Reprinted with permission from *Journal of Marketing Research*, published by the American Marketing Association, Kinshuk Jerath, Liye Ma, and Young-Hoon Park, 51, no. 4 (2014): 480-486.

KINSHUK JERATH, LIYE MA, and YOUNG-HOON PARK*

The authors study consumers' click behavior on organic and sponsored links after a keyword search on an Internet search engine. Using a data set of individual-level click activity after keyword searches from a leading search engine in Korea, the authors find that consumers' click activity after a keyword search is low and heavily concentrated on the organic list. However, searches of less popular keywords (i.e., keywords with lower search volume) are associated with more clicks per search and a larger fraction of sponsored clicks. This indicates that, compared with more popular keywords, consumers who search for less popular keywords expend more effort in their search for information and are closer to a purchase, which makes them more targetable for sponsored search advertising.

Keywords: sponsored search advertising, organic search listing, click-through behavior, keyword search volume

Online Supplement: http://dx.doi.org/10.1509/jmr.13.0099

Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity

Consumers use online search engines as tools to search for information on the World Wide Web. Examples of popular search engines include Google, Yahoo, and Bing in many countries worldwide; Yandex in Russia; Baidu in China; and Daum and Naver in Korea. When a user searches using a keyword on a search engine, he or she is typically presented with two lists of search results of web pages relevant to the search query: a list of "organic" results and a list of "sponsored" results.[1] The search engine determines the organic results by finding web pages relevant to the search query, typically information-based web pages. The sponsored links are determined using online auctions in which advertisers bid to be placed in response to queries by consumers and therefore are more commercial. This type of advertising is called "sponsored search" or "paid search" advertising.

When presented with lists of organic and sponsored links in response to their search queries, how do consumers respond in terms of clicks on both types of links? How does click behavior on the two lists vary across keywords? Are there systematic patterns in variations across keywords? Which keyword characteristics can help inform these patterns? What is the pattern of heterogeneity across searches by consumers, and how can this heterogeneity be explained? Because Internet users depend heavily on search engines to find information on the Web, it is crucial for advertising firms, researchers, and search engine companies to obtain answers to these questions.

The commercial success of sponsored search advertising in the past decade has motivated a significant body of work studying its different aspects (e.g., Agarwal, Hosanagar, and Smith 2011, 2012; Amaldoss, Jerath, and Sayedi 2014; Chan and Park 2014; Edelman, Ostrovsky, and Schwarz 2007; Ghose and Yang 2009; Goldfarb and Tucker 2011; Jerath et al. 2011; Jerath and Sayedi 2012; Joo et al. 2014; Katona and Sarvary 2010; Rutz and Bucklin 2011; Sayedi, Jerath, and Srinivasan 2014; Varian 2007; Yang and Ghose

[1]Throughout the article, we use "keyword," "query," "search query," "search phrase," and so on, interchangeably. A "keyword" may be a phrase with more than one word in it.

*Kinshuk Jerath is Associate Professor of Marketing, Columbia Business School, Columbia University (e-mail: jerath@columbia.edu). Liye Ma is Assistant Professor of Marketing, Robert H. Smith School of Business, University of Maryland (e-mail: liyema@rhsmith.umd.edu). Young-Hoon Park is AMOREPACIFIC Professor of Management and Associate Professor of Marketing, Samuel Curtis Johnson Graduate School of Management, Cornell University (e-mail: yp34@cornell.edu). All authors contributed equally and are listed in alphabetical order. The authors thank the company (which has asked to remain anonymous) that provided the data used in this study, as well as Eric Bradlow, Pete Fader, Mingyu Joo, Carl Mela, Wendy Moe, Ken Wilbur, Song Yao, and Yi Zhu for their valuable comments. Avi Goldfarb served as associate editor for this article.

2010; Yao and Mela 2011). This literature stream has primarily focused on which keywords advertisers should bid on, what their bidding strategies should be, and how advertisers can improve the performance (in terms of click-through and conversion rates) of their sponsored advertisements. Yang and Ghose (2010) and Agarwal, Hosanagar, and Smith (2012) empirically study how the presence of an advertising firm's own link and competitors' links in the organic listing influence click and conversion behavior for the focal firm's sponsored ads, and vice versa. Broadly speaking, both studies find complementarities between click-through rates on firms' organic and sponsored links. Existing studies, however, use data from a single advertising firm and therefore lack data on clicks on sponsored and organic links of other entities on the search results page. In other words, they do not have sufficient data to present a comprehensive picture of user activity on the search results page. In this study, we use data obtained from a search engine and have information on clicks on the full lists of sponsored and organic links presented after a keyword search. Using these data, we are able to paint a more complete picture of user activity on the search results page.

We analyze data on approximately 1.63 million keyword searches over a one-month period for 120 keywords. We model the click counts on both the organic and the sponsored lists, incorporating both observed and unobserved heterogeneity at both the keyword level and the search instance level. We ask the key question: Which keyword characteristics serve as good indicators of consumer response after a keyword search?

Previous studies have typically studied keyword characteristics such as whether the search phrase includes the name of a brand or a retailer, the length of the search phrase, and so on (e.g., Agarwal, Hosanagar, and Smith 2011; Ghose and Yang 2009; Rutz and Bucklin 2011; Rutz, Bucklin, and Sonnier 2012; Rutz and Trusov 2011; Rutz, Trusov, and Bucklin 2011; Yang and Ghose 2010). Other studies (e.g., Rutz, Bucklin, and Sonnier 2012) have examined the semantic characteristics of the search phrase, whereby the semantic characteristics are determined using managerial knowledge of the business domain. Such keyword characteristics are inherent to the keyword searched (i.e., they can be directly determined from the keyword). The aforementioned studies find these characteristics to be correlated with click and conversion behavior. In our study, in addition to aforementioned keyword characteristics, we include a measure of the popularity of a keyword. Keyword popularity is a fundamentally different type of characteristic from those used in previous research because it cannot be determined by inspecting the keyword itself; rather, it depends on how many search engine users searched for the focal keyword relative to other keywords.

Our novel finding is that keyword popularity is a key determinant of consumer click behavior after a keyword search. Specifically, on average, both the number of clicks per search and the share of sponsored clicks are larger for less popular keywords than for more popular keywords. Furthermore, we find that consumers can be classified into segments on the basis of their click patterns, which can be interpreted as corresponding to different stages of consumer involvement with the topic they are searching about or the product they want to purchase. Specifically, we find that lower-involvement consumers search more for popular keywords, whereas higher-involvement consumers search more for less popular keywords. This finding is consistent with the "purchase funnel" theory of consumer purchasing (Howard and Sheth 1969) and also resonates with the results in Moe (2003), who shows that different consumers conduct online activity in different stages of the purchase process.

We organize the rest of this article as follows: In the next section, we provide an overview of our data. Following this, we develop and estimate our formal model and discuss the results and insights we obtain. We conclude with a discussion of the implications of our research and directions for further research. We provide additional details and analyses in the Web Appendix.

DATA OVERVIEW

We obtained a data set of individual-level consumer click activity after keyword searches from a leading search engine firm in Korea. Subsequent to a keyword search, we observe which sponsored ads are displayed in response to the consumer's search query and which sponsored ads and organic links the user clicked. We do not have information on the full list of organic links displayed to the consumer; however, we do have data on how many organic links the user clicked.

The page layout used by the Korean search engine is similar to the layout used by popular U.S. search engines such as Google, Yahoo, and Bing. In response to a keyword search, a list of sponsored ads is placed at the top of the results page, with a maximum of five ads displayed. A list of organic links appears below the list of sponsored links. This layout is similar to the layout of popular U.S. search engines, which display several (typically, up to three) sponsored links on the top of the results page, followed by organic links, with the remaining sponsored links (if any) on the right-hand side of the page. The organic links on the Korean search engine are typically grouped according to the source of the content (e.g., news, blogs, images, videos), which is similar to the practice followed by U.S. search engines. Given the similarities in page layout across search engines, the insights we obtain from our analysis can shed light to a large extent on user click behavior on other search engines as well.

The search engine provided us data on search and click activity for 1,200 keywords over the one-month period (specifically, 28 days) of February 2011. The search engine chose and provided us the keywords considered in this research, which represent products and services for which the search engine expects consumers to be relatively active; therefore, firms also advertise on these keywords. Given keywords that pass this criterion, the search engine provided keywords to ensure significant variation in keyword search volume. Note that a "keyword" used in a query may be a single word or a phrase of a few words. The total number of search instances for 1,200 keywords add up to more than 30 million. This is a prohibitively large data set given the complexity of estimation of the model we use. Therefore, for our research, we sample 120 keywords from the 1,200 keywords uniformly at random. An assessment using exploratory methods indicated that the data set with the 120 sampled keywords is representative of the full data set with 1,200 keywords. We ensure that there is exactly one (randomly chosen) search instance per unique IP address in the data, which guarantees, with high certainty, that no more than one search instance per person is included in the estimation data. Because we focus on a cross-sectional analysis rather than on individual-level behavior across search instances, this is an appropriate data structure.

We list the 120 keywords, along with their search volumes (after processing the data as described previously), in Section I of the Web Appendix. This new data set contains 1,631,336 total searches across the 120 keywords. On average, there are 13,595 search queries per keyword, ranging from a minimum of 1,241 to a maximum of 278,458 search queries. An average of 4.39 ads are displayed per search, with a relatively low standard deviation of 1.21.

MODEL DEVELOPMENT

Our objective is to model the number of clicks by a user on the organic and sponsored lists that he or she is presented with after a keyword search. Each search instance is associated with a user. To enable users in different search instances to have different click behavior, we allow for search instance–level heterogeneity. As we discussed previously, our data structure practically ensures that we do not have consumers in the data who conduct multiple searches. Therefore, although we can account for unobserved search instance–level heterogeneity in our model, we refrain from making inferences about consumers at the individual level. For expositional ease, we use "consumer" and "search instance" interchangeably.

We posit that consumers who search the keywords are from several latent segments (Kamakura and Russell 1989); this approach allows for different behaviors on both the overall click propensity and the propensity to click sponsored or organic links. We assume that, for keyword k, the segment s for search instance i in which keyword k is searched is a random draw from a multinomial distribution with probabilities given by the vector $\pi_k = (\pi_{k,1}, ..., \pi_{k,S})$, where $\pi_{k,s}$ is the probability of keyword k being in segment s, and $\Sigma_{s=1}^{S}\pi_{k,s} = 1$. Note that this probability vector is specific to each keyword because we account for the possibility that different keywords attract consumers from different segments in different proportions. Furthermore, the same keyword may belong to different segments in the context of different search instances.

Let $y_{ki} \in \mathbb{Z}_{\geq 0}$ denote the total number of links clicked by consumer i after searching keyword k (including clicks on both sponsored and organic links). We assume that y_{ki} follows a Poisson distribution with rate parameter $\lambda_{ki,s}$. Next, we assume that for each click, there is probability $p_{ki,s}$ that the click will be on a sponsored link (i.e., we assume a Bernoulli process with parameter $p_{ki,s}$). The subscripts "ki, s" for λ and p indicate that these are consumer i's mean propensities for keyword k given that she belongs to segment s when conducting this search. We model $\lambda_{ki,s}$ and $p_{ki,s}$ as follows:

$$(1)\quad \ln(\lambda_{ki,s}) = \beta^{\lambda}_{k,s} + \beta^{\lambda}_{POP} \times \ln(\text{Popular}_k) + \beta^{\lambda}_{X}X_k + \beta^{\lambda}_{Z}Z_i\text{, and}$$

$$(2)\quad \text{logit}(p_{ki,s}) = \beta^{p}_{k,s} + \beta^{p}_{POP} \times \ln(\text{Popular}_k) + \beta^{p}_{X}X_k + \beta^{p}_{Z}Z_i.$$

In Equations 1 and 2, Popular_k is a measure of the popularity of keyword k and is defined as the rank of keyword k on the basis of the search query volume during the data period, with the most-searched keyword ranked at the top as 1. Therefore, a larger value of Popular_k indicates that the keyword is less popular.[2] X_k is a vector of keyword-specific covariates. Through X_k, we include three important observed keyword-specific characteristics: (1) Retailer_k—whether the keyword has retailer-specific information (i.e., whether a seller/retailer name appears in the query); (2) Brand_k—whether the keyword has brand-specific information (i.e., whether a brand name appears in the query); and (3) Length_k—the length (number of words) of the search phrase. Table A1 in Section II of the Web Appendix reports the descriptive statistics of the keyword popularity measure and the three keyword characteristics. We employ these keyword-specific covariates to control for observed heterogeneity across keywords. Z_i includes covariates specific to the instance of the search by consumer i. We incorporate two covariates here. First, we include Num_Sponsored_i, which denotes the number of sponsored links displayed after the keyword search is conducted (the number of available sponsored links can be expected to influence the number of sponsored links clicked and, therefore, the overall number of clicks as well). Second, we include Weekend_i, which denotes a weekend indicator to account for day-of-the-week effect.[3]

With respect to the parameters in Equation 1, the parameter $\beta^{\lambda}_{k,s}$ represents the baseline click propensity for a consumer in latent segment s after searching keyword k. $\beta^{\lambda}_{k,s}$ is a keyword–segment-specific intercept that captures heterogeneity across consumers in different segments in their tendency to click after searching a particular keyword. The parameter β^{λ}_{POP} indicates how click propensity changes as a function of the keyword popularity: if β^{λ}_{POP} is negative, the average number of clicks (per search) is greater for more popular (higher search volume) keywords, whereas if β^{λ}_{POP} is positive, the average number of clicks (per search) is lesser for more popular keywords. β^{λ}_{X} is a vector of three coefficients measuring the impact of Retailer_k, Brand_k, and Length_k, respectively, on the propensity to click. β^{λ}_{Z} is a vector of two coefficients measuring the impact of Num_Sponsored_i and Weekend_i, respectively, on the propensity to click. In Equation 2, the $\beta^{p}_{\cdot}$ parameters play the same roles as described for the corresponding $\beta^{\lambda}_{\cdot}$ parameters except that the effect is on the probability of clicking a sponsored link.

We adopt a hierarchical Bayesian framework and assume that the keyword–segment-specific intercepts for click and sponsored propensity for each keyword–segment pair are random draws from segment-specific normal distributions given by the following equation:

$$(3)\qquad \beta^{\lambda}_{k,s} \sim N\left[\bar{\beta}^{\lambda}_{s}, \left(\sigma^{\lambda}_{s}\right)^2\right] \text{ and } \beta^{p}_{k,s} \sim N\left[\bar{\beta}^{p}_{s}, \left(\sigma^{p}_{s}\right)^2\right],$$

[2]We note two points here: (1) for the 120 sampled keywords, we continue to use the ranks according to the entire set of 1,200 keywords, and (2) using keyword search volume directly instead of the rank yields qualitatively similar results. We use the rank of a keyword as the observed popularity measure because this is a direct indicator of relative popularity.

[3]We do not have data on individual-level demographics (e.g., age, income, sex).

where $\overline{\beta}_s^{\lambda}$ and $\overline{\beta}_s^{p}$ are the respective segment-level means and $(\sigma_s^{\lambda})^2$ and $(\sigma_s^{p})^2$ are the respective segment-level variances. We note that for keyword k, except for the baseline parameters (i.e., $\beta_{k,s}^{\lambda}$ and $\beta_{k,s}^{p}$), which are segment specific, we assume all parameters to be population specific to maintain simplicity of interpretation of the results.

In summary, click behavior on the search results page is governed by two components of the model: the overall propensity to click and the likelihood to search for information in the sponsored versus organic listings. Furthermore, the model accounts for (1) observed heterogeneity in keywords (through keyword popularity and other keyword characteristics), (2) observed heterogeneity in consumers (through characteristics of the search instances), and (3) unobserved heterogeneity among consumers (through latent segments).

ESTIMATION AND RESULTS

Estimation and Model Fit

We have a total of 1,631,336 search instances for the 120 keywords we consider. To allow for a shorter estimation time, we randomly sample 20% of the consumer searches from the sample; this 20% subsample contains 326,080 search instances. As Table A2 in Section II of the Web Appendix shows, the values of key summary statistics between the full data of the 120 chosen keywords and the 20% sample are very close.

We adopt a Bayesian approach and use the Markov chain Monte Carlo (MCMC) method to estimate our proposed model. Section III in the Web Appendix provides the details of the MCMC procedure. We draw samples from the posterior distribution of 40,000 iterations from two independent MCMC chains following a burn-in of 40,000 iterations.

We estimate the model with different numbers of consumer segments, ranging from one to six. We report the log-marginal density (LMD) for the models with different numbers of consumer segments in Table 1, which shows that as the number of consumer segments increases from one to four, LMD increases significantly; after four segments, however, LMD practically levels off. On examining the mean absolute error (MAE) in predicted total number of clicks for the models with four, five, and six segments, we find that the MAE has the value of 1.28 for these three models. In other words, increasing the number of segments from four to six leads to a very small increase in LMD and does not improve MAE. We conclude that the model with four segments is appropriate, and we therefore focus on this model hereinafter. (The results with five or six segments are qualitatively the same.)

As an additional measure of the accuracy of the model with four segments, we calculate, for each keyword, the expected number of organic and sponsored clicks over the data period and compare them with the actual numbers of clicks. Across the 120 keywords, mean absolute percentage errors weighted by search volume are 2.33%, 1.08%, and 2.09% for organic, sponsored, and total clicks, respectively. These statistics provide strong evidence that the proposed model with four segments performs well in capturing click behavior for both organic and sponsored links at the keyword level.

Table 1

LMD FOR DIFFERENT NUMBERS OF LATENT SEGMENTS

Number of Segments	*LMD*
1	–525,573.0
2	–494,724.7
3	–491,879.7
4	–491,099.1
5	–490,650.8
6	–490,555.4

Results

We organize the reporting of our results and the associated insights about user click behavior into three main parts. We report all parameter estimates in Table 2.

Characteristics of segments. As discussed previously, we obtain four latent segments. We first examine the keyword–segment-specific intercepts for click propensity. The population-level mean estimates, denoted by $\overline{\beta}_s^{\lambda}$, s ∈ {1, 2, 3, 4} are –.128, .478, 1.227, and 2.124, respectively; these correspond to .88, 1.61, 3.41, and 8.36 average clicks per search for Segments 1, 2, 3, and 4, respectively. Keeping the same order of segments, the population-level mean estimates for the keyword–segment-specific intercepts for the propensity to click sponsored links, denoted by $\overline{\beta}_s^{p}$, s ∈ {1, 2, 3, 4} , are –3.574, –3.362, –1.660, and –2.346, respectively; these correspond to sponsored click probabilities of 2.73%, 3.35%, 15.98%, and 8.74% for Segments 1, 2, 3, and 4, respectively. These estimates imply the following: First, the average number of clicks after a keyword search is quite small,[4] and the share of sponsored clicks is also quite small (the weighted averages across segments being 1.44 and 3.67%, respectively). Second, in the ordering we impose, for higher-numbered segments (i.e., Segments 3 and 4, compared with Segments 1 and 2), consumers are inclined to click more links per search and also to click sponsored links with higher probability. These results indicate that Segments 1 and 2 are lower-involvement segments, and Segments 3 and 4 are higher-involvement segments. We note that the behavioral differences among the segments are significant—a consumer in Segment 4 clicks almost ten times more than a consumer in Segment 1 after a search and is more than three times as likely to click a sponsored link.

Next, on the basis of inferred search instance memberships, we find that Segments 1, 2, 3, and 4 have relative sizes of 49.11%, 44.96%, 4.20%, and 1.73%, respectively (i.e., 49.11% of all keyword searches fall in Segment 1, 44.96% of all searches fall in Segment 2, and so on). In other words, lower-numbered segments are larger than higher-numbered segments. Furthermore, search instances for individual keywords are not distributed across segments in proportion to overall segment sizes. Indeed, less popular (i.e., lower search volume) keywords in general have a larger portion of their searches by consumers from higher-numbered segments compared with more popular keywords. This can be observed from Table 3, in which we report the average percentage of consumers from each segment for the top 30 most popular keywords, the next 30 most popular keywords, and so on in our 120-keyword data set. We find that as the keyword ranks increase (i.e., as the keywords become

[4]This result resonates with the results of Johnson et al. (2004), who report that the amount of online search conducted by consumers across websites is quite limited.

Table 2
PARAMETER ESTIMATES

	Segment 1	Segment 2	Segment 3	Segment 4
A: Parameters Describing Latent Segments				
$\overline{\beta}_s^{\lambda}$	–.128 (–.267, .013)	.478 (.338, .623)	1.227 (1.105, 1.341)	2.124 (1.996, 2.248)
$(\sigma_s^{\lambda})^2$	.529 (.393, .699)	.383 (.287, .513)	.260 (.182, .361)	.414 (.298, .559)
$\overline{\beta}_s^{p}$	–3.574 (–3.884, –3.297)	–3.362 (–3.755, –2.998)	–1.660 (–1.978, –1.369)	–2.346 (–2.547, –2.139)
$(\sigma_s^{p})^2$	1.714 (1.135, 2.685)	3.216 (2.316, 4.333)	2.176 (1.494, 3.033)	.946 (.641, 1.329)
B: Coefficients for Keyword-Level Covariates				
β_{POP}^{λ}	.012 (.003, .020)			
β_{POP}^{p}	.299 (.095, .425)			
	Retailer$_k$	*Brand$_k$*	*Length$_k$*	
β_X^{λ}	–.168 (–.233, –.069)	–.044 (–.150, .027)	.035 (–.012, .087)	
β_X^{p}	.178 (–.243, .556)	–.170 (–.497, .037)	–.009 (–.147, .107)	
C: Coefficients for Search-Level Covariates				
	Num_Sponsored$_i$	*Weekend$_i$*		
β_Z^{λ}	–.018 (–.029, –.007)	–.007 (–.016, .001)		
β_Z^{p}	.413 (.360, .463)	–.007 (–.048, .038)		

Notes: Values in parentheses are the 95% credible intervals.

Table 3
SEGMENT PROPORTIONS FOR KEYWORDS (GROUPED BY KEYWORD POPULARITY RANK)

Keyword Rank	*Segment 1*	*Segment 2*	*Segment 3*	*Segment 4*
1–30	56.22%	37.85%	3.81%	2.11%
31–60	54.61%	33.22%	9.26%	2.91%
61–90	56.12%	28.32%	11.54%	4.01%
91–120	49.54%	29.46%	14.98%	6.02%

less popular), the proportions of Segment 1 and Segment 2 decrease while the proportions of Segment 3 and Segment 4 increase. We observe a clear pattern indicating that Segments 1 and 2—the low-involvement segments—have a larger proportion of keywords with higher search volume than Segments 3 and 4—the high-involvement segments. Consequently, more popular keywords have fewer clicks per search and a larger proportion of clicks on the organic listing in comparison with less popular keywords.

It is widely accepted in marketing that consumers move toward a purchase through a hierarchical sequence of events, captured by a "purchase funnel" model (Howard and Sheth 1969)—for example, the Awareness-Interest-Desire-Action (or AIDA) model. Notably, the consumer segments we identify correspond to a purchase funnel model—consumers are in different stages of involvement, and the segments representing higher involvement are smaller than segments representing lower involvement.

Effect of popularity. From the posterior means of β_{POP}^{λ} and β_{POP}^{p}, we find that more popular keywords receive fewer clicks per search and receive a smaller fraction of clicks on sponsored links, an effect that is in agreement with, but in addition to, the analysis discussed previously. To further understand the effect of popularity, we also estimated the model with only one segment (i.e., we do not allow for consumer heterogeneity through latent segments). As Table 1 reveals, the LMD of the one-segment model is significantly worse than that of models with two or more segments. This shows that allowing for multiple consumer segments is indeed crucial for understanding consumers' click behavior. More importantly, in the one-segment model, compared with the four-segment model, both β_{POP}^{λ} and β_{POP}^{p} have values that are considerably larger and have the same signs (i.e., both effects are in the same direction but are stronger). (In the model with one segment, β_{POP}^{λ} has the value .053 with a 95% credible interval of [.019, .097], and β_{POP}^{p} has the value .606 with a 95% credible interval of [.506, .719].) This finding leads to a noteworthy conclusion: a large part of the impact of keyword popularity occurs through consumer selection into different segments. In other words, the finding that inclusion of multiple segments weakens the direct impact of popularity indicates that the effect of popularity on click behavior occurs through the different stages of involvement of consumers.

Effect of covariates in X_k and Z_i. In examining the other parameters for observed heterogeneity, we find that the covariates Retailer_k, Brand_k, and Length_k largely have no impact on either the overall click propensity after a search or the propensity to click a sponsored link. The only exception is that the click propensity is lower if the searched keyword contains retailer-specific information. Turning to the search instance–specific covariates, we find that the number of sponsored links displayed at the time of search is positively correlated with the propensity to click sponsored links. This may be due to an agglomeration effect (i.e., more sponsored links draw greater attention from the user). We also find that a weekend search is not different from a weekday search in terms of either the overall click propensity or the propensity to click sponsored links.

Comparison with Previous Literature

Existing articles that study the impact of observed keyword characteristics on click behavior (e.g., Agarwal, Hosanagar, and Smith 2011, 2012; Ghose and Yang 2009; Rutz and Bucklin 2011; Rutz, Bucklin, and Sonnier 2012; Rutz and Trusov 2011; Rutz, Trusov, and Bucklin 2011; Yang and Ghose 2010) have found that search phrases that (1) include a retailer name or a brand name and (2) are longer have higher click-through rates on sponsored links. (We note that these articles do not view popularity as a covariate.) Notably, we do not find strong effects of observed keyword characteristics on either the overall click propensity or the propensity to click sponsored versus organic links. However, we do find a strong effect of keyword popularity (i.e., the relative search volume) on both the overall click propensity and the propensity to click sponsored versus organic links. In our data, the correlation between keyword search volumes and presence of a retailer name in the keyword is .182, the correlation between keyword search volume and presence of a brand name in the keyword is .169, and the correlation between keyword search volume and length of the keyword is .003; all these correlations are weak. In this light, our results indicate that keyword popularity is an important characteristic that determines click behavior after a keyword search and that, in general, this characteristic is different from the observed keyword characteristics mentioned previously.[5]

CONCLUSIONS AND DISCUSSION

In this article, we study consumers' click behavior on the organic and sponsored results presented to them after a keyword search on an Internet search engine. We analyze a unique data set obtained from a Korean search engine for more than 1.5 million keyword searches of 120 different keywords over the span of one month.

Our novel empirical finding is that the total clicks and the proportion of sponsored clicks after a keyword search is greater for less popular keywords. Furthermore, we find that users can be grouped into latent segments of low-involvement and high-involvement consumers, in which the latter has more clicks per search and a larger fraction of sponsored clicks than the former. Segments representing low-involvement consumers are composed of those who typically search for more popular keywords, and vice versa.

Previous studies have typically used observed keyword characteristics (e.g., the presence of a retailer name or a brand name in the search phrase, the length of the search phrase). In contrast, popularity is a fundamentally different type of keyword characteristic in that it cannot be determined by inspecting the keyword itself (because it depends on how many search engine users searched the focal keyword relative to other keywords).[6]

The result that consumers searching for more popular keywords focus relatively more on the organic results whereas consumers searching for less popular keywords focus relatively more on sponsored results suggests that advertisers might want to focus their sponsored search advertising efforts on less popular keywords and focus their search engine optimization efforts on more popular keywords. In addition, the insights that we uncover into consumers' click behavior can also help search engines design better responses to consumer queries and, therefore, better serve both search engine users and advertisers.

Further research can build on our findings by replicating our analysis on similar data sets obtained from other sources, which might lead to empirical generalizations regarding postsearch consumer click activity. Future studies can also address certain shortcomings of our research. First, consumers' click behavior depends on the relevance of the results presented to the keyword searched. Given our data, we are unable to address this aspect. Richer data are needed to explicitly incorporate relevance into the model; for example, we may need data on the identities of all firms that are displayed in the organic and sponsored lists, the ad copy used, the landing pages that the user is directed to, and so on. Second, we show that keyword popularity is an important indicator of searchers' click tendencies, and we argue that consumers in the segments we identify occupy different stages of involvement with respect to the relevant search, on the basis of their click behavior. However, we stop short of making causality arguments. Further research could control for stages of consumer involvement more carefully by running experiments, possibly in the manner of Lambrecht, Seim, and Tucker (2011) and Lambrecht and Tucker (2013), to infer causality. Further research could also account for possible endogeneity issues. For example, if advertisers anticipate that less popular keywords are more conducive to sponsored search advertising, they may place their advertisements accordingly. Third, future studies could conduct a deeper analysis of the process through which consumers decide on the keywords to search in the different stages over time. For this analysis, a richer longitudinal data set of search behavior at the individual consumer level may be required. Fourth, the relative popularity of keywords will change over time. It would be useful to obtain temporal data to study the length of time for which keyword popularity is stable and how this varies across keywords. We hope that our study can motivate further research in these directions.

REFERENCES

Agarwal, Ashish, Kartik Hosanagar, and Michael D. Smith (2011), "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets," *Journal of Marketing Research*, 48 (December), 1057–73.

———, ———, and ——— (2012), "Do Organic Results Help or Hurt Sponsored Search Performance?" working paper, University of Texas at Austin.

Amaldoss, Wilfred, Kinshuk Jerath, and Amin Sayedi (2014), "Keyword Management Costs and 'Broad Match' in Sponsored Search Advertising," working paper, Duke University.

Chan, Tat Y. and Young-Hoon Park (2014), "The Value of Consumer Search Activities for Sponsored Search Advertisers," working paper, Washington University in St. Louis.

Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz (2007), "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review*, 97 (1), 242–59.

[5]Although it would be reasonable to argue that consumers search less often for longer search phrases or those that contain the name of a retailer or brand, this does not imply that all keywords that are searched less have one or more of the aforementioned observed characteristics.

[6]It can be argued that the popularity of a keyword is a latent construct, rather than one that can be calculated directly from the observed keyword search volume (as we have done in our analysis). The observed number of searches for the keyword would then be determined on the basis of this latent construct, with some stochasticity in the outcome. In Section IV in the Web Appendix, we develop a model to estimate the latent popularity score for a keyword and find that the results on the effect of popularity on click behavior are qualitatively the same.

Ghose, Anindya and Sha Yang (2009), "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, 55 (10), 1605–1622.

Goldfarb, Avi and Catherine Tucker (2011), "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science*, 57 (3), 458–70.

Howard, John A. and Jagdish N. Sheth (1969), *The Theory of Buyer Behavior*. New York: John Wiley & Sons.

Jerath, Kinshuk, Liye Ma, Young-Hoon Park, and Kannan Srinivasan (2011), "A 'Position Paradox' in Sponsored Search Auctions," *Marketing Science*, 30 (4), 612–27.

——— and Amin Sayedi (2012), "Exclusive Display in Sponsored Search Advertising," working paper, Columbia University.

Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, and Jerry Lohse (2004), "On the Depth and Dynamics of World Wide Web Shopping Behavior," *Management Science*, 50 (3), 299–308.

Joo, Mingyu, Kenneth C. Wilbur, Bo Cowgill, and Yi Zhu (2014), "Television Advertising and Online Search," *Management Science*, 60 (1), 56–73.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probablistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–90.

Katona, Zsolt and Miklos Sarvary (2010), "The Race for Sponsored Links: Bidding Patterns for Search Advertising," *Marketing Science*, 29 (2), 199–215.

Lambrecht, Anja, Katya Seim, and Catherine Tucker (2011), "Stuck in the Adoption Funnel: The Effect of Interruptions in the Adoption Process on Usage," *Marketing Science*, 30 (2), 355–67.

——— and Catherine Tucker (2013), "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, 50 (October), 561–76.

Moe, Wendy W. (2003), "Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream," *Journal of Consumer Psychology*, 13 (1/2), 29–40.

Rutz, Oliver and Randolph E. Bucklin (2011), "From Generic to Branded: A Model of Spillover in Paid Search Advertising," *Journal of Marketing Research*, 48 (February), 87–102.

———, ———, and Garrett P. Sonnier (2012), "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising," *Journal of Marketing Research*, 49 (June), 306–319.

——— and Michael Trusov (2011), "Zooming In on Paid Search Ads: A Consumer-Level Model Calibrated on Aggregated Data," *Marketing Science*, 30 (5), 789–800.

———, ———, and Randolph E. Bucklin (2011), "Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?" *Marketing Science*, 30 (4), 646–65.

Sayedi, Amin, Kinshuk Jerath, and Kannan Srinivasan (2014), "Competitive Poaching in Sponsored Search Advertising and Its Strategic Impact on Traditional Advertising," *Marketing Science*, (published electronically March 24), [DOI: 10.1287/mksc.2013.0838].

Varian, Hal R. (2007), "Position Auctions," *International Journal of Industrial Organization*, 25 (6), 1163–78.

Yang, Sha and Anindya Ghose (2010), "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" *Marketing Science*, 29 (4), 602–623.

Yao, Song and Carl F. Mela (2011), "A Dynamic Model of Sponsored Search Advertising," *Marketing Science*, 30 (3), 447–68.