baye et al. 2009), *Price* has a negative sign, showing the higher the price of a hotel, the lower the willingness of consumers to click on that hotel. *Class* has a positive sign, showing the higher the hotel class, the lower the CTR.

Interestingly, we found that the interaction effect between *Rank* and *Class* is negative and statistically significant (i.e., −0.026). The interaction effect between *Rank* and *Price* is also statistically significantly and negative (i.e., −0.019). However, the interaction effect between *Rank* and *Rating* is statistically significant and positive (i.e., 0.020). These findings indicate that higher-class or more expensive hotels are more sensitive to the online ranking effect. They tend to be more adversely affected by an inferior screen position (e.g., at the lower part of the screen). On the other hand, hotels with lower online user ratings are more likely to benefit from being placed on the top of the search results, an effect that also benefits the underlying search engine that is typically paid by clickthrough or conversion.[10] This finding illustrates the need for product search engines to directly incorporate signals from online social media into the ranking algorithms.

4.2. Conversion-Rate Model

The coefficient estimates from the conversion model are presented in panel (b) of Table 2. Most of the coefficients are statistically significant. *Rank* and *Page* have a negative and statistically significant effect, indicating that screen position not only affects clickthroughs, but they also significantly affect conversions. Consumers are more likely to book a hotel that is positioned on an earlier page in the search results and at the top of a webpage. In particular, a one-position increase in rank corresponds to a 5.63% increase in conversions on an average. Similarly, we found a positive coefficient on the quadratic term of

[9] In this paper, we follow prior literature (e.g., Ghose and Yang 2009, Yang and Ghose 2010, Agarwal et al. 2011) with regard to the likelihood function.

[10] We found similar trends in the interaction effects in the conversion-rate model as well, which we briefly discuss in the next subsection.

Table 2 Main Results from Model Estimation

Panel (a): Coefficient estimates from clickthrough model

	Mean	*Class*	*Price*$^{(L)}$	*Rating*
Intercept	1.049 (0.054)*	0.040 (0.011)*	—	—
Rank	−0.062 (0.007)*	−0.026 (0.004)*	−0.019 (0.004)*	0.020 (0.003)*
Rank^2	0.004 (0.000)*	—	—	—
Page	−0.035 (0.004)*	−0.007 (0.001)*	−0.011 (0.005)*	0.016 (0.002)*
Price$^{(L)}$	−0.141 (0.021)*	0.002 (0.000)*	—	0.004 (0.000)*
Rating	0.078 (0.015)*	0.001 (0.002)	—	—
ReviewCnt$^{(L)}$	0.033 (0.009)*	0.029 (0.032)	−0.002 (0.023)	0.017 (0.003)*
$H^{(L)}$ (total no. of hotels)	−0.007 (0.000)*	—	—	—
Brand		Yes		
SpecialSort$^{(L)}$		Yes		

Unobserved heterogeneity estimates (covariance matrix Σ^β)

	Intercept	*Rank*	*Page*	*Price*	*Rating*	*ReviewCnt*$^{(L)}$
Intercept	1.012 (0.041)*	—	—	—	—	—
Rank	−0.029 (0.003)*	0.118 (0.045)*	—	—	—	—
Page	0.016 (0.001)*	−0.025 (0.002)*	0.102 (0.032)*	—	—	—
Price	−0.156 (0.029)*	−0.020 (0.008)*	0.031 (0.101)	1.443 (0.058)*	—	—
Rating	0.025 (0.006)*	−0.051 (0.206)	−0.042 (0.067)	−0.039 (0.012)*	0.067 (0.003)*	—
ReviewCnt$^{(L)}$	0.003 (0.000)*	−0.109 (0.099)	0.037 (0.008)*	0.060 (0.297)	−0.116 (0.004)*	0.217 (0.040)*

Panel (b): Coefficient estimates from conversion model

	Mean	*Class*	*Price*$^{(L)}$	*Rating*
Intercept	1.087 (0.166)*	0.057 (0.011)*	—	—
Rank	−0.021 (0.003)*	−0.009 (0.002)*	−0.010 (0.001)*	0.015 (0.005)*
Rank^2	0.002 (0.000)*	—	—	—
Page	−0.029 (0.004)*	−0.008 (0.001)*	−0.006 (0.002)*	0.003 (0.002)
Price$^{(L)}$	−0.156 (0.047)*	0.014 (0.011)*	—	0.009 (0.001)*
Rating	0.037 (0.001)*	0.002 (0.003)	−0.007 (0.016)	—
ReviewCnt$^{(L)}$	0.019 (0.001)*	0.013 (0.028)	−0.005 (0.017)	0.012 (0.001)*
$H^{(L)}$ (total no. of hotels)	−0.008 (0.001)*	—	—	—
Brand		Yes		
SpecialSort$^{(L)}$		Yes		

Unobserved heterogeneity estimates (covariance matrix Σ^γ)

	Intercept	*Rank*	*Page*	*Price*	*Rating*	*ReviewCnt*$^{(L)}$
Intercept	1.225 (0.032)*	—	—	—	—	—
Rank	−0.041 (0.012)*	0.089 (0.022)*	—	—	—	—
Page	0.038 (0.007)*	−0.070 (0.031)*	0.216 (0.088)*	—	—	—
Price	−0.203 (0.056)*	0.104 (0.051)*	0.044 (0.093)	2.005 (0.262)*	—	—
Rating	−0.159 (0.234)	0.137 (0.419)	0.028 (0.036)	0.077 (0.032)*	0.108 (0.024)*	—
ReviewCnt$^{(L)}$	0.015 (0.003)*	−0.089 (0.106)	0.020 (0.001)*	0.111 (0.183)	0.165 (0.052)*	0.304 (0.086)*

rank, suggesting the negative effect of rank order on conversion rate also increases at a decreasing rate.

As expected, *Price* has a negative effect on hotel demand, whereas *Class* has a positive effect on hotel demand. The online WOM-related variables, *Rating* and *ReviewCnt*, have a statistically significant and positive effect on hotel demand. We also found similar trends in the interaction effects between *Ranking* and *Price/Class/Rating*, suggesting higher-class hotels and more expensive hotels are more sensitive to the online ranking effect. And hotels that receive lower ratings from users benefit more when placed on the top of the screen. The total number of hotels in a certain market, H, has a negative effect on hotel-level conversion rate. Intuitively, the higher the number of choices there are available to consumers, lower the probability of buying from any given hotel. Thus, on average, the conversion rate for each hotel decreases.

4.3. Ranking Model

The coefficient estimates from the ranking model are presented in panel (c) of Table 2. This third model

Table 2 (Continued)

Panel (c): Coefficient estimates from ranking model

	Mean	*Class*	*Price*$^{(L)}$	*Rating*
Intercept	1.487 (0.059)*	−0.017 (0.002)*	—	—
CR_{t-1}	−0.121 (0.014)*	−0.005 (0.010)	−0.004 (0.001)*	0.017 (0.022)
Price$^{(L)}$	0.114 (0.023)*	0.002 (0.003)	—	−0.012 (0.001)*
Rating	−0.019 (0.000)*	0.019 (0.027)	—	—
ReviewCnt$^{(L)}$	−0.017 (0.000)*	−0.003 (0.000)*	−0.006 (0.002)*	−0.002 (0.000)*
$H^{(L)}$ (total no. of hotels)	0.010 (0.001)*	—	—	—
Brand		Yes		
SpecialSort$^{(L)}$		Yes		

Unobserved heterogeneity estimates (covariance matrix Σ^{ω})

	Intercept	CR_{t-1}	*Price*	*Rating*	*ReviewCnt*$^{(L)}$
Intercept	2.246 (0.117)*	—	—	—	—
CR_{t-1}	−0.107 (0.033)*	0.282 (0.057)*	—	—	—
Price	0.114 (0.012)*	−0.095 (0.040)*	0.332 (0.056)*	—	—
Rating	−0.201 (0.023)*	0.037 (0.013)*	−0.002 (0.027)	0.838 (0.126)*	—
ReviewCnt$^{(L)}$	−0.032 (0.002)*	−0.043 (0.155)	0.054 (0.118)	−0.069 (0.033)*	0.078 (0.023)*

Panel (d): Coefficient estimates from rating model

	Mean	*Class*	*Price*$^{(L)}$
Intercept	2.198 (0.056)*	0.035 (0.008)*	—
Rank	−0.028 (0.007)*	0.001 (0.005)	0.003 (0.002)
Rank^2	0.004 (0.001)*	—	—
Page	−0.007 (0.000)*	−0.002 (0.000)*	−0.004 (0.000)*
Price$^{(L)}$	0.005 (0.001)*	0.001 (0.003)	—
ReviewCnt$^{(L)}$	0.003 (0.000)*	0.006 (0.011)	0.017 (0.015)
$H^{(L)}$ (total no. of hotels)	0.004 (0.000)*	—	—
Brand		Yes	
SpecialSort$^{(L)}$		Yes	

Unobserved heterogeneity estimates (covariance matrix Σ^{ρ})

	Intercept	*Rank*	*Page*	*Price*	*ReviewCnt*$^{(L)}$
Intercept	4.123 (0.287)*	—	—	—	—
Rank	0.195 (0.046)*	0.086 (0.030)*	—	—	—
Page	0.086 (0.025)*	0.127 (0.053)*	0.326 (0.068)*	—	—
Price	−0.211 (0.078)*	0.061 (0.080)	−0.155 (0.189)	2.017 (0.235)*	—
ReviewCnt$^{(L)}$	0.001 (0.003)	−0.098 (0.105)	0.072 (0.034)*	−0.209 (0.276)	0.174 (0.060)*

Panel (e): Covariance across clickthrough, conversion, rank and rating Ω_{jt}

	Clickthrough	*Conversion*	*Rank*	*Rating*
Clickthrough	2.721 (0.087)*	—	—	—
Conversion	2.006 (0.043)*	0.773 (0.060)*	—	—
Rank	−0.214 (0.022)*	−0.626 (0.051)*	0.521 (0.060)*	—
Rating	0.835 (0.067)*	0.304 (0.038)*	−0.409 (0.079)*	0.339 (0.036)*

Note. SpecialSort is a vector of six control variables indicating the frequency of use of different sorting criteria.
$^{(L)}$The natural logarithm form of the variable.
$^*p < 5\%$.

sheds light on how search engines' ranking decisions are related to different product inherent characteristics, social media influences, and certain performance metrics such as previous conversions. Not surprisingly, we found that *Price* has a positive sign and *Class* has a negative sign. All else equal, a hotel with a higher price is more likely to appear in a better screen position. A higher-class hotel is also more likely to appear in a higher screen position, after controlling for the sorting criteria. Both *Rating* and *ReviewCnt* have a significant and negative effect, showing that hotels with a higher user rating and with more reviews are more likely to appear at the top of a page, controlling for everything else.

Table 3 Model Fit Comparison Results

	Main model (quadratic rank term, full heterogeneity)	Model with quadratic rank term, partial heterogeneity	Model with linear rank term, full heterogeneity	Model with linear rank term, partial heterogeneity	Model with ordered probit for rank
			In-sample model prediction (clickthrough rate)		
RMSE	0.0665	0.0759	0.0732	0.0968	0.1020
MSE	0.0044	0.0058	0.0054	0.0094	0.0104
MAD	0.0102	0.0165	0.0152	0.0282	0.0345
			Out-of-sample model prediction (clickthrough rate)		
RMSE	0.0939	0.1068	0.1134	0.1247	0.1601
MSE	0.0088	0.0114	0.0129	0.0156	0.0256
MAD	0.0361	0.0427	0.0464	0.0505	0.0963
			In-sample model prediction (conversion rate)		
RMSE	0.0816	0.0996	0.0925	0.1127	0.1445
MSE	0.0067	0.0099	0.0086	0.0127	0.0209
MAD	0.0183	0.0237	0.0208	0.0389	0.0490
			Out-of-sample model prediction (conversion rate)		
RMSE	0.1164	0.1218	0.1292	0.1573	0.1867
MSE	0.0135	0.0149	0.0167	0.0247	0.0349
MAD	0.0386	0.0523	0.0479	0.0688	0.1102

Note. RMSE, root mean square error; MSE, mean square error; MAD, mean absolute deviation.

4.4. Rating Model

Finally, the coefficient estimates from the rating model are shown in panel (d) of Table 2. The rating model allows us to account for the potential endogenous nature of the customer ratings. We found that both *Rank* and *Page* have a negative and statistically significant effect, suggesting screen position is also correlated with a hotel's rating. Hotels with higher ratings are more likely to be positioned on an earlier page in the search results and at the top of a webpage. We also found a similar positive effect from the quadratic term of rank, which suggests that the marginal effect of ranking on rating is decreasing.

Note that, in the main model, we assume consumer evaluation (e.g., rating of a hotel, utility of clicking, or booking a hotel) is a quadratic function of the rank order. As a robustness check, we also tried using a simple linear form. We excluded the quadratic term of the rank order from the clickthrough, conversion, and rating models. The qualitative nature of the estimation results stays consistent. The corresponding estimation results are shown in Online Appendix B. We also conducted model fit comparisons between the different alternative models. We found the main model provides a better performance in both in- and out-of-sample predictions. The model fit comparison results are provided in Table 3.

4.5. Policy Experiment: Effect of Ranking on Revenue

Previous work has shown that a consumer-utility-based search engine ranking system can lead to an increase in consumer surplus (Ghose et al. 2012). However, how such a ranking system affects the search engine's revenues is unclear. Therefore, one question in which we are interested is how different ranking mechanisms would affect search engine revenues.

Toward this goal, we conduct a set of policy experiments. In particular, we consider and compare six different ranking designs based on consumer utility, conversion rate (CR), clickthrough rate (CTR), price, customer rating, and the Travelocity default algorithm.[11] We define the ranking equation in the simultaneous equation model as being based on each of these six ranking criteria to reflect different search engine ranking systems. For the consumer-utility-based ranking, we define the ranking equation based on Equation (8). For the other five ranking designs, we define the ranking equation to contain only the corresponding variable on the right-hand side. For example, in the case of the price-based ranking mechanism, we define the ranking equation to contain the price variable as the independent variable. All other control variables remain the same in each of the six scenarios.

We estimate the simultaneous equation model under each different ranking equation using data from the previous $t-1$ periods. Based on the estimates, we predict the CTR and CR correspondingly for the tth period under each case. This process allows us to predict the future revenue for the search engine

[11] The default ranking algorithm used by Travelocity at the time of our data collection was based on a fixed commission rate (10%) of the last period's revenue. Therefore, in the policy experiment, we use the last period's revenue as the ranking criterion to approximate the Travelocity default ranking.

Table 4 Policy Experiment Results for Search Engine Revenue Prediction

Ranking mechanism	Predicted revenues from top-1 ranked hotel ($)	Predicted overall revenues from all hotels ($)
Utility	1,846	423,401
CR	1,866	415,678
Travelocity default	2,210	402,349
Rating	1,739	367,662
Price	2,003	361,096
CTR	1,476	312,757

under various ranking mechanisms. The overall revenue for the search engine is as follows:

$$Revenue = \sum_{j=1}^{J} (CR_j * CTR_j * Price_j). \qquad (15)$$

From our prediction results, we find that although the Travelocity default ranking and price-based ranking mechanisms lead to higher search engine revenue received from the top-ranked hotel, the consumer-utility-based ranking mechanism leads to the highest overall revenue received from all hotels. This finding suggests that a utility-based ranking mechanism not only maximizes the surplus for consumers (Ghose et al. 2012) but also maximizes the revenue for search engines.

The main reason for this finding is likely due to the diversity provided in the utility-based ranking. Consistent with the previous results by Ghose et al. (2012), consumers prefer the diversity in the ranking results. More importantly, we find that under the utility-based ranking mechanism consumers are more likely to click and purchase products that are ranked lower in the list, compared with all the other competing ranking mechanisms. This finding seems to explain why the utility-based ranking outperforms the others (especially the price-based or short-term revenue-based mechanisms) in the overall search engine revenue—the additional conversions received from the lower-ranked products are able to dominate the overall compromise in price. We provide the detailed prediction results in Table 4.

5. Randomized Experimental Design

Our Bayesian analysis provides important insights into the relationship between search engine ranking mechanism and consumer behavior. However, to fully understand how consumers make decisions in the product search engine context, we designed and conducted randomized experiments. Specifically, we tested the effectiveness of four ranking mechanisms and two personalization designs—active (customizable) personalized ranking and passive (noncustomizable) personalized ranking—on influencing consumer behavior and search engine revenues.

In a randomized experiment, a study sample is divided into two groups: one receiving the intervention being studied (the treatment group) and the other not receiving it (the control group).[12] Randomized experiments have major advantages over observational studies in making causal inferences. Randomization of subjects to different treatment conditions ensures the treatment groups are, on average, identical with respect to all possible characteristics of the subjects, regardless of whether those characteristics can be measured. In our first experiment, we designed four treatment groups. Each group is exposed to the same search-ranking mechanism except for a different default ranking method. In the second experiment, we have two treatment groups and one control group. The control group is granted full access to the search mechanism with active personalization that allows them to interact with and customize the search engine recommendation algorithm. By contrast, the two key personalization features are disabled for the two treatment groups (which we refer to as passive personalization). Our experimental participants come from Amazon Mechanical Turk (AMT, https://www.mturk.com), which is an online marketplace used for crowdsourcing microtasks that require human intervention (i.e., cannot be fully automated using machine learning tools).[13] We discuss the experimental procedure in §§5.1–5.5.

5.1. Hotel Search Engine Design

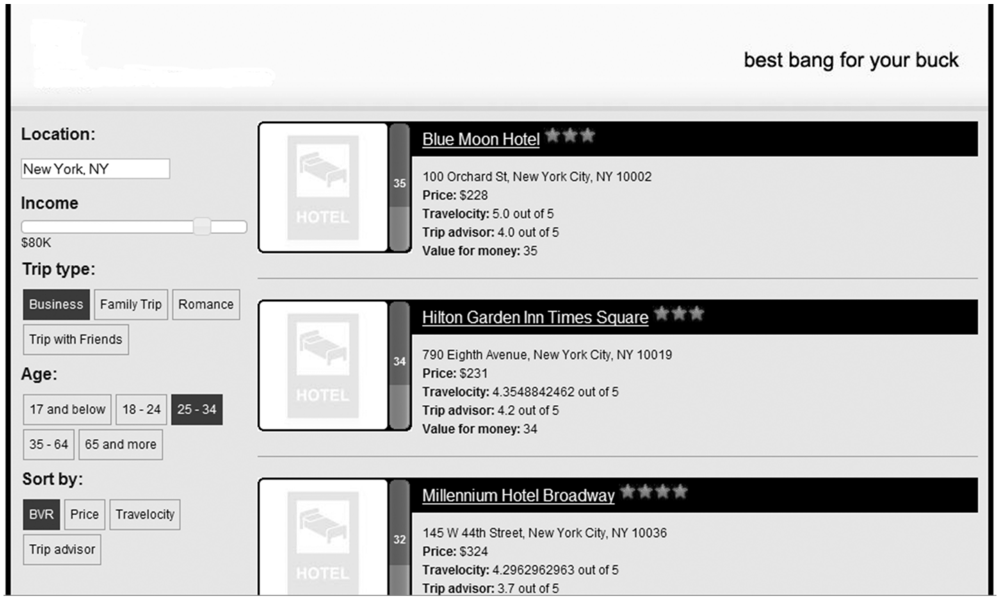
First we designed and built a real-world hotel search engine. This application served as the main instrument for our experimental studies. The main interface of this search engine consists of three components: (1) search criteria, including travel destination and search context (e.g., demographics such as income, trip type, and age); (2) sorting methods; and (3) resulting hotel list, on the right-hand side as the response to (1) and (2). A screenshot of the main search interface is provided in Figure 2.

When consumers start to search for hotels, they are able to define the travel destination, income level, trip type, and age group. We classify consumer trip type into four major categories: *business trip*, *family trip*, *romantic trip*, and *trip with friends*. We classify consumer age into five groups: 17 and below, 18–24, 25–34, 35–64, and 65 and older. Meanwhile, we provide consumers with four different sorting methods: *BVR*, *price*, *TripAdvisor.com customer rating*, and

[12] In some cases, rather than be compared with the control group, multiple treatment groups can be compared with each other (Ranjith 2005). We use this method in our first experimental study.

[13] Based on a pilot study, we found that the AMT population is generally representative of the overall U.S. Internet population. We provide more details of the pilot study in Online Appendix E.

Figure 2 Screenshot of the Main Search Interface of the Hotel Search Engine



Travelocity.com customer rating. BVR denotes the best-value ranking adapted from the utility-based ranking in Ghose et al. (2012). The value-for-money score represents how much additional value consumers can obtain from a hotel after paying the nightly reservation rate. We use the acronym BVR on the search engine to minimize the potential experimenter-expectancy bias that can accrue from displaying the full, expanded label. For each hotel listed on the right-hand side, we provide the summarized hotel information, including the hotel class (i.e., in pink stars), address, price, customer ratings from both Travelocity.com and TripAdvisor.com, and the value for the money (i.e., both in text and indicated by a vertical pink bar).

Users view the summary information in the hotel list and decide whether they want to click on a hotel's URL to acquire more detailed information. If a user chooses to click on a hotel's URL, he or she is directed to that hotel's landing page. A sample hotel landing page is provided in Figure 3. Generally speaking, the landing page consists of three components: (1) search criteria, similar to those on the main search page, where consumers can refine the travel destination and search context; (2) value-for-the-money scores, including the hotel's overall value for the money and the breakdown value score for each hotel feature (e.g., price, location, and service and customer reviews); and (3) consumer decision: a "buy now with one-click" button that allows consumers to make a simulated purchase or a "back" button that takes consumers back to the main search-result page to continue searching.

Note that the value-for-the-money score on the landing page exists in two forms: the population's *average* value score and the *personalized* value score. The former represents how much value a hotel feature provides to the overall population, whereas the latter represents the personalized value to a specific consumer based on the search context and demographics. Moreover, each hotel feature is associated with a "weight" that ranges from -1 to $+1$, representing consumer preference from "strongly dislike" to "strongly favor." A consumer can adjust the weight of his or her preference for each hotel feature to obtain a personalized value that most closely represents his or her preference. Overall, by choosing different search criteria or/and weights of preferences, a consumer is able to personalize the ranking results provided by the search engine.

5.2. Consumer-Behavior-Tracking System

To better understand the complete decision-making process, we keep track of the exact searching and purchasing behavior of users. This tracking system records the detailed information of every online activity by every consumer. For example, such activity information includes click behavior (e.g., a hotel URL being clicked, corresponding rank position, time spent on the landing page), usage of the search functions (e.g., search criteria changed, sorting methods chosen), hotel landing page browsing behavior (e.g., preference weights adjusted, search criteria changed), and purchase behavior (e.g., corresponding hotel being booked, corresponding ranking position, sorting method). Furthermore, each activity is recorded with a time stamp capturing when the activity occurs.

Figure 3 **Screenshot of a Sample Hotel Landing Page**

Note. There are a total of 25 hotel features on the landing page. For brevity, we only list seven features here: price, beach, downtown, hotel class, internal amenities, online rating, and review count.

5.3. Experiment I: Evaluating the Impact of the Ranking Mechanism

We now discuss the design of our first randomized experiment, which aims to examine consumer behavior and search engine revenue under different ranking mechanisms. The basic procedure is as follows. We ask the subjects to visit our hotel search engine website, conduct a hotel search using a set of randomly assigned search criteria, and make a simulated purchase at the end. The independent variable is the default ranking method. We are interested in how the ranking mechanism affects the breadth, depth, concentration, and final decision of consumer search. Moreover, we are interested in the resulting revenue for the search engine. Therefore, the dependent variables we focus on are (i) number of clicks, (ii) time spent on evaluation, (iii) number of online activities, (iv) number of conversions (0 or 1), and (v) search engine revenues.

We use a mixed experimental design. First, for the between-subjects design, we use a completely randomized setting with four treatment conditions. We manipulate the independent variable by changing the default ranking method for each of the four treatment groups. Each treatment group is exposed to a different default ranking method. We then randomly assign each subject to only one of the four groups. Meanwhile, to control for the error variance associated with individual subject-level differences, we propose a within-subjects design considering hotel search in two major U.S. cities: New York City and Los Angeles. We allow each subject to participate in two experiments corresponding to the two cities but only in the same treatment group. We summarize the design of this study in Table 5(a).

5.4. Experiment II: Evaluating the Impact of Personalization

In our second study, we examine consumers' responses to different personalized ranking mechanisms. In particular, we focus on two independent variables that capture two different levels of personalized ranking: (1) whether it allows consumers to change their personalized search context and (2) whether it allows consumers to adjust their weights or preferences for different hotel features.

Table 5(a) Experimental Design—Study I

	Within subject	
	New York City	Los Angeles
Between subject		
Treatment group 1	BVR	BVR
Treatment group 2	Price	Price
Treatment group 3	TripAdvisor rating	TripAdvisor rating
Treatment group 4	Travelocity rating	Travelocity rating

Table 5(b) Experimental Design—Study II

	Within subject	
	New York City	Los Angeles
Between subject		
Control group	Full access	Full access
Treatment group 1	No search context	No search context
Treatment group 2	No weight	No weight

The dependent variables we look into are the CTR and CR at both the subject and group levels. Moreover, we are also interested in the resulting search engine revenue. As before, we propose a mixed experimental design. For the between-subjects design, we apply a completely randomized setting with two treatment groups and one control group. We define the control group as subjects who have full access to our search engine website. For the two treatment groups, everything else is the same as in the control group, except that we remove the two personalization features—the user's ability to change the search context and to adjust weights of preferences—one at a time. Meanwhile, we control for the subject-level fixed effect by using a within-subjects design, similar to that in the first study. We summarize the design of the second study in Table 5(b).

5.5. Implementation

We have 900 unique user responses in the experiments, with 100 for each experimental group. We recruit users from the AMT platform. To control for quality, we allow only those AMT workers with a prior approval rate higher than 95% to participate in the experiments. AMT provides an approval rate for each worker based on the frequency with which buyers have approved tasks. This approval rate can provide information on the quality of the workers. Moreover, we design an additional survey at the end of the experiment asking the subjects to provide (1) a verification ID that is automatically generated once the experiment is properly finished and (2) a short explanation of why they made their final decision, using at least 20 characters. This two-step process helps us avoid negligent participants who have not gone through the entire experiment seriously. With regard to the experimental procedure, we first provide a short introduction about the experiment, as shown in Figure 4. To familiarize subjects with how to use the hotel-search website, we provide a quick two-page demo of the website prior to the experiment. Figure 5 shows the final introduction page leading to the start of the experiment.

Figure 4 Screenshot of the Introduction Page (1)

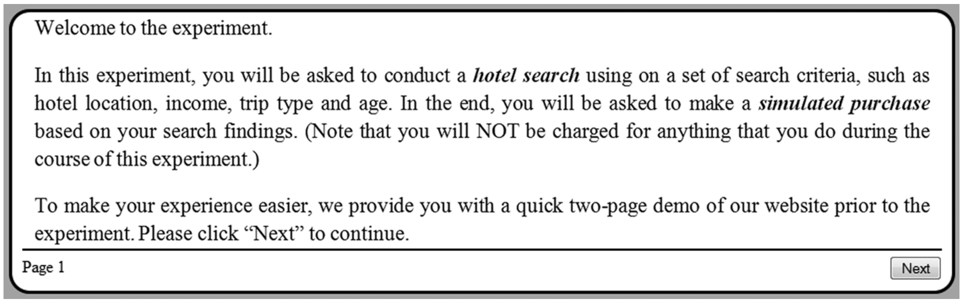
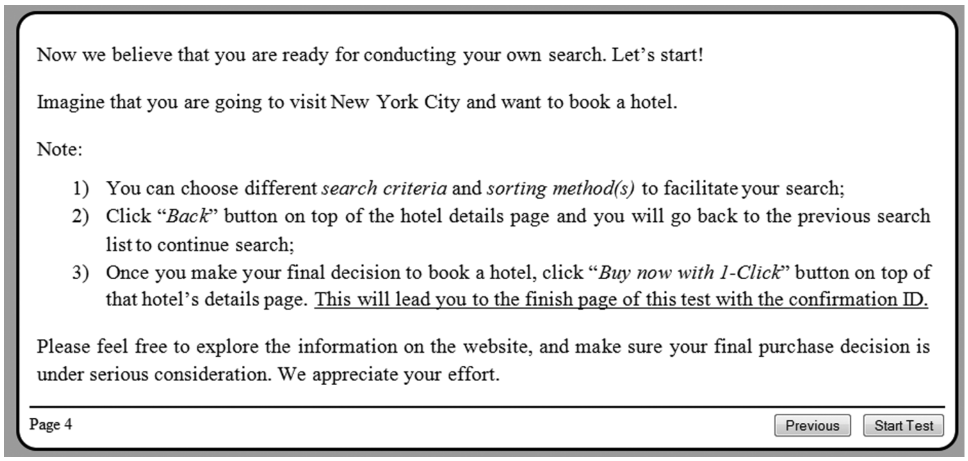


Figure 5 Screenshot of the Introduction Page (2)



6. Results from Randomized Experiments

6.1. Direct Ranking Effect

6.1.1. Ranking Effect on Click and Purchase Propensities. First we look into how the design of ranking mechanisms affects different aspects of user behavior on search engines. We examine the total time spent, number of online activities and number of clicks at the subject level, and the overall purchase propensity[14] from each of the four treatment groups in Study I. Table 6 shows the final purchase propensities under different ranking mechanisms. Subjects who get to see BVR as the default ranking pay more attention and display higher purchase propensities than subjects from other groups. This result is significant at the $p = 0.05$ level based on a post hoc ANOVA test. Price-based ranking provides the second-best performance on these two dimensions, followed by the rankings based on TripAdvisor and Travelocity ratings, respectively. Moreover, this finding is consistent across the two cities, New York City and Los Angeles. This result shows how the design of ranking mechanisms affects the performance of a product search engine.

We also find a significant ranking effect at the individual hotel level. Hotels ranked at the top of the search result list received, on average, 2.39 times more clicks compared with the second-ranked hotels and 3.42 times more compared with the third-ranked hotels. This trend stays consistent across the two cities and regardless of the default ranking method. Table 7 shows the number of clicks received for hotels ranked in the top 10.

We also examine CTR for the same hotel that appeared in different ranking positions under different default ranking mechanisms. Controlling for everything else, the same hotel in a higher screen

Table 6 Experiment Results—Average User Behavior Under Different Ranking Mechanisms

	Purchase propensity (NYC)	Purchase propensity (LA)
BVR (utility)	0.88	0.93
Price	0.65	0.69
TripAdvisor rating	0.54	0.44
Travelocity rating	0.47	0.41

Notes. Group mean over all users. Significant ($p < 0.05$), post hoc ANOVA.

[14] The purchase propensity is defined as the number of subjects who have made a purchase, divided by the total number of subjects in each group.

Table 7 Experiment Results—Number of Clicks Received at Top-10 Ranking Positions

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9	Rank 10
BVR										
NYC	56	24	13	10	9	11	8	2	1	1
LA	68	20	14	11	10	7	5	4	1	2
Price										
NYC	25	10	9	9	7	5	2	1	0	0
LA	34	15	10	8	6	4	3	2	0	1
TripAdvisor										
NYC	31	12	8	8	5	4	4	0	1	1
LA	23	15	10	9	4	2	0	1	0	1
Travelocity										
NYC	23	11	9	6	7	4	4	1	0	2
LA	17	9	8	8	5	3	2	2	0	0

position received significantly more clicks. For example, the Blue Moon Hotel in New York City received a total of 56 clicks under the BVR ranking, in which it was ranked at position 1. However, the same hotel received zero clicks under the price-based ranking, in which it was ranked 31.

6.1.2. Ranking Effect on Search Engine Revenue. Recall we are interested in how different ranking systems affect overall search engine revenues. We compute the overall search engine revenues by multiplying the unit price by the number of conversions for each hotel and then summing over all hotels in the experiments. We provide the detailed results in Table 8.

Our experimental results are highly consistent with the policy experiment results from the previous archival data analysis (i.e., §4.5). We find that price-based ranking leads to the highest search engine revenue received from the top-ranked hotel. However, BVR (consumer-utility-based) ranking leads to the highest overall revenue from all the hotels. Moreover, we find experimental evidence that under the BVR ranking, a significant part of the overall revenue comes from hotels that are ranked lower on the computer screen, which is different from the other competing ranking mechanisms.

These experimental findings support our previous policy experiment. They indicate consumers prefer the diversity in the utility-based ranking. Diversity presented in the ranking list can lead to a significant increase in conversions, especially from the lower-ranked products. Moreover, these additional conversions can contribute significantly to the overall revenue for search engines.

Table 8 Experiment Results—Search Engine Revenue Under Different Ranking Mechanisms

	Revenues from top-1 ranked hotel ($)	Overall revenues from all hotels ($)
BVR (utility)	2,052	7,162
Price	2,876	6,898
TripAdvisor rating	1,738	4,350
Travelocity rating	1,486	4,002

Notes. Revenue summed over two cities (NYC and LA). Significant ($p < 0.05$), post hoc ANOVA.

6.2. Interaction Effect Between Ranking and Product Rating

6.2.1. Interaction Effect between Ranking and Hotel Class Rating. We examine the differences in CTR from different ranking positions for two different classes of hotels—luxury- and budget-class hotels. In particular, we look into the changes in CTR at different ranking positions for either 4- or 5-star hotels (i.e., luxury hotels) and for 3-star or lower hotels (i.e., budget hotels). We find that as one moves down from the top-ranked position to a lower-ranked position, the decrease in CTR for luxury hotels is much larger than that for budget hotels. For example, moving down from the top to the fifth position leads to a 75% drop in CTR for the luxury hotels compared with a 54% drop for the budget ones. We test different ranking positions using a robustness check and find the results to be consistent. Table 9(a) shows the changes in CTR of hotels when moving down from the top position to the third, fifth, and 10th position.

6.2.2. Interaction Effect between Ranking and Customer Rating. Similarly, we also examine the differences in CTR from different ranking positions for hotels with higher customer ratings compared with those with lower customer ratings. In particular, we compare CTR at different ranking positions for 4- to 5-star hotels, as rated by reviewers, versus 1- to 2-star hotels. We find the increase in CTR resulting from hotels moving from a lower- to a higher-ranked position is greater for hotels with a poor reputation than for hotels with good reputation. For example, moving up from the 10th-ranked position to the top position increases CTR by 245% for hotels with low user ratings compared with an increase of 83% for hotels with

Table 9(a) Experiment Results—Interaction Effect Between Ranking Positions and Hotel Class Ratings

Rank	Luxury (4-, 5-star) (%)	Budget (1-, 2-, 3-star) (%)
1 → 3	−69	−43
1 → 5	−75	−54
1 → 10	−99	−80

Note. Results are based on average CTR.

Table 9(b) Experiment Results—Interaction Effect Between Ranking Positions and Hotel Customer Ratings

Rank	Good (4-, 5-star) (%)	Poor (1-, 2-star) (%)
10 → 5	11	45
10 → 3	49	166
10 → 1	83	245

Note. Results are based on average CTR.

high user ratings. Table 9(b) shows the corresponding changes in the CTR of hotels moving up from the 10th position to the fifth, third, and top position.

Findings in Tables 9(a) and 9(b) provide important insights and additional support to the archival data analysis, indicating luxury hotels are more sensitive to the ranking effect and are more adversely affected by an inferior screen position. Meanwhile, hotels that receive a lower reputation from online WOM are benefiting more when placed at the top of the search results. Our findings strongly illustrate the need for product search engines to directly incorporate signals from online social media into the ranking algorithms.

6.3. Effect of Active versus Passive Personalized Ranking

6.3.1. Effect of Personalized Ranking on Click and Purchase Propensities. Another important goal of our research is to examine how different personalized ranking mechanisms influence the way consumers behave on product search engines. In Study II, we consider three levels of personalization: active personalized ranking with full access (control group, henceforth "FULL_ACCESS"), passive personalized ranking without search context (treatment group 1, henceforth "NO_SEARCH"), and passive personalized ranking without weights of individual preferences (treatment group 2, henceforth "NO_WEIGHT"). Table 10 summarizes the average user behavior, in terms of total time spent and total number of activities, under the three different personalization mechanisms.

We find the active personalized ranking mechanism results in more user time and more activities than the two passive mechanisms. Each user, on average, spends approximately 351 seconds and conducts 19 activities per session when exposed

Table 10 Experiment Results—Average User Time and Activities Under Different Personalized Ranking Mechanisms

	Time spent (seconds)	Total no. of activities
Active personalized ranking with full access	351.23	19.36
Passive personalized ranking with no search context or demographics[a]	228.52	16.78
Passive personalized ranking with no weights of individual preferences[b]	127.01	8.24

Notes. Group mean over all users, across two cities (NYC and LA). Significant ($p < 0.05$), post hoc ANOVA.
[a]This passive personalized ranking only allows users to personalize their weights of individual preferences.
[b]This passive personalized ranking only allows users to personalize their search contexts and demographics.

to active personalized ranking. This finding suggests that an active personalized ranking can generate higher online engagement on the search engine. The NO_WEIGHT group with passive personalized ranking demonstrates the lowest level of user engagement. This step provides a sanity check that these different personalization features indeed influence user behavior in our experiments.

Table 11 displays the average number of clicks made by a user and the overall purchase propensity for the two different cities under the three personalized ranking mechanisms. Interestingly, we find that a travel search engine with an active personalized ranking mechanism can attract significantly more clicks than those with passive mechanisms. However, active personalized ranking leads to a significantly lower purchase propensity. This finding is consistent across the two different cities and is interesting because one would expect the active personalized ranking mechanism to increase, rather than decrease, the purchase propensities. One possible explanation is related to consumer expectations. In most online

Table 11 Experiment Results—User Behavior and Search Engine Revenues Under Different Personalized Ranking Mechanisms

	No. of clicks (NYC)	No. of clicks (LA)	Purchase propensity (NYC)	Purchase propensity (LA)	Overall revenues ($)
Active personalized ranking with full access	2.17	2.36	0.51	0.55	5,103
Passive personalized ranking with no search context or demographics	1.38	1.40	0.77	0.83	6,631
Passive personalized ranking with no weights of individual preferences	1.62	1.67	0.72	0.73	6,254

Notes. Group mean over all users. Significant ($p < 0.05$), post hoc ANOVA.

shopping environments, consumers find active personalization especially useful because it helps them discover what they want to buy before they know it themselves. In other words, the active personalized ranking is more likely to increase sales when consumers have not planned their purchase beforehand. In our setting, we focus on the type of consumers who have planned their purchase before the search starts. Under such a scenario, the major advantage of active personalized ranking is lost on consumers because they already have in mind what they are searching for. What is worse, if the personalization results do not meet consumers' expectations, they may easily stop the sale. This finding is in line with previous findings by Lambrecht and Tucker (2013), who show the mismatch between the specificity of the ad content and whether a consumer has well-defined preferences can lead to ineffective personalization. Another plausible explanation is related to consumers' cognitive limitations. The ability to extensively search and change their current consideration sets under the active personalized ranking mechanism can lead to information overload during the decision-making process. As a consequence, consumers may end up being confused or frustrated and therefore skip buying completely.

Comparing the NO_SEARCH group with the FULL_ACCESS group, the additional personalization based on search context and demographics (i.e., "search-based" personalization) results in a larger negative effect on purchase propensity (i.e., 6% larger for LA and 3% larger for NYC) than when we compare the NO_WEIGHT group with the FULL_ACCESS group. This finding provides a plausible explanation: two types of personal information can apparently be used in the personalization process in our context—(i) user-identity-related (i.e., who are you?) and (ii) user-preferences-related (i.e., what do you like?). Search context and demographic information lie closer to the former category, whereas weights of location and service preferences belong to the latter. Our results suggest that when designing a personalized ranking mechanism, using the identity-related information is less beneficial, not only for privacy-preserving purposes, but also for the economic outcomes such as conversions.

The findings above are directly observed at the search engine level. To verify the effects of active and passive personalized ranking mechanisms, we conduct two further analyses at the individual-subject level.

First, we consider the user-level number of clicks as the dependent variable in our analysis. The independent variables we are interested in are two dummies: *NOSEARCH* and *NOWEIGHT*, corresponding to the two passive personalized ranking treatment groups,

Table 12 Experiment Results—Negative Binomial Model on Number of Clicks

	Coeff.	Coeff.	Coeff.
NOSEARCH	−0.891* (0.362)	−0.889* (0.371)	−0.773* (0.242)
NOWEIGHT	−0.577* (0.230)	−0.569* (0.238)	−0.494* (0.201)
City	No	Yes	Yes
Activities	No	No	Yes
Log pseudolikelihood	−176.56322	−176.54825	−155.10346

*$p < 5\%$.

respectively. Because the number of clicks is a non-negative integer, we use a count data model, the negative binomial model with robust error. For estimation, we apply the maximum likelihood method. To control for the location effect, we include a city dummy variable denoting whether it is New York City (NYC) or Los Angeles (LA). Moreover, from the previous analysis, we notice the number of consumer activities drops significantly in the case of *NOWEIGHT*. Therefore, to control for the level of online attention, we include the number of total activities at subject level as an additional control variable. The results are qualitatively consistent as displayed in columns 2–4 in Table 12. Both *NOSEARCH* and *NOWEIGHT* show a significant and negative effect on the number of clicks, which means the presence of personalization in search context and weights of preferences has significant positive effects on the clicks at the individual level. The ability to define their search criteria on specific contexts and adjust their preferences toward product features leads to more clicks.

Second, we consider the user-level purchase propensity as the dependent variable in our analysis. As before, we are interested in two independent variables: *NOSEARCH* and *NOWEIGHT*. Note that in our experiment, we ask each subject to make a purchase at the end. However, subjects can still decide not to do so. Thus, the purchase outcome is a binary variable: 0 or 1. Therefore, we apply the probit model with maximum likelihood method for estimation. Again, we include two additional control variables: city dummy and number of total activities. We display the results in columns 2–4 in Table 13. Both *NOSEARCH* and *NOWEIGHT* have a statistically significant positive

Table 13 Experiment Results—Probit Model on Purchase Propensity

	Coeff.	Coeff.	Coeff.
NOSEARCH	0.587* (0.233)	0.581* (0.228)	0.591* (0.219)
NOWEIGHT	0.076 (0.096)	0.080 (0.089)	0.167* (0.093)
City	No	Yes	Yes
Activities	No	No	Yes
Log pseudolikelihood	−341.00704	−340.88529	−318.09032

*$p < 5\%$.

sign. This finding suggests that the presence of personalization in search context and individual preferences has significant negative effects on the purchase propensity at the individual level. This result is highly consistent with our previous analysis at the search engine level. It indicates the active personalized ranking mechanism can lead to a significant decrease in consumer purchase propensity.

6.3.2. Effect of Personalized Ranking on Search Engine Revenue. Finally, we are interested in how active and passive personalized ranking mechanisms affect the revenue for search engines. Consistent with the previous definition, we sum over all hotels in the experiment to compute the overall search engine revenue. We find the active personalized ranking mechanism can lead to significantly lower overall revenues than the two passive mechanisms in our travel search engine. This finding provides further insight that the decrease in purchases because of the improper use of the active personalized ranking strategy can result in a decrease in the overall revenue for product search engines. Thus, implementing the active personalized ranking mechanism may not always be profitable for product search engines. We provide the corresponding results in the last column in Table 11.

6.4. Robustness Tests

To further test the validity of our results, we conduct two robustness tests by considering two additional situations. First, we consider a setting with an even higher level of active personalization. Consumers who are randomly assigned to this setting are granted full access to active personalized ranking, as in the previous setting. Moreover, they can adjust their individual weights of preferences not only on the hotel landing page but also on the main search page. The value score for each hotel and the corresponding BVR ranking will be adjusted instantly based on the weight preferences consumers choose on the search page. The search interface for this robustness test is shown in Figure 6.

We found a similar trend when comparing the case of active personalized ranking with passive personalized ranking. In the new setting, users tend to spend even more time (i.e., an average of 343.02 seconds)

Figure 6 Screenshot of the Main Search Interface (Robustness Test)

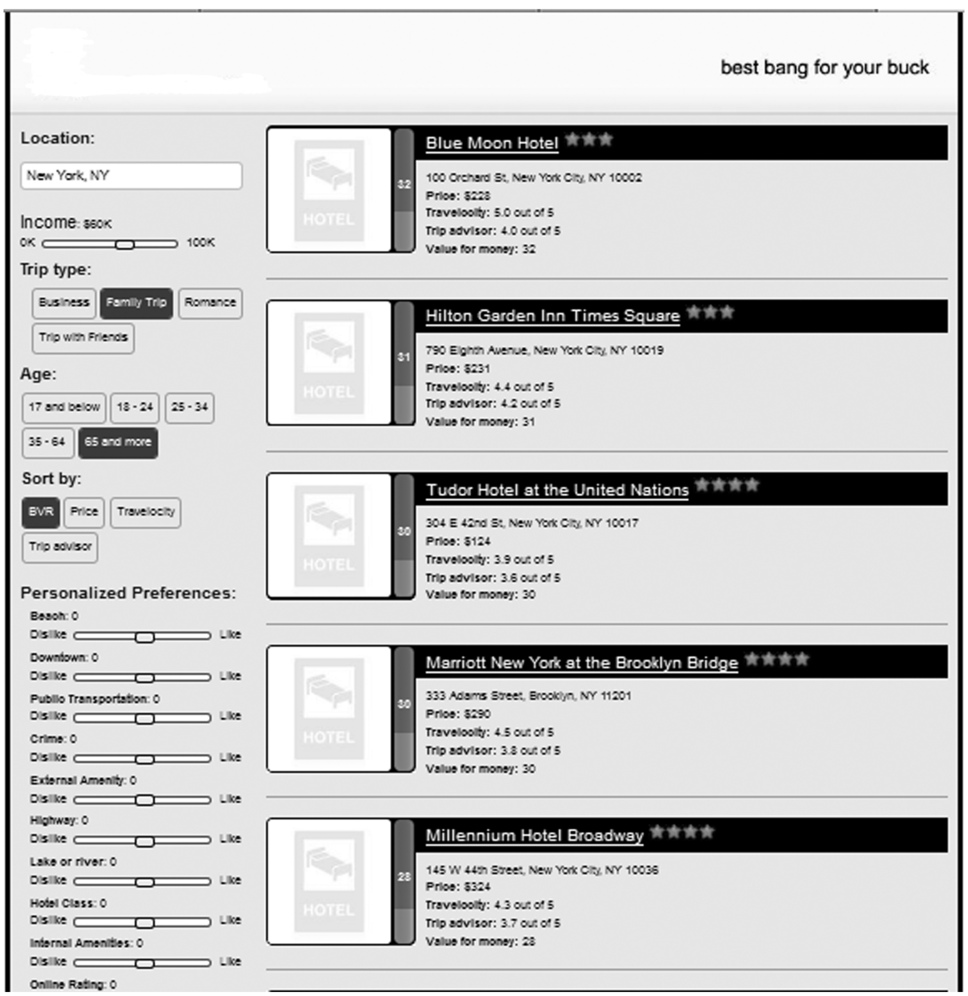


Table 14 Experiment Results—Robustness Test (1)

	Time spent (seconds)	Total no. of activities	No. of clicks (NYC)	No. of clicks (LA)	Purchase propensity (NYC)	Purchase propensity (LA)	Overall revenues ($)
High-level active personalized ranking with full access	343.02	19.27	2.28	2.42	0.45	0.44	4,622
Passive personalized ranking with no search context or demographics	228.52	16.78	1.38	1.40	0.77	0.83	6,631
Passive personalized ranking with no weights of individual preferences	127.01	8.24	1.62	1.67	0.72	0.73	6,254

and conduct even more activities (i.e., an average of 19.27 activities) on the search engine than in the two passive personalized ranking scenarios. These two statisticsSP again serve as good manipulation checks, indicating users are indeed using the personalization features. Furthermore, the high-level active personalization leads to a significantly lower purchase propensity and lower search engine revenue compared with the two passive mechanisms. This result strongly supports our previous findings obtained from both the archival data analysis and the experiment regarding whether excess information discourages consumers from making final decisions. Improper use of the active personalized ranking mechanism can lead to a loss of profit for product search engines. The detailed results are shown in Table 14.

Second, to test consumers' behavior when they have a less structured purchase plan in mind, we consider a more general purchase situation in which, rather than having to make a planned purchase at the end of each search session, consumers can choose to leave the search session without making a purchase. For comparison, consumers who are randomly assigned to this setting receive full access to the active personalized ranking recommendation.

We found that in the case of active personalization with an "unplanned purchase," the average time users spend on the site drops to nearly half of that in the case of active personalization with a "planned purchase" (i.e., 177.01 versus 351.23 seconds). However, the average number of activities in which users engage in the two cases remains similar (i.e., 18.18 versus 19.36 activities). Furthermore, in the case of active personalization with an "unplanned purchase," purchase propensities increase compared with the case of a "planned purchase." The results are consistent across the two cities.

This finding suggests that active personalized ranking may be more effective when consumers generally do not have a well-structured purchase plan. In such cases, they are more likely to discover potentially relevant products. However, this scenario is not the case when consumers already have a clear purchase plan.

Table 15 Experiment Results—Robustness Test (2)

	Time spent (seconds)	Total no. of activities	Purchase propensity (NYC)	Purchase propensity (LA)
Active personalized ranking with a planned purchase	351.23	19.36	0.51	0.55
Active personalized ranking with an unplanned purchase	177.01	18.18	0.75	0.69

Consumers can be highly discouraged and terminate the search completely if the active personalized ranking results mismatch their original expectations. This test provides additional insights into our main findings, suggesting that active personalized ranking should not be adopted blindly, and the level of personalization should be carefully designed based on the search context. The detailed results are provided in Table 15.

7. Conclusions and Implications

In this paper, we focus on investigating three major issues that product search engines are increasingly facing: the direct effect of ranking mechanism on consumer behavior and search engine revenue; the interaction effect of ranking and product ratings; and what kind of personalized ranking mechanism, if any, to adopt. Toward these objectives, we combine archival data analysis with randomized experiments based on a hotel search engine application that we designed. By manipulating the default ranking method and enabling or disabling a variety of active personalization features on the hotel search engine website, we are able to analyze consumer behavior and search engine revenue under different scenarios.

Our experimental results on ranking are consistent with those from the Bayesian model-based archival data analysis, suggesting a significant and causal effect of search engine ranking on consumer click and purchase behavior. In addition to a significant surplus gain found by a previous study (Ghose et al. 2012), a consumer-utility-based ranking mechanism yields the highest purchase propensity and the highest search engine overall revenue compared with existing benchmark systems, such as ranking based on price or star ratings. Moreover, an inferior screen position tends to more adversely affect luxury hotels and more expensive hotels. Hotels with lower reputations benefit more from being placed at the top of the search results. This finding illustrates the need for product search engines to directly incorporate signals from online social media into the ranking algorithms. We are beginning to see much of this interplay between search and social media happening in information search engines. Google began to incorporate tweets and other social media status updates into its real-time search function and then decided to create its own version of the Facebook "Like" button—the Google +1—and have it show up in search results. In another example of the interplay between social media and search, Microsoft's search engine Bing is now incorporating Facebook updates in its results.

Our experimental results on personalized ranking show the availability of excess personalization capabilities during the decision-making process may discourage consumers from searching, evaluating, and making final choices. In particular, we find that although active personalized ranking, compared with passive personalized ranking, can attract more online attention from consumers, it leads to a lower purchase propensity and lower search engine revenue. This finding suggests that personalized ranking should not be adopted blindly and the level of personalization should be carefully designed based on the search context. Our research sheds light on how consumers search, evaluate choices, and make purchase decisions in response to differences in product search engine designs. We provide empirical and experimental evidence for future studies to build on when designing an efficient ranking system and dynamically modeling consumer behavior on product shopping sites. A good ranking mechanism can reduce consumers' search costs, improve clickthrough rates and conversion rates of products, and improve revenue for search engines.

Our work has some limitations, some of which we are striving to address in our ongoing work. First, although the AMT platform provides an efficient and cost-friendly framework for randomized experimental design, the inherent heterogeneity in the Internet population makes controlling for subject characteristics across different treatment groups difficult. The randomization process can alleviate this concern to a large extent. However, robustness tests based on offline subjects as well would be helpful. Our current experiments focus on the type of consumers who can make, at most, one purchase in each online shopping session. To better understand the counterintuitive finding that an active personalized ranking mechanism leads to lower conversion rates, one can extend our experimental design to make a comparison with consumers who are allowed to make *multiple* purchases in a given session. In addition, a study of how the content of consumer search, such as the length and type of search keyword, interacts with the ranking effect would be interesting. Moreover, with regard to examining the ranking mechanism, one can expand the research scope by taking into account consumers' social network neighbors' search and purchase behavior. This expansion would allow one to test the impact of social-signal-based ranking mechanisms on product search engines. Notwithstanding these limitations, we believe our paper paves the way for future research in this exciting area at the intersection of social media and search engines.

Acknowledgments
For helpful comments, the authors thank Jason Chan, Sam Hui, Prasanna Tambe, Duncan Simester, Yong Tan, and Russ Winer, as well as seminar participants at Carnegie Mellon University, Cheung Kong Graduate School of Business, Columbia University, Harvard Business School, the Massachusetts Institute of Technology, University of Texas at Austin, University of Connecticut, University of Texas at Dallas Marketing Conference, the Wharton School, University of California, Los Angeles, and New York University. Anindya Ghose and Panagiotis G. Ipeirotis acknowledge funding from the Google Faculty Research Award in 2012.

References
Abhishek V, Hosanagar K, Fader P (2013) Aggregation bias in sponsored search data: The curse and the cure. Working paper, Carnegie Mellon University, Pittsburgh.

Agarwal A, Hosanagar K, Smith M (2011) Location, location, location: An analysis of profitability of position in online advertising markets. *J. Marketing Res.* 48(6):1057–1073.

Animesh A, Viswanathan S, Agarwal R (2011) Competing "creatively" in sponsored search markets: The effect of rank, differentiation strategy, and competition on performance. *Inform. Systems Res.* 22(1):153–169.

Ansari A, Mela CF (2003) E-customization. *J. Marketing Res.* 40(2):131–145.

Aral S, Walker D (2011) Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Sci.* 57(9):1623–1639.

Arora N, Henderson T (2007) Embedded premium promotion: Why it works and how to make it more effective. *Marketing Sci.* 26(4):514–531.

Baye MR, De los Santos B, Wildenbeest MR (2012) What's in a name? Measuring prominence, and its impact on organic traffic from search engines. Working paper, Indiana University, Bloomington.

Baye MR, Gatti JRJ, Kattuman P, Morgan J (2009) Clicks, discontinuities, and firm demand online. *J. Econom. Management Strategy* 18(4):935–975.

Chen Y, Yao S (2012) Search with refinement. Working paper, Northwestern University, Evanston, IL.

Chevalier J, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345–354.

Chib S, Greenberg E (1995) Understanding the Metropolis–Hastings algorithm. *Amer. Statistician* 49(4):327–335.

Chittor V (2010) Online retail: Getting the right product in front of your customers. *Inc.* (March 8), http://www.inc.com/internet/articles/201002/chittoor.html.

Dellarocas C (2012) Double marginalization in performance-based advertising: Implications and solutions. *Management Sci.* 58(6):1178–1195.

De los Santos B, Koulayev S (2013) Optimizing click-through in online rankings for partially anonymous consumers. Working paper, Indiana University, Bloomington.

Dodds WB, Monroe KB, Grewal D (1991) Effects of price, brand, and store information on buyers' product evaluations. *J. Marketing Res.* 28(3):307–319.

Dzyabura D (2014) Preference discovery in product search: When novice consumers don't know what they don't know. Working paper, New York University, New York.

Ghose A, Yang S (2009) An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Sci.* 55(10):1605–1622.

Ghose A, Goldfarb A, Han S (2013a) How is the mobile Internet different? Search costs and local activities. *Inform. Systems Res.* 24(3):613–631.

Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowd-sourced content. *Marketing Sci.* 31(3):493–520.

Ghose A, Ipeirotis PG, Li B (2013b) Surviving social media overload: Predicting consumer footprints on product search engines. Working paper, Carnegie Mellon University, Pittsburgh.

Goldfarb A, Tucker C (2011) Online display advertising: Targeting and obtrusiveness. *Marketing Sci.* 30(3):389–404.

Greene W (1999) *Econometric Analysis*, 4th ed. (Prentice Hall, Upper Saddle River, NJ).

Hausman JA (1975) An instrumental variable approach to full information estimators for linear and certain nonlinear econometric models. *Econometrica* 43(4):727–738.

Jerath K, Ma L, Park Y, Srinivasan K (2011) A "position paradox" in sponsored search auctions. *Marketing Sci.* 30(4):612–627.

Jeziorski P, Segal I (2012) What makes them click: Empirical analysis of consumer demand for search advertising. Working paper, University of California, Berkeley, Berkeley.

Koulayev S (2014) Estimating demand in online search markets, with application to hotel bookings. *RAND J. Econom.* Forthcoming.

Kuksov D, Villas-Boas JM (2010) When more alternatives lead to less choice. *Marketing Sci.* 29(3):507–524.

Lahiri K, Schmidt P (1978) On the estimation of triangular structural systems. *Econometrica* 46(5):1217–1221.

Lambrecht A, Tucker C (2013) When does retargeting work? Information specificity in online advertising. *J. Marketing Res.* 50(5):561–576.

Li X, Hitt LM (2010) Price effects in online product reviews: An analytical model and empirical analysis. *MIS Quart.* 34(4): 809–831.

Narayanan S, Kalyanam K (2011) Measuring position effects in search advertising: A regression discontinuity approach. Working paper, Stanford University, Stanford, CA.

Nevo A (2001) Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69(2):307–342.

Ranjith G (2005) Interferon-α-induced depression: When a randomized trial is not a randomized controlled trial. *Psychother Psychosom* 74(6):387.

Rossi PE, Allenby GM (2003) Bayesian statistics and marketing. *Marketing Sci.* 22(3):304–328.

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4):321–340.

Rutz OJ, Trusov M (2011) Zooming in on paid search ads—A consumer-level model calibrated on aggregated data. *Marketing Sci.* 30(5):789–800.

Rutz OJ, Bucklin RE, Sonnier GP (2012) A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *J. Marketing Res.* 49(3):306–319.

Yang S, Ghose A (2010) Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Sci.* 29(4):602–623.

Yao S, Mela CF (2011) A dynamic model of sponsored search advertising. *Marketing Sci.* 30(3):447–468.

Zhang J, Wedel M (2009) The effectiveness of customized promotions in online and offline stores. *J. Marketing Res.* 46(2):190–206.