Examining the Impact of Search Engine Ranking and Personalization on Consumer Behavior: Combining Bayesian Modeling with Randomized Field Experiments

Anindya Ghose [1,2], Panagiotis G. Ipeirotis [2], Beibei Li [2]

[1] Department of Operation and Information Management
Wharton School of Business, University of Pennsylvania

[2] Department of Information Systems
Stern School of Business, New York University

Abstract

In this paper, we examine how different ranking and personalization mechanisms on product search engines influence consumer online search and purchase behavior. To investigate these effects, we combine archival data analysis with randomized field experiments. Our archival data analysis is based on a unique dataset containing approximately 1 million online sessions from Travelocity over a 3-month period. Using a hierarchical Bayesian model, we first jointly estimate the relationship among consumer click and purchase behavior, and search engine ranking decisions. To evaluate the causal effect of search engine interface on user behavior, we conduct randomized field experiments. The field experiments are based on a real-world hotel search engine application designed and built by us. By manipulating the default ranking method of search results, and by enabling or disabling a variety of personalization features on the hotel search engine website, we are able to empirically identify the causal impact of search engines on consumers' online click and purchase behavior.

The archival data analysis and the randomized experiments are consistent in demonstrating that ranking has a significant effect on consumer click and purchase behavior. We find that hotels with a higher reputation for providing superior services are more adversely affected by an inferior screen position. In addition, a consumer utility-based ranking mechanism yields the highest click and purchase propensities in comparison to existing benchmark systems such as ranking based on price or customer ratings. Our randomized experiments on the impact of active vs. passive personalization mechanisms on user behavior indicate that although *active* personalization (wherein users can interact with the recommendation algorithm) can lead to a higher click-through rate compared to *passive* personalization, it leads to a lower conversion rate when consumers have a planned purchase beforehand. This finding suggests that active personalization strategies should not be adopted ubiquitously by product search engines. On a broader note, our inter-disciplinary approach provides a methodological framework for how econometric modeling, randomized field experiments, and IT-based artifacts can be integrated in the same study towards deriving causal relationships between variables of interest.

1. Introduction

Many businesses today have started looking at consumers' online search queries and click log data, to understand how consumers seek and evaluate relevant information during their online shopping forays. In fact, the knowledge created from customer interactions with product search engines allows firms to customize their business and services in an interactive way to gain and retain customers (Henzinger 2007, Gretzel et al. 2006). Consequently, product search engines have evolved into one of the most important strategic platforms for information seeking and marketing-communications. Moreover, because of the information overload reinforced by the recent explosion of social media (e.g., online word-of-mouth, social communities, geo-/social-tagging, photo/video sharing and blogs), product search engines perhaps provide the best way for consumer to seek information and act upon it.

Outside of search, one of the most important ways for shoppers to discover products has been through recommendation engines (Chittor 2010). Personalization and recommendation engines have been around for a while and have been a strong driver of sales. For example, Amazon's recommendation system was said to account for up to 35 percent of sales in 2006. However, while individual online retailers have increased their usage of recommendation systems, product search engines have still not made any headway into providing personalized results in response to consumer queries for products.

Over the last few years, a tremendous amount of research has focused on how to improve the content quality of the search results, for example, by optimizing retrieval of relevant documents from the Web, mainly as a response to a keyword query (e.g., Lavrenko and Croft 2001, Pang and Lee 2008). Nevertheless, due to the multi-dimensional preferences of consumers for many products and services, several questions remain unanswered in this space. How can product search engines present their results in a manner that facilitates efficient information exchange and effective marketing activities? Should product search engines allow consumers to interact with the recommendation algorithm to personalize their search results? Therefore, two challenges appear to be crucial for product search engines today. First, what *ranking mechanism* should be used to effectively present the search results? Second, what *personalization mechanism* should be applied to deliver the search results to the population of heterogeneous consumers? These are the goals of our research. More specifically, first, we aim to examine how differences in search engine ranking mechanisms affect consumer search and purchase behavior online. Second, we examine how different levels of personalization affect consumer behavior and search engine performance. In particular, we compare between two types of personalization mechanisms: active personalization and passive personalization. In our context, a ranking system that personalizes results based on the average utility from a given hotel and enables consumers to proactively interact with the recommendation algorithm prior to the display of results from a search query are classified as "active". In contrast, a ranking system that personalizes results based on the average utility from a given hotel, but does not allow customers to interact with the recommendation algorithm prior to displaying results is classified as passive.

Towards examining these questions, we combine Bayesian modeling on archival data analysis with randomized field experiments. Our research focuses on the hotel industry. We apply archival data analysis to gain insights towards our first research objective of studying the impact of ranking mechanisms on consumer click and purchase behavior. Using a panel data set from 2008/11 to 2009/1, containing approximately 1 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 inter-relationship between consumer click and purchase behavior, and search engine ranking decisions.

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 this is useful or not. Hence, there is no known archival data in any product search engine that has information on the effect of personalization on user behavior. Therefore, we design and conduct randomized field experiments based on a unique hotel search engine application designed and built by us. This also helps us make causal claims about the relationship between the search-based personalization strategies and consumers' purchase behavior.

In a randomized experiment, a study sample is divided into one group that will receive the intervention being studied (the treatment group) and another group that will not receive the intervention (the control group) [1]. Randomized experiments have major advantages over observational studies in making causal inferences. Randomization of subjects to different treatment conditions ensures that the treatment groups, on average, are identical with respect to all possible characteristics of the subjects, regardless of whether those characteristics can be measured or not. In our first experiment, we have 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 the search engine recommendation algorithm. In contrast, for the treatment groups, the two key personalization features are disabled for each group (which we refer to as passive personalization).

Our randomized experimental results are based on a total of 730 unique user responses over two-week period via Amazon Mechanical Turk (AMT) crowd-sourcing platform. We use a customized behavior tracking system to observe the detailed information of consumer search, evaluation and purchase decision making process. The use of randomized experimental design should allow a degree of certainty that the research findings cited in studies that employ this methodology reflect the effects of the interventions being measured and not some other underlying variable or variables. Hence, we need to be careful in designing these experiments. 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 extract the causal effect of search engine ranking and personalization on consumer behavior.

---

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

Our main findings are the following. First, we find a significant ranking effect on both click-throughs and conversions. A hotel that appears on a higher position on the screen and on an earlier webpage attracts a more clicks and conversions from consumers. On average, a one position increase on the screen is associated with a 7.31% increase in hotel click-throughs and a 4.56% increase in conversions. Moreover, we find that hotels with a higher reputation for providing superior services are more adversely affected by an inferior screen position (i.e., being ranked on the bottom part of the screen) than others.

Second, we find that the total number of hotels in a certain market has a negative effect on hotel click-throughs and conversions. This suggests that the more hotels available for a consumer to choose from, the less likely the consumer will choose any of them. A plausible explanation is related to theories of consumer cognitive cost. Prior theoretical work has shown that information overload and non-negligible search cost can discourage decision makers of searching, and end up with not searching or not choosing (Kuksov and Villas-Boas 2010). Our empirical finding nicely dovetails with the theoretical conclusion by Kuksov and Villas-Boas in that "more alternatives can lead to fewer choices."

Third, our experimental results on ranking mechanism are highly 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. Specifically, a consumer utility-based ranking mechanism yields the highest click and purchase propensities in comparison to existing benchmark systems such as ranking based on price or customer ratings.

Finally, we find active personalization mechanism that requires consumer interactions to specify both search context and individual preference can attract higher online attention from consumers and leads to higher click-through rate for search engine, compared to the two passive mechanisms where the two personalization choices are disabled one at a time. Surprisingly, search engine with active personalization mechanism performs the worst in the conversion rate. This finding suggests although active personalization helps consumers *discover* what they want to buy hence increasing the sale, it should not be adopted ubiquitously. When consumers already have a planned purchase in mind (as in our setting), active personalization may actually cause the conversion rate to drop.

2. Literature Review

Our research is related to the fields of search engine ranking and online position effect. Over the past few years, two opposite views have been held towards the position effect in product search. On one hand, consumers are endowed with cognitive limitation. Eye-tracking studies have long shown that people tend to scan the search results in order (e.g., Aula and Rodden 2009). Hence, the same link will have a higher click-through rate (CTR) if it is positioned towards the top of the page versus the bottom (e.g., Srikant et al 2010). Studies have also found empirical evidence suggesting significant effect of rank order in the context of search engine-based keyword advertising (e.g., Ghose and Yang 2009, Rutz and Bucklin 2007).

However, very little empirical work actually examines the rank order effect on product demand in searching and purchasing commercial products. A few existing studies such as Baye et al. (2009) examine the ranking effect on click-through rate as a substitute for the actual demand (conversions). Other studies tend to focus only on a single search dimension, for example, examining the competition of retailers ranked on *price* search engines (e.g., Ellison and Ellison 2009).

In contrast to the above theoretical work, consumers have been found to be "variety-seeking" in their economic choice making process (e.g., McAlister 1982, Givon 1984). Especially, recent studies have shown that when consumers search and shop for commercial products online, they tend to examine the variety reflected in the set of product search results as a whole for their choice decision (Agrawal et al 2009, Panigrahi and Gollapudi 2011). This is different from the traditional web search (i.e., which returns web pages) where people often examine the results in a top-down order. As a consequence, the rank order of the product search results (i.e., which contain normally commercial products) may not have significant effects as in the web page search context (Bhattacharya et al 2011), whereas only the *diversity* of products in the search results set matters. Therefore, one of our major goals in this research is to examine whether there exists a significant ranking effect in product search. By combining archival data analysis with a set of randomized experiments, our research can thus provide critical insights on the impact of search engine ranking and design on users' search and purchase behavior from a causal perspective.

Existing research also holds two different opinions toward the effects of personalization: supportive and skeptical (Arora et al. 2008). From the supportive perspective, Malthouse and Elsner (2006) show in a field test that personalizing the copy used in a book offer increases response rates significantly. Rossi et al. (1996) quantified the benefits of adopting one-to-one pricing by utilizing household purchase history data and empirically found that individual personalization improves 7.6% over mass optimization. Ansari and Mela (2003) found that targeting the content can potentially increase the expected number of click through by 62%. Arora and Henderson (2007) showed targeting at individual level can enhance the efficiency of embedded premium. From the skeptical perspective, Zhang and Wedel (2009) investigate the profit potential of various promotion programs customized at different levels in online and offline stores. They found that the incremental benefits of one-to-one promotions over segment- and market-level customized promotions were small in general, especially in offline stores.

Furthermore, one major concern in one-to-one marketing is invasion of privacy (Chellappa and Sin 2005, Arora et al. 2008). Chellappa and Sin (2005) developed a parsimonious model to predict consumers' usage of online personalization as a result of the tradeoff between their value for personalization and concern for privacy. They found that a consumer's intent to use personalization services is positively influenced by her trust in the vendor. A recent experimental study by Aral and Walker (2011) looked at application adoptions among 1.4 million friends of over 9,000 users on Facebook.com, and found that active-personalized invitations are less effective in generating peer influence and social contagion compared to passive-broadcast notifications. In summary, existing studies indicate although personalization can lead to customer satisfaction and profits, it may not work universally. Moreover, the level of personalization design is sensitive to the context and consumer behavior. Therefore, another goal of our research is to examine consumer online search and purchase behavior under different levels of personalization mechanisms on product search engines.

3. Data

Our dataset 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 over 3 months from 2008/11 to 2009/1. Besides, we supplement our search and transaction data with hotel service-, location- and customer review-based information collected using various machine learning techniques such as image classification and text mining tools. This provides us a final dataset with a total of 29,222 weekly observations for 2117 hotels in the US. More specifically, our dataset combines four major sources:

3.1. Consumer Search, Click and Conversion Data from Travelocity.com

We have complete information on consumer searching and shopping behavior. 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 there exists any, the login and actual transaction(s) if any conversion occurs, and the termination of the session.

We count a "display" for a hotel if that hotel appears visible to a consumer on the web page in an online search session. Meanwhile, a "click" is counted if the hotel is selected by a consumer, and a "conversion" is counted only if a consumer has finished the payment in that online session. Since our major goal is to exam the effect of rank order displayed on a page, we focus only on the sessions with at least one display[2]. A display often leads to a click, but it may not lead to an actual 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. Notice that when Travelocity displays the hotel search results on a web page, it

[2] In some cases, users may initiate a session and look for general travel information, for example the area of the city, rather than search for any hotels, thus there will be no hotels displayed on any web page. We excluded such sessions in our analysis.

only shows 25 hotels per page[3]. This 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. Besides, 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 control variable in our study to indicate how frequently a hotel appears in a result list under a "special sort."

In addition, we also have supplemental data collected from three other sources. We only briefly discuss them below.

3.2. Hotel Characteristics

Location Characteristics: We used geo-mapping search tools (Bing Maps API) and social geo-tags (from geonames.org) to identify the external amenities (e.g., shops, bars) and public transportation in the area around the hotel. We also used image classification together with Mechanical Turk to examine whether there is a nearby beach, a nearby lake, a downtown area, and whether the hotel is close to a highway. We extracted these characteristics within an area of 0.25-mile, 0.5 mile, 1-mile, and 2-mile radius.

Service Characteristics: This category contains hotel class, number of internal amenities and number of rooms. Hotel class is an internationally accepted standard ranging from 1-5 stars, representing low to high hotel grades. Number of internal amenities is the aggregation of hotel internal amenities, such as bed quality, hotel staff, food quality, bathroom amenities and parking facility. We extracted this information from the Tripadvisor website using fully automated parsing. Since hotel amenities are not directly listed on the Tripadvisor website, we retrieved them by following the link provided on the hotel web page, which randomly directs the user to one of its cooperating partner websites (e.g., Travelocity, Orbitz).

Review Characteristics: We collected customer reviews from Travelocity.com. The online reviews and reviewers' information were collected on a daily basis up to January 31, 2009 (the last

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

date of transactions in our database). In addition to the total number of reviews and the numeric reviewer rating, we extracted indicators that measure the stylistic characteristics of the reviews for robustness checks. We examined two text-style features: subjectivity and readability of reviews (Ghose and Ipeirotis 2011). Also, since prior research suggested that disclosure of identity information is associated with changes in subsequent online product sales (Forman et al 2008), we measured the percentage of reviewers for each hotel who reveal their name or location information on their profile pages.

Table 1. Definitions and Summary Statistics of Variables

Variable	Definition	Mean	Std. Dev.	Min	Max
	Search, Click and Conversion Data				
PRICE	Transaction price per room per night	120.45	73.25	25.77	978
DISPLAY	Number of displays	213.65	382.28	1	4849
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
	Hotel Location-Related Characteristics				
BEACH	Beachfront within 0.6 miles	.18	.38	0	1
LAKE	Lake or river within 0.6 miles	.22	.42	0	1
TRANS	Public transportation within 0.6 miles	.30	.46	0	1
HIGHWAY	Highway exits within 0.6 miles	.74	.44	0	1
DOWNTOWN	Downtown area within 0.6 miles	.67	.47	0	1
EXTAMENITY	Number of external amenities within 1 mile, i.e., restaurants, shopping malls, or bars	4.57	7.92	0	27
CRIME	City annual crime rate	193.19	126.70	3	1310
	Hotel Service-Related Characteristics				
CLASS	Hotel class	3.36	1.37	1	5
AMENITYCNT	Total number of hotel amenities	11.54	7.56	2	23
ROOMS	Total number of hotel rooms	212.30	250.70	12	2900
	Hotel Review-Related Characteristics				
REVIEWCNT	Total number of reviews	21.06	29.28	1	202
RATING	Overall reviewer rating	3.84	.85	1	5
	Control Variables				
SPECIALSORT	Number of times using a sorting method	204.64	377.26	0	4810
H	Total number of hotels in a city	24.03	56.48	1	922
BRAND	Dummies for 9 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**			

In summary, each observation in our dataset contains the hotel id, week id, number of displays, number of clicks, number of conversions, average screen position (i.e., rank on the result page), average page number, and the corresponding service-/location-/review-related characteristics for that hotel 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.

4. Hierarchical Bayesian 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 estimate the impacts from the search engine ranking mechanism on consumer search and purchase behavior.

Our model is motivated by the work in (Ghose and Yang 2009). The general idea is the following. We propose to build a simultaneous model of click-through, conversion, and rank. We model the click-through and conversion behavior as a function of hotel brand, price, rank, page, sorting criteria, and hotel characteristics (available from either the hotel search summary page or the hotel landing page, depending on the stage in a search process). The rank of a hotel is modeled as a function of hotel brand, price, sorting criteria, hotel characteristics that are available from the hotel landing page, and performance metrics like previous conversion rate. Each function contains an unobserved error that is assumed to be normally distributed with mean zero. To capture the unobserved co-variation among click-throughs, conversions, and rank, we assume that the three error terms are correlated and follow the multivariate normal distribution with mean zero. More specifically, our model can be described as follows.[4]

4.1. Model Setup

First, we define our unit of observation to be "hotel-week." Thus, for hotel *j* in week *t* assume that there are n_{jt} clicks-throughs among N_{jt} displays ($n_{jt} \leq N_{jt}$ and $N_{jt} > 0$). Meanwhile, assume that

[4] For robustness check, we also tried a count data model, the Poisson Model. The qualitative nature of our results stays consistent. Due to brevity, we do not describe it in this paper. The results are available upon request.

among the n_{jt} click-throughs, there exist m_{jt} conversions ($m_{jt} \le n_{jt}$). We further define the probability of having a click-through to be p_{jt} and the probability of having a conversion conditional on a click-through to be q_{jt}. A consumer's decision process involves two steps: In the first step, she sees a hotel displayed on the search result web page and decides whether to click it; in the second step, if she clicks on the hotel, she will decide whether 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 event is $1-p_{jt}$. (2) A consumer sees a hotel, clicks through, but does not purchase. The probability of such event is $p_{jt}(1-q_{jt})$. (3) A consumer sees a hotel, clicks through and makes a purchase. The probability of such event is $p_{jt}q_{jt}$.

Therefore, we can derive the likelihood function of observing the joint occurrence of n_{jt} click-throughs and m_{jt} conversions, (n_{jt}, m_{jt}), to be the following

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

4.2. A Simultaneous Equation Model of Click-Through, Conversion, and Rank

We model the click-through, conversion and rank simultaneously in a hierarchical Bayesian framework. In particular, we divide our model into three interactive components.

(1) Click-Through Rate Model

First, we model the probability that a consumer clicks on hotel *j* in week *t* to be a function of *rank order, page number, hotel price,* and *hotel characteristics* that are available from the search result *summary* page (i.e., *hotel class, customer rating,* and *customer review count*). In addition, to control for the size of the local market, we include the *total number of hotels* in *j*'s city, H_j, as a control variable. We also include hotel *brand* dummies to control for the unobserved hotel characteristics. Finally, to capture consumers' particular sorting preferences we include an additional control variable, *SpecialSort*$_{jt}$, representing the total number of times a special sorting algorithm is used by consumers during the search process for hotel *j* in week *t*. This gives us the following equation.

$$p_{jt} = \frac{\exp(U_{jt}^{p})}{1+\exp(U_{jt}^{p})}$$

$$\text{where,} \quad \begin{aligned} U_{jt}^{p} &= \beta_{j0} + \beta_{j1}\text{Rank}_{jt} + \beta_{2}\text{Page}_{jt} + \alpha_{1}\text{Price}_{jt} + \alpha_{2}\text{Class}_{j} + \alpha_{3}\text{Rating}_{jt} \\ &+ \alpha_{4}\text{ReviewCount}_{jt} + \alpha_{5}\text{H}_{j} + \alpha_{6}\text{Brand}_{j} + \alpha_{7}\text{SpecialSort}_{jt} + \varepsilon_{jt}. \end{aligned} \tag{2}$$

To capture the unobserved heterogeneity, we model β_{j0}, β_{j1} to be random coefficients:

$$\beta_{j0} = \overline{\beta_{0}} + \sigma_{j0}^{\beta} \quad \text{and} \tag{3}$$

$$\beta_{j1} = \overline{\beta_{1}} + \delta_{1}\text{Class}_{j} + \sigma_{j1}^{\beta}, \tag{4}$$

where we assume the intercept β_{j0} to vary along its population mean $\overline{\beta_{0}}$, and β_{j1} to vary along the population mean $\overline{\beta_{1}}$ and the hotel-specific characteristic (i.e., hotel class). Moreover, we model the two error terms in (3) and (4) to be correlated in the following way:

$$[\sigma_{j0}^{\beta}, \sigma_{j1}^{\beta}]' \sim MVN(0, \Sigma^{\beta}), \quad \text{where } \Sigma^{\beta} \text{ is a } 2\times 2 \text{ covariance matrix.} \tag{5}$$

(2) Conversion Rate Model

Second, we model the probability of a consumer's conversion as a function of *rank order*, *page number*, *hotel price,* and *hotel characteristics* that are available from the search result *landing* page (i.e., in addition to the ones used in the click-through model, all other hotel characteristics obtained from the detailed hotel descriptions, images, maps and online customer review information on the hotel landing web page). Similarly, we include the *total number of hotels, brand, and special sort* as control variables. The conversion equation is written as

$$q_{jt} = \frac{\exp(U_{jt}^{q})}{1+\exp(U_{jt}^{q})},$$

$$\text{where,} \quad \begin{aligned} U_{jt}^{q} &= \gamma_{j0} + \gamma_{j1}\text{Rank}_{jt} + \gamma_{2}\text{Page}_{jt} + \theta_{1}\text{Price}_{jt} + \theta_{2}\text{Class}_{j} + \theta_{3}\text{Rating}_{jt} \\ &+ \theta_{4}\text{ReviewCount}_{jt} + \theta_{5}\text{H}_{j} + \theta_{6}\text{Brand}_{j} + \theta_{7}\text{SpecialSort}_{jt} + \theta_{8}\text{HotelFeatures} + \eta_{jt}. \end{aligned} \tag{6}$$

Note that *HotelFeatures* here represents the set of hotel characteristics that are visible to consumers only after they click the hotel and go to the detailed hotel landing page, including total number of rooms, total number of hotel internal amenities (e.g., gym, pool), near a beach, near a lake/river, easy access to an interstate highway, easy access to public transportations, near the downtown area, total number of external amenities (e.g., restaurants, shops), local crime rate, etc. Thus, θ_8 is a vector of coefficients, with each coefficient corresponding to each of the above hotel features. Similarly, we model γ_{j0}, γ_{j1} as random coefficients with the following properties:

$$\gamma_{j0} = \overline{\gamma_0} + \sigma^{\gamma}_{j0} \text{ and} \tag{7}$$

$$\gamma_{j1} = \overline{\gamma_1} + \tau_1 \text{Class}_j + \sigma^{\gamma}_{j1}. \tag{8}$$

Moreover, we model the two error terms in (7) and (8) to be correlated in the following way:

$$[\sigma^{\gamma}_{j0}, \sigma^{\gamma}_{j1}]' \sim MVN(0, \Sigma^{\gamma}), \quad \text{where } \Sigma^{\gamma} \text{ is a } 2 \times 2 \text{ covariance matrix.} \tag{9}$$

(3) Ranking Model

Equations (2) ~ (9) model consumer's behavior of click-through and conversion. Meanwhile, we can model search engine's ranking decision. We model the rank order of hotel *j* in week *t* as being dependent on the set of hotel characteristics and the control variables used in the previous consumer behavior models. Besides, we include a performance metric, *the previous conversion rate,* $CV_{j,t-1}$. The model is written as follows:

$$\begin{aligned} \ln(Rank_{jt}) &= \omega_{j0} + \omega_1 \text{CV}_{j,t-1} + \omega_2 \text{Price}_{jt} + \omega_3 \text{Class}_j + \omega_4 \text{Rating}_{jt} + \omega_5 \text{ReviewCount}_{jt} \\ &+ \omega_6 \text{H}_j + \omega_7 \text{Brand}_j + \omega_8 \text{SpecialSort}_{jt} + \omega_9 \text{HotelFeatures} + \upsilon_{jt}. \end{aligned} \tag{10}$$

Similarly, we model the intercept ω_{j0} as random coefficient to vary along the population mean, and with a variance that follows normal distribution with mean zero:

$$\omega_{j0} = \overline{\omega_0} + \sigma^{\omega}_{j0}. \tag{11}$$

Finally, to capture the unobserved co-variation and the potential endogenous relationship among click-through, conversion and rank, we assume the three error terms in equations (2), (6) and (10) to be correlated as follows:

$$[\varepsilon_{jt}, \eta_{jt}, \upsilon_{jt}]' \sim MVN(0, \Omega_{jt}), \text{ where } \Omega_{jt} \text{ is a } 3 \times 3 \text{ covariance matrix.} \tag{12}$$

5. Empirical Analysis 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 50,000 iterations and used the last 20,000 iterations to compute the mean and standard deviation of the posterior distribution of the model parameters.

5.1. Click-Through Rate Model

We present the estimation results on the coefficients of the click-through model in Table 2. First of all, all coefficients are statistically significant at 5% level. The coefficients of both *Rank* and *Page* are negative and significant, indicating that position bias does exist. A hotel that appears on an earlier web page in the search results and from a higher position on the screen will receive significantly more clicks from consumers. One position higher on the screen will correspond to an average of 7.31% increase in click-throughs. Meanwhile, we estimated the effects from other hotel characteristics on click-throughs. Consistent with the theory and existing empirical findings (e.g., Baye et al. 2009), *Price* has a negative sign showing that the higher the price of a hotel, the lower the willingness of consumers to click on that hotel. *Class* presents a positive sign (i.e., .051), which suggests a positive relationship between hotel class and click-throughs. Besides, we found the interaction effect between *Rank* and *Class* is also negative and significant (i.e., -.014). This suggest that hotels with higher reputation for fancy services are more adversely affected by an inferior screen position (e.g., on the bottom part of the screen) than others.

Interestingly, we noticed that the total number of hotels in a certain market has a negative sign. This suggests that the more hotels available for a consumer to choose from, the less likely the consumer will click on any of them. An intuitive reason for this is related to the local competition among the alternatives. When there are more hotels in a market, the rivalry within the local market become more severe. Thus, on average the click-through rate for each hotel decreases. Another plausible explanation for this relates to the cognitive cost literature, and the idea that decision makers may only be able to process a limited amount of information (e.g., Simon 1955, Miller

1956, Shugan 1980, Gourville and Soman 2005). Information overload and non-negligible search cost can discourage decision makers of searching, and end up with not searching or not choosing (Kuksov and Villas-Boas 2010). Kuksov and Villas-Boas were able to theoretically show that indeed "more alternatives can lead to less choice." Our empirical finding in the hotel search context stands nicely in accordance with the theory and literature.

Table 2. Coefficient Estimates from Click-Through Model

		Intercept		*Class*
Intercept	$\overline{\beta_0}$	1.021(.086)*	α_2	.051(.010)*
Rank	$\overline{\beta_1}$	-.045(.007)*	δ_1	-.014(.002)*
Page	β_2	-.029(.001)*		--
Price$^{(L)}$	α_1	-.120(.018)*		--
Rating	α_3	.053(.011)*		--
ReviewCnt$^{(L)}$	α_4	.015(.000)*		--
$H^{(L)}$*(Total Number Of Hotels)*	α_5			--
Brand	α_6	Yes		--
SpecialSort$^{(L)}$	α_7	Yes		--

Unobserved Heterogeneity Estimates

	σ^{β}_{j0}	*Intercept*	σ^{β}_{j1}	*Rank*
σ^{β}_{j0} *Intercept*		.993(.076)*		--
σ^{β}_{j1} *Rank*		-.055(.005)*		.068(.012)*

$^{(L)}$: The natural logarithm form of the variable.

*: Significance level at $p < 5\%$.

5.2. Conversion Rate Model

The coefficient estimates from the conversion model are presented in Table 3. Recall that one major difference between the click-through model and the conversion model is that, the latter incorporates not only the hotel features that are directly available on the search result summary page (i.e., "perceived quality"), but also the features that are available on the hotel landing page (i.e., "actual quality"). A consumer's click decision depends only on the first set of hotel features, whereas a consumer's purchase decision depends on both sets of hotel features.

As we noticed from Table 3, most of the coefficients are statistical significant at 5% level. In particular, we found negative and significant effects from *Rank* and *Page*, indicating that the screen position not only affects the click-throughs, but also significantly affects the demand (i.e., conversions). A hotel that is positioned on an earlier web page in the search results and on top of a web page is more likely to be purchased by consumers. In particular, one position higher on the screen corresponds to an average of 4.56% increase in demand. Meanwhile, we found that *Price* has a significant negative effect on hotel demand, whereas *Class* has a significant positive effect on hotel demand. Two online word-of-mouth related variables, *Rating* and *Review Count* both present significant positive effects on hotel demand.

Table 3. Coefficient Estimates from Conversion Model

		Intercept		*Class*
Intercept	$\bar{\gamma}_0$	1.008(.179)*	θ_2	.045(.014)*
Rank	$\bar{\gamma}_1$	-.017(.002)*	τ_1	-.006(.001)*
Page	γ_2	-.022(.003)*		--
Price$^{(L)}$	θ_1	-.108(.042)*		--
Rating	θ_3	.030(.001)*		--
ReviewCnt$^{(L)}$	θ_4	.009(.002)*		--
$H^{(L)}$*(Total Number Of Hotels)*	θ_5	-.005(.001)*		--
Brand	θ_6	Yes		--
SpecialSort$^{(L)}$	θ_7	Yes		--
Rooms$^{(L)}$	θ_8	.002(.000)*		--
AmenityCnt$^{(L)}$		.015(.002)*		--
Beach		.077(.005)*		--
Lake		-.036(.021)		--
Trans		.082(.012)*		--
Downtown		.034(.006)*		--
Highway		.019(.002)*		--
ExtAmenity$^{(L)}$		.022(.007)*		--
Crime$^{(L)}$		-.004(.000)*		--
Unobserved Heterogeneity Estimates				
	σ^{γ}_{j0}	*Intercept*	σ^{γ}_{j1}	*Rank*
σ^{γ}_{j0} *Intercept*		1.227(.066)*		--
σ^{γ}_{j1} *Rank*		-.031(.005)*		.057(.009)*

$^{(L)}$: The natural logarithm form of the variable.
*: Significance level at $p < 5\%$.

5.3. Ranking Model

The coefficient estimates from the ranking model are presented in Table 4. This third model sheds lights on how search engines' ranking decisions are related to different product inherent characteristics, social media influences, as well as certain performance metrics like previous conversions.

Table 4. Coefficient Estimates from Ranking Model

		Intercept		ω_9 *Intercept*
Intercept	$\bar{\omega}_0$	1.231(.061)*	*Rooms*$^{(L)}$	-.005(.000)*
CV_{t-1}	ω_1	-.107(.018)*	*AmenityCnt*$^{(L)}$	-.004(.000)*
Price$^{(L)}$	ω_2	.088(.016)*	*Beach*	-.042(.012)*
Class	ω_3	-.015(.004)*	*Lake*	-.033(.006)*
Rating	ω_4	-.012(.000)*	*Trans*	-.037(.002)*
ReviewCnt$^{(L)}$	ω_5	-.011(.001)*	*Downtown*	-.065(.010)*
$H^{(L)}$*(Total Number Of Hotels)*	ω_6	.009(.003)*	*Highway*	-.028(.002)*
Brand	ω_7	Yes	*ExtAmenity*$^{(L)}$	-.005(.001)*
SpecialSort$^{(L)}$	ω_8	Yes	*Crime*$^{(L)}$	.011(.003)*
Unobserved Heterogeneity Estimates				
σ^{ω}_{j0} *Intercept*		.887(.104)*		

$^{(L)}$: The natural logarithm form of the variable.
*: Significance level at $p < 5\%$.

As we can see from Table 4, the majority of coefficient estimates are statistical significant at 5% level. Particularly, we noticed a significant and large negative effect from CV_{t-1}, which indicates that previous conversions have a dominate impact on search engines' ranking decisions. In particular, the higher the previous conversions of a hotel, the more likely it will be positioned on top of a page. Not surprisingly, we found *Price* has a positive sign and Class has a negative sign. This shows that all else being equal, a hotel with a higher price is more likely to appear at a lower screen position, while a hotel from a higher class is more likely to appear at a higher screen position, controlling for the sorting criteria. Besides, all hotel service- and location related features (except for *Crime*, which is the opposite) appear to be negatively associated with the rank order (i.e., positively associated with the screen position). This further indicates that quality is indeed the major factor that is considered by search engines in organic ranking decisions. For word-of-mouth features, both *Rating* and *Review Count* have a significant and negative effect, showing that hotels with higher rating and more reviews are more likely to appear on top of a page, controlling for everything else.

Table 5. Covariance Across Click-Through, Conversion and Rank Ω_{jt}

	ε_{jt} *Click-Through*	η_{jt} *Conversion*	υ_{jt} *Rank*
ε_{jt} *Click-Through*	2.004(.097)*	--	--
η_{jt} *Conversion*	1.652(.063)*	1.233(.072)*	--
υ_{jt} *Rank*	-.237(.005)*	-.715(.036)*	.951(.103)*

*: Significance level at p < 5%.

Notably, we found a statistical significance on the heterogeneity estimate in all three models, suggesting that individual-level hotel heterogeneity may affect the click-through, conversion and ranking prediction for each hotel. This heterogeneity is driven by unobserved factors beyond the observed hotel characteristics. Finally, recall that in order to model the unobserved co-variation and the potential endogeneity among click-through, conversion and rank, we assume the three error terms from the three models to be correlated. Table 5 shows the estimates for the covariance across the click-through, conversion, and rank. The unobserved covariance among the three factors is statistically significant, suggesting ranking can be endogenous. Ignoring the endogenous relationship will lead to biased estimates on the impact of ranking on click-throughs and conversions.

6. Randomized Field Experimental Design

Our Bayesian analysis alleviates concerns about unobserved heterogeneity and endogeneity by using a simultaneous equation modeling framework with random coefficients. It provides important insights on the *correlations* between the search engine ranking mechanism and consumer behavior. However, to fully understand how consumers make decisions under different search engine design features, it is important to identify a *causal* relationship. Unfortunately, evaluating the causal effects of search engine design features is difficult because search and purchase behavior are typically endogenous as shown by Ghose and Yang (2009) and Yang and Ghose (2010). We therefore designed and conducted two randomized field experiments testing the effectiveness of four search engine ranking mechanisms and two of the most widely used search engine personalization approaches – active personalization and passive personalization – in influencing the search and purchase behavior of consumers.

To examine the effect of search engine design and personalization activities on consumer behavior from a *causal* perspective, we supplement our archival data analysis with a series of randomized online experiments, using Google App Engine[5] and Amazon Mechanical Turk (AMT)[6]. Google App Engine is an online developing environment for building scalable web applications that run on Google's infrastructure. AMT is an online marketplace, used for crowd-sourcing on-demand micro-tasks that require human intervention (i.e., cannot be fully automated using machine learning tools)[7]. To study consumer behavior under different search mechanisms, we conduct two independent experimental studies to examine the ranking mechanism and personalization mechanism, respectively. We discuss the experimental procedure in subsections 6.1-6.5.

6.1 Hotel Search Engine Design

First, we design a unique hotel search engine using Google App Engine. This hotel search engine serves as the main instrument for us to study consumer online search and purchase behavior. 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; 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 1.

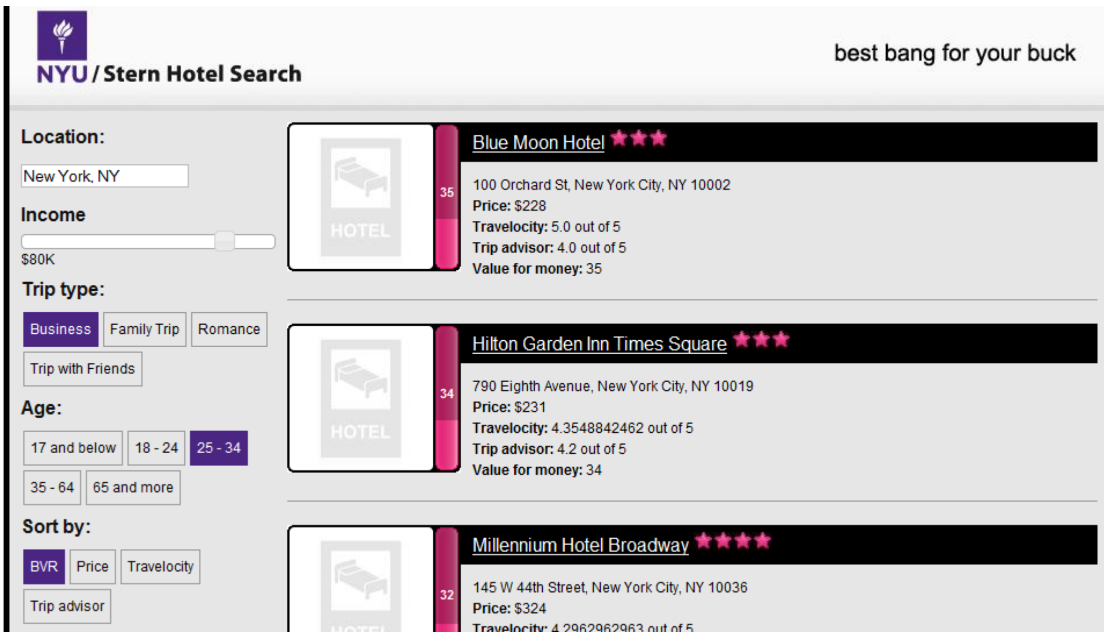
When consumers start to search for hotels, they are able to define the travel destination, income level, trip type, and age group. We classify consumer trip type into four major categories: *business trip*, *family trip*, *romantic trip*, and *trip with friends*. We classify consumer age into five groups: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

---

[5] https://appengine.google.com/start
[6] https://www.mturk.com/mturk/welcome
[7] Based on a pilot study, we found the AMT population is generally representative of the overall US Internet population.

minimize the potential experimenter-expectancy bias. 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 money (i.e., in both text and vertical pink bar). The value for money score represents how much additional value consumers can obtain from a hotel after paying for the price. It is rigorously derived based on the concept of consumer utility surplus using economic demand estimation[8].

Figure 1. Screenshot of the Main Search Interface of the Hotel Search Engine



Thus, consumers view the summary information in the hotel list and decide whether they like to click a hotel's URL to acquire more detailed information. If consumers choose to click a hotel's URL, they will be directed to the landing page of that hotel. A sample hotel landing page is provided in Figure 2. In particular, the lading page consists of three components: 1) Search Criteria: similar as that on the main search page, where consumers can refine the travel destination and search context; 2) Value for Money Scores: including overall value for money of the hotel and the breakdown value score for each individual hotel feature (e.g., price, location, service and customer review features); 3) Consumer Decision: "buy now with 1-click" button that allows consumers to

[8] More details are available in Ghose, Ipeirotis and Li (2011).

make a simulated purchase, or "back" button that takes consumers back to the main search result page to continue searching.

Note that for each value for money score on the landing page, it exists in two forms: the population *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 counterpart 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 "a strong dislike" to "a strong favor." A consumer can adjust the weight of her preference towards each hotel feature to obtain a personalized value further tuned for herself.

6.2. Consumer Behavior Tracking System

To understand the complete decision process of consumers, we need to keep in track of the exact behavior path how consumers search and make purchases. For this purpose, we design an online behavior tracking system that is specially tailored for our hotel search engine. This tracking system enables us to record the detailed information of every online activity by every consumer. For example, such activity information includes clicks (e.g., corresponding hotel URL being clicked, corresponding rank position and sorting method, time spent on the landing page, etc), search (e.g., search criteria changed, sorting methods chosen, etc), landing page browse (e.g., preference weights adjusted, search criteria changed, etc.), and purchase (e.g., corresponding hotel being purchased, corresponding ranking position and sorting method, etc.). Furthermore, each activity is recorded with a time stamp capturing when such activity occurs.

6.3. Field Experiment I: Evaluating the Impact of the Ranking Mechanism

In this subsection, we discuss the design of our first randomized experiment to examine how consumers behave in response to different search engine ranking mechanisms. The basic procedure is the following. 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.

Figure 2. Screenshot of a Sample Hotel Landing Page[9]

The independent variable is the default ranking method. We are interested in how ranking mechanism affects the breadth, depth, concentration, and final decision of consumer search.

[9] There are totally 25 hotel features on the landing page. For brevity, we only list 7 features here: price, beach, downtown, hotel class, internal amenities, online rating, and review count.

Therefore, the dependent variables we focus on are: (i) number of clicks, (ii) time spent on evaluation, (iii) number of online activities, and (iv) number of conversions (0 or 1).

We propose a mixed experimental design. First, regarding the between-subject design, we use a completely randomized setting with four treatment conditions. We manipulate the independent variable by changing the default ranking method for the four treatment groups, each with a different default ranking method. Each subject is then randomly assigned to only one of the four groups. Meanwhile, to control for the error variance associated with individual subject-level differences, we propose a within-subject design considering hotel search in two major US 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. The design of this study is summarized in Table 6.

Table 6. 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	Travelocity Rating	Travelocity Rating
	Treatment Group 4	TripAdvisor Rating	TripAdvisor Rating

6.4. Field Experiment II: Evaluating the Impact of Personalization

In this subsection, we discuss the design of our second study to examine consumer online behavior in response to the personalization mechanism. In particular, we focus on two independent variables that capture different levels of personalization design: (1) whether it allows consumers to change their personalized search context; (2) whether it allows consumers to adjust their own weights of preferences. Meanwhile, the dependent variables we look into are the click-through and conversion rates at both subject-level and group-level.

Similarly, we propose a mixed experimental design. For the between-subject design, we apply a completely randomized setting with two treatment groups and one control group. We define as control group subjects who have full access to our search engine website with BVR as the default ranking. Regarding the two treatment groups, we provide everything else the same, except that we manipulate the independent variables by removing two personalization features: user ability to change search context and user ability to adjust weights of preferences, one at a time. Meanwhile, we control for the subject-level fixed effect by using a within-subject design, similar as in the first study. The design of the second study is summarized in Table 7.

Table 7. Experimental Design – Study II

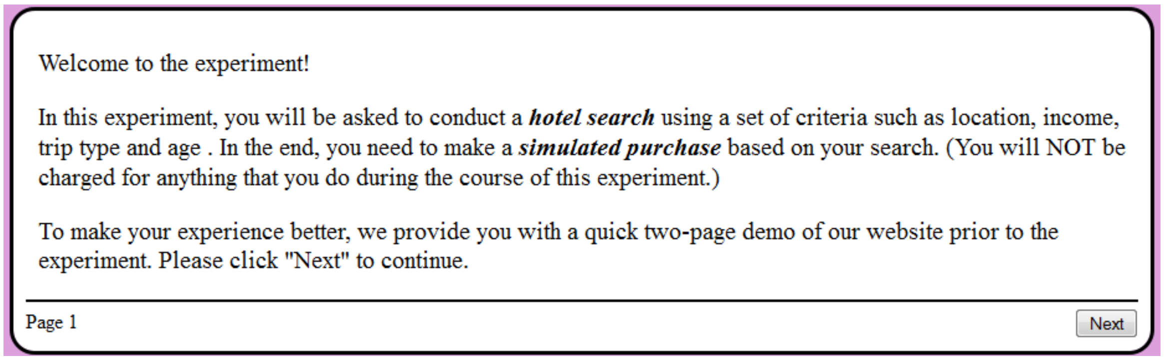
		(Within-Subject)	
		New York City	**Los Angeles**
(Between -Subject)	**Control Group**	Full Access	Full Access
	Treatment Group 1	No Search Context	No Search Context
	Treatment Group 2	No Weight	No Weight

6.5. Implementation

We simulate actual consumers by recruiting subjects from the AMT platform. To control for the quality of the responses, we allow only those AMT workers with a prior approval rate higher than 95% to participate in the experiment. AMT provides an approval rate for each worker based on the frequency with which tasks have been approved by the buyer. 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 reason why they made their final choice decision with at least 20 characters. This extra step helps us to prevent spammers who have not gone through the entire experiment seriously.

In regards to the experimental procedure, we first provide a short introduction about the experiment, as shown in Figure 3. To familiarize subjects how to use the hotel search website, we provide a quick two-page demo of the website prior to the experiment. To make sure the demographic distribution of the experiment subjects is consistent with that of the real world online consumers, we randomly assign each subject with a set of pre-defined search context and demographics, which we derive based on the real world traveler distribution. The assigned search context and demographic information is then used by the subjects as search criteria during the hotel search process. Figure 4 shows the final introduction page with a sample assignment of search criteria leading to the start of the experiment.

Figure 3. Screenshot of the Introduction Page (1)



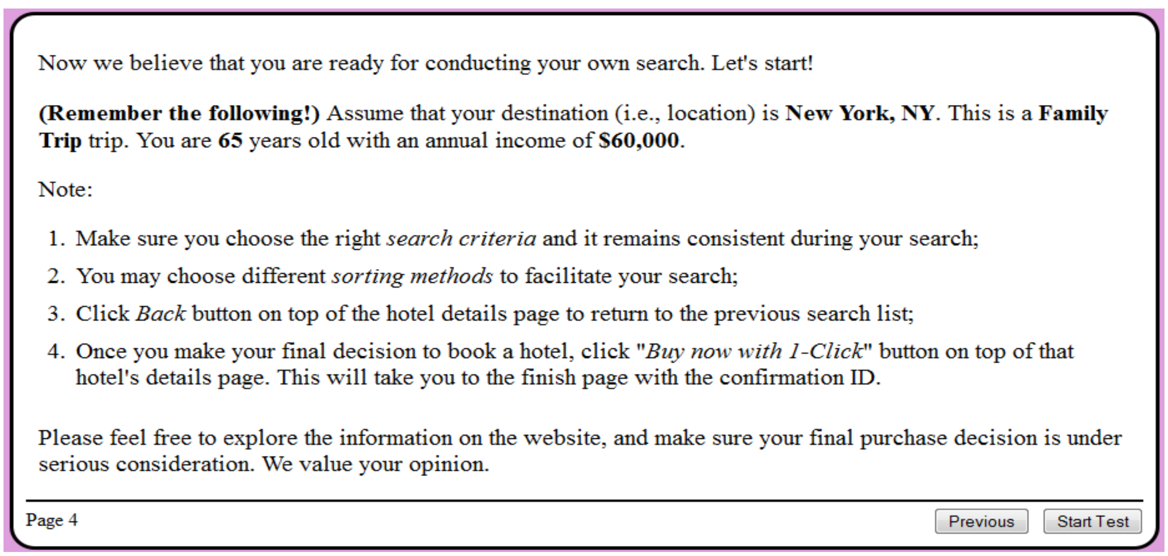
Welcome to the experiment!

In this experiment, you will be asked to conduct a ***hotel search*** using a set of criteria such as location, income, trip type and age . In the end, you need to make a ***simulated purchase*** based on your search. (You will NOT be charged for anything that you do during the course of this experiment.)

To make your experience better, we provide you with a quick two-page demo of our website prior to the experiment. Please click "Next" to continue.

Page 1 Next

Figure 4. Screenshot of the Introduction Page (2)



Now we believe that you are ready for conducting your own search. Let's start!

(Remember the following!) Assume that your destination (i.e., location) is **New York, NY**. This is a **Family Trip** trip. You are **65** years old with an annual income of **$60,000**.

Note:

1. Make sure you choose the right *search criteria* and it remains consistent during your search;
2. You may choose different *sorting methods* to facilitate your search;
3. Click *Back* button on top of the hotel details page to return to the previous search list;
4. Once you make your final decision to book a hotel, click "*Buy now with 1-Click*" button on top of that hotel's details page. This will take you to the finish page with the confirmation ID.

Please feel free to explore the information on the website, and make sure your final purchase decision is under serious consideration. We value your opinion.

Page 4 Previous Start Test

7. Results from Randomized Field Experiments

7.1. Effect of Ranking Mechanism on Search Engine

First, we look into how the design of ranking mechanism affects the performance and economic outcomes of search engine. We examine the total time spent, number of online activities, number of clicks, and purchase propensity at subject level from each of the four treatment groups in study I. Table 8 shows the average user behavior in an online session under different ranking mechanisms.

We find subjects who get to see BVR as the default ranking spend significantly larger amount of time in their search sessions, compared to subjects from other treatment groups. Meanwhile, subjects from the BVR group are engaged in significantly more activities, more clicks, and higher purchase propensity than subjects from other groups. Price-based ranking provides the second best performance, and followed by the rankings based on TripAdvisor and Travelocity ratings.

Table 8. Average User Behavior under Different Ranking Mechanisms

	Time Spent (seconds)	*Total # of Activities*	*# of Clicks*	*Purchase Propensity*
BVR	349.19	19.04	3.42	0.85
Price	163.01	12.59	1.62	0.79
Travelocity	130.33	11.33	1.17	0.54
TripAdvisor	244.97	12.19	1.52	0.61

Average over all subjects, across two cities (NYC and LA).

This result strongly suggests the design of ranking mechanism can significantly affect product search engine's performance. The best value/consumer utility-based ranking outperforms the other three popular ranking mechanisms. It can better motivate consumer online engagement and improve consumer click and purchase propensities.

7.2. Effect of Ranking Mechanism on Individual Product

In addition to the ranking effect at search engine level, we also found significant ranking effect at individual hotel level. More specifically, hotels ranked on top of the search result list received on average 2.04 times more clicks compared to the second-ranked hotels and 3.31 times more compared to the third-ranked ones. The decrease in clicks of hotels from the third-ranked position beyond diminishes. This trend stays consistent across two different cities, and regardless of the default ranking method. Table 9 shows the detailed number of clicks received for hotels at top-10 ranked positions. Moreover, we also examined the click-through rates for the same hotel that appeared at different rank positions under different default ranking mechanisms. Controlling for everything else, the same hotel with a higher screen position received significantly more clicks. For example, the "Blue Moon Hotel" in New York City received a total of 52 clicks under the BVR ranking where it was ranked on top. However, the same hotel received zero click under the price ranking where it was ranked the 31st.

Our findings from the randomized experiment highly correspond to the empirical results from the Bayesian archival data analysis, suggesting the effect from search engine ranking mechanism on consumer search and purchase behavior is causal and significant.

Table 9. Number of Clicks Received at Top-10 Ranking Positions

		Rank1	*Rank2*	*Rank3*	*Rank4*	*Rank5*	*Rank6*	*Rank7*	*Rank8*	*Rank9*	*Rank10*
BVR	NYC	52	25	14	16	11	12	10	4	1	2
	LA	64	24	14	12	10	6	7	5	3	1
Price	NYC	28	13	9	8	8	5	3	0	0	1
	LA	33	16	11	10	5	3	4	2	0	0
Travelocity	NYC	21	11	7	8	5	4	2	0	0	2
	LA	19	12	6	7	6	2	1	2	1	2
TripAdvisor	NYC	30	15	10	9	8	4	0	4	2	0
	LA	26	14	9	9	5	1	1	1	0	1

7.3. Effects of Active vs. Passive Personalization on Search Engine

Another important goal of our research is to examine how different personalization mechanisms influence the way consumers behave on product search engine. In study II, we consider three levels of personalization: active personalization with full access (control group), passive personalization without search context (treatment group 1), and passive personalization without weights of individual preferences (treatment group 2). Table 10 summarizes the average user behavior on the total time spent and total number of activities under the three different personalization mechanisms.

Table 10. Average User Behavior under Different Personalization Mechanisms

	Time Spent (seconds)	*Total # of Activities*
Active Personalization with Full Access	349.19	19.04
Passive Personalization w/o Search Context and Demographics	225.10	17.47
Passive Personalization w/o Weights of Individual Preferences	131.88	8.98
Average over all subjects, across two cities (NYC and LA).		

We find the active personalization mechanism attracts significantly higher user time and activities than the two passive mechanisms, with each user spending approximately 350 seconds and conducting 19 activities per session on average. This finding indicates that active personalization mechanism can generate higher online attentions. Meanwhile, passive mechanism without personalization on weights of individual preferences attracts the lowest user time and activities. This suggests that the distribution of consumer online activities tend to skew towards the hotel landing page, where the majority of consumer attention focuses on the personalization in the weights of preferences.

Furthermore, we look into the overall search engine click-through rate (CTR) and conversion rate (CR). Table 11 displays the total number of hotel impressions, number of clicks on hotel URLs, and number of conversions across all subjects and two cities, under the three personalization mechanisms. Based on these data, we can derive the overall search engine CTR and CR. Interestingly, we find search engine with an active personalization mechanism has a significantly higher CTR than the ones with passive personalization mechanisms (i.e., 0.09 vs. 0.03/0.04). However, active personalization leads to a significantly lower CR (i.e., 0.25 vs. 0.52/0.58).

We find this result interesting because one would expect the active personalization mechanism to increase, rather than decrease, the CR. However, in most online 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 personalization mechanism is more likely to increase sales when consumers do not have a planned purchase beforehand. Nevertheless, in our experimental setting, we focus on the type of consumers who have a planned purchase before the search starts. Under such scenario, the major advantage of active personalization is not dominant to the consumers, since consumers already have in mind what they are searching for. What is worse, if the personalization results do not meet consumers' expectation, they may easily kill the sale.

Regarding the magnitude of the effects, search context-based personalization presents an 11.5% larger negative effect, compared to weight-based personalization. This finding provides a plausible indication. It seems there are two types of personal information often used in the personalization: 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 lies closer with the former category, whereas weights of preferences belong to the latter. Our results suggest when designing the personalization mechanism, it is less recommended to use the identity related information, not only for privacy-preserving purpose, but also for the economic outcomes (i.e., conversion-based).

Table 11. Overall Search Engine Performance under Different Levels of Personalization

	Total # of Hotel Impressions	*Total # of Clicks on Hotel URLs*	*Total # of Conversions via Clicks*	*CTR*	*CR*
Active Personalization with Full Access	5530	481	119	0.09	0.25
Passive Personalization w/o Search Context and Demographics	3160	120	69	0.04	0.58
Passive Personalization w/o Weights of Individual Preferences	3555	121	63	0.03	0.52

Total over all subjects, across two cities (NYC and LA).

8. Conclusions and Implications

In this paper, we focus on investigating two major design issues increasingly faced by next-generation product search engines: which *ranking mechanism* to deploy in response to consumer queries and what kind of *personalization mechanism* to adopt, if at all. To examine the *causal* effect of different search mechanisms on consumer online search and purchase behavior, we combine archival data analysis with randomized experimental approaches.

Our archival data analysis is based on a unique dataset consisting of detailed information on 1 million online sessions from Travelocity.com over three months from 2008/11 to 2009/1. Besides, we supplement our search and transaction data with hotel service-, location- and customer review-based information collected using various machine learning techniques such as image classification and text mining tools. To study the relationship between search engine ranking and consumer search and purchase behavior, we propose a simultaneous equation model under a hierarchical Bayesian framework, and estimate it using the MCMC methods with a Metropolis-Hastings algorithm and a random walk chain. We are able to quantify the ranking effect on consumer click-through and purchase propensities.

However, evaluating the causal effects of search engine design features is difficult because search and purchase behavior are typically endogenous. Therefore, we design and conduct randomized field experiments based on a hotel search engine application designed and built by us to draw causal inferences about the effects from the ranking and personalization mechanisms, respectively. 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 under different search mechanisms.

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. Specifically, a consumer utility-based ranking mechanism yields the highest click and purchase propensities in comparison to existing benchmark systems such as ranking based on price or star ratings.

Our experiments on personalization shows that active personalization tools can attract higher online attention from consumers and lead to a higher click-through rates compared to passive personalization. Nevertheless, active personalization leads to a lower conversion rate, suggesting that it should not be adopted blindly. When consumers already have a planned purchase in mind, active personalization may cause the conversion rate to drop.

Our research sheds lights on understanding how consumers search, evaluate choices, and make purchase decisions in response to differences in search engine designs. We provide important empirical and experimental evidence for future studies to build on, in the process of designing an efficient ranking system and dynamically modeling consumer behavior on product shopping sites.

A good ranking mechanism can reduce consumers' search costs, improve the click-through rate and conversion rate of products, and improve the return-on-investments for search engines.

On a broader note, our inter-disciplinary approach provides a methodological framework for how econometric modeling, randomized field experiments, and IT-based artifacts can be integrated in the same study towards deriving causal relationships between variables of interest.

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 it difficult to control for subject characteristics across different treatment groups. The randomization process can alleviate such concern to a large extent. However, it would be helpful to conduct robustness tests based on offline subjects as well. We plan to conduct similar experiments using a pool of subjects drawn from non-AMT, offline sources. In our ongoing research we also plan to expand the scope and scale of these randomized experiments. Our current experiments focus on the type of consumers who have a planned purchase beforehand and can make at most one purchase in each online shopping session. To better understand the counter-intuitive finding that active personalization leads to a lower conversion rates, we plan to extend our experimental design and compare with consumers who do not have a planned purchase and are allowed to make N purchases in one session (N=0, 1, 2, 3, …). Our current experiments are based on a relatively small sample of participants. Because the experiments are still ongoing, we are confident that we will be able to validate our results on a much larger scale by the time we present this work at WISE, provided we are given the opportunity.

References

Agrawal, R., Gollapudi, S., Halverson, A., and Leong, S. 2009. Diversifying Search Results. In *Proceedings of WSDM' 2009*. Barcelona, Spain.

Ansari, A., and Mela, C. F. (2003). E-customization. JMR, *Journal of Marketing Research*, 40(2), 131–145. doi:10.1509/jmkr.40.2.131.19224.

Aral, S. and D. Walker. 2011. Creating social contagion through viral product design: A Randomized trial of peer influence in networks. *Management Science*. Forthcoming.

Archak, N., A. Ghose, P. Ipeirotis. 2011. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*. Forthcoming.

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

Arora, Neeraj, Xavier Dreze, Anindya Ghose, James D. Hess, Raghuram Iyengar, Bing Jing, Yogesh Joshi, V. Kumar, Nicholas Lurie, Scott Neslin, S. Sajeesh, Meng Su, Niladri Syam, Jacquelyn Thomas, and Z. John Zhang (2008), Putting One-to-One Marketing to Work: Personalization, Customization, and Choice, *Marketing Letters*, 19, 305–21.

Aula, A. and K. Rodden. 2009. Eye-tracking studies: More than meets the eye. http://googleblog.blogspot.com/ 2009/02/eye-tracking-studies-more-than-meets.html

Baye, M.R., Gatti, J.R.J., Kattuman, P. and Morgan, J. 2009. Clicks, Discontinuities, and Firm Demand Online. *Journal of Economics & Management Strategy*. 18(4), 935-975.

Bhattacharya, S., Gollapudi, S., and Munagala, K. 2011. Consideration Set Generation in Commerce Search. In *Proceedings of WWW 2011*. Hyderabad, India.

Chatterjee, P., D. Hoffman, and T. Novak. 2003. Modeling the clickstream: Implications for Web-based advertising efforts. *Marketing Science*. 22(4) 520–541.

Chellappa, R.K. and Sin, R.G. 2005. Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. *Information Technology and Management*, Volume 6, Numbers 2-3, 181-202.

Chib, S., E. Greenberg. 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4), pp. 327-335.

Chittor, V. 2010. Online Retail: Getting the Right Product in Front of Your Customers. http://www.inc.com/internet/articles/201002/chittoor.html.

Ellison, G. and Ellison, S.F. 2009. Search, Obfuscation, and Price Elasticities on the Internet. *Econometrica*, 77, 427-452.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. *Information Systems Research.* 19, 3, 291–313.

Ghose, A. and Ipeirotis, P. G. 2011. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE TKDE*.

Ghose, A., Ipeirotis, P. G. and Li, B. 2011. Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowd-Sourced Content. *Working paper*.

Ghose A, and Yang, S. 2009. An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *Management Science*. 55(10), pp. 1605-1622.

Givon, Moshe M. 1984, Variety Seeking Through Brand Switching. *Marketing Science*, 3 (Winter), 1-22.

Gourville, J. T., D. Soman. 2005. Overchoice and assortment type: When and why variety backfires. *Marketing Science*. 24(3) 382–395.

Gretzel, U., D.R. Fesenmaier and J.T. O'Leary. 2006. The Transformation of Consumer Behavior. *Tourism Business Frontiers*. Edited by D. Buhalis and C. Costa. Oxford, UK: Butterworthh-Heinemann.

Henzinger, M. 2007. Search Technologies for the Internet. *Science*, 317 (5837): 468-71.

Iyengar, S. S., G. Huberman,W. Jiang. 2004. How much choice is too much? Contributions to 401(k) retirement plans. O. S. Mitchell, S. P. Utkus, eds. Pension Design and Structure: New Lessons from Behavioral Finance. *Oxford University Press*, Oxford, UK, 83–96.

Kuksov, D. and Villas-Boas, J.M. 2010. When More Alternatives Lead to Less Choice. *Marketing Science*. Vol. 29, No. 3, pp. 507–524.

Lavrenko, V. and Croft, W.B.. 2001. Relevance based language models, in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 120-127.

Malthouse, E. C., and Elsner, R. (2006). Customization with cross-basis sub-segmentation. *Database Marketing & Customer Strategy Management*, 14(1), 40–50.

McAlister, Leigh. 1982, A Dynamic Attribute Satiation Model of Variety-Seeking Behavior. *Journal of Consumer Research*, 9, 141-150.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psych. Rev*. 63 81–97.

Moe, W. and P. S. Fader. 2003. Dynamic conversion behavior at e-commerce sites. *Management Science*. 50(3) 326–335.

Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2, 1-2.

Panigrahi, D. and Gollapudi, S. 2011. Result Enrichment in Commerce Search using Browse Trails. In *Proceedings of WSDM' 2011*. Hong Kong, China.

Ranjith, G. 2005. Interferon-α-Induced Depression: When a Randomized Trial Is Not a Randomized Controlled Trial, *Psychother Psychosom* 74:387.

Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*, 15(4), 321–340.

Rossi, P. E. and Greg M. Allenby. 2003. Bayesian Statistics and Marketing. *Marketing Science*. Vol. 22, No. 3, pp. 304-328.

Rutz, O. and R.E. Bucklin. 2007. A Model of Individual Keyword Performance in Paid Search Advertising. *Working paper*, Yale University, New Haven, CT.

Shugan, S. M. 1980. The cost of thinking. *Journal of Consumer Research*. 7(2) 99–111.

Simon, H. A. 1955. A behavioral model of rational choice. *Quart. J. Econom*. 69(1) 99–118.

Srikant, R., Basu, S., Wang, N. and Pregibon, D. 2010. User Browsing Models: Relevance versus Examination. In *Proceedings of ACM Conference on Knowledge Discovery and Data Mining (KDD'10)*, Washington, DC.

Yang, S. and A. Ghose. 2010. Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science*, 29, 602-623.

Zhang, J. and Wedel, M. 2009. The effectiveness of customized promotions in online and offline stores. *Journal of Marketing Research.* Vol. 46, No. 2, pp. 190-206.