Advanced Database Marketing

Paid Search Advertising by Oliver J. Rutz and Randolph E. Bucklin

Oliver J. Rutz is Assistant Professor of Marketing at the Michael G. Foster School of Business at the University of Washington (e-mail: orutz@uw.edu)
Randolph E. Bucklin is Peter W. Mullin Professor, Anderson School of Management, University of California, Los Angeles (e-mail: rbucklin@anderson.ucla.edu).

Abstract

This chapter discusses paid search advertising, where Internet advertisers reach customers in the midst of their product search process. We address the direct and indirect impacts of paid search. The direct effect is the immediate impact of a paid search ad, whereas indirect effects are longer term. We review the history of paid search advertising and institutional issues such as the bidding process for ad placement. We then turn to a summary of empirical studies and models pertaining to direct effects, including the determinants of click-through rates and conversion. We next discuss indirect effects including the impact of generic search on future branded search, the impact of click-through visits on future visits, the value of search advertising as a customer acquisition channel, and search ad copy design. We conclude with a discussion of emerging topics such as the long tail in paid search, and the relationship between organic search and paid search click-throughs.

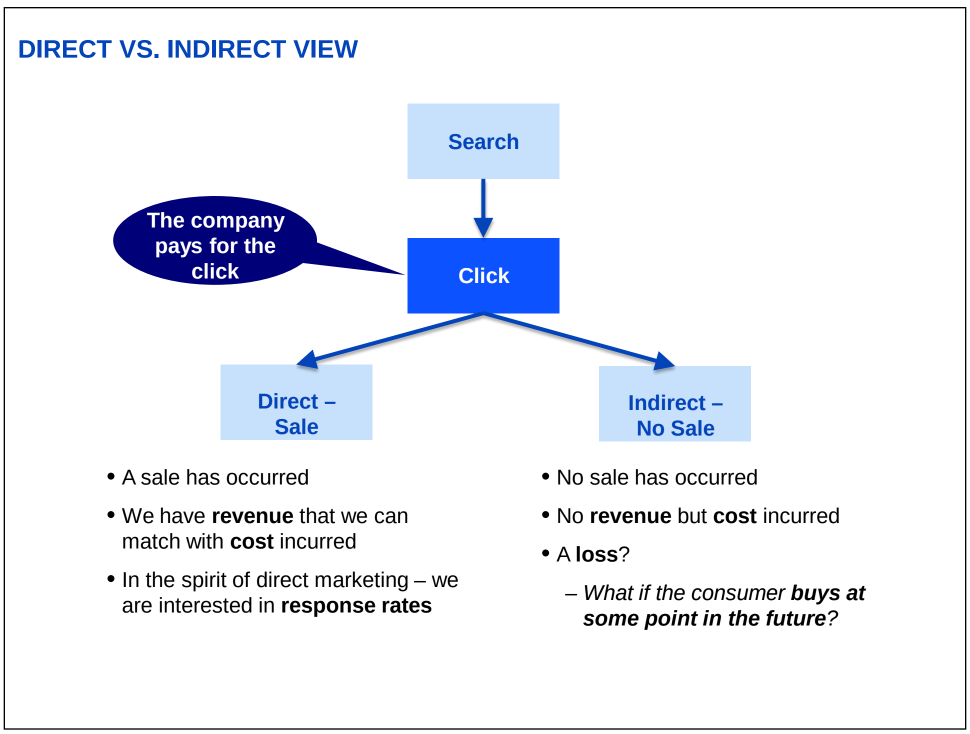
1. Introduction

The rise of the Internet – and with it the rise of Internet search engines such as Google, Yahoo! and Bing – has revolutionized the way information can be accessed by consumers. In the wake of this seismic shift in information search comes a new tool that allows marketers to intercept and target consumers during their search process – search engine marketing (SEM). Search engine marketing breaks down into two key parts, search engine optimization (SEO) and paid search advertising (paid search). The goal of SEO is to optimize a firm's position in the so-called organic listings provided by search engines, i.e., the search results based on the algorithms search engines use. On the other hand, paid search allows firms to buy a placement in the so-called sponsored or paid listings of the search engine results page (SERP). The purpose of this chapter is to discuss how paid search functions as an advertising vehicle and to show how firms can use data and analysis to better leverage their spending on it.

Why is this topic important? Internet advertising has been growing rapidly with a compound annual growth rate (CAGR) of 20 percent over the past 10 years in the United States. Total spending amounted to $31.7 billion in 2011 (PriceWatherhouseCoopers 2012). Of this overall spending on Internet advertising, 47 percent is now related to search engine marketing, followed by display (or banner) advertising with 22 percent.. It's noteworthy that search's share of the total has grown to its current 47 percent from 15 percent in 2002. Thus, not only has the total size of the online advertising market grown dramatically, but search engine marketing has been able to triple its share.

Our discussion begins with a brief overview of the history of paid search which is important for understanding the underpinnings of how the serving of paid search ads works and how costs and positions are determined. After describing the data that are typically provided to advertisers by the search engines, we turn to its analysis. We consider this in two stages (please see Figure 1). First, we examine the productivity of paid search advertising using a short-run or direct marketing perspective. This is based on the notion that the advertising data on impressions, clicks, costs, and conversions can be used to assess – and at least partially optimize – a paid search campaign. Second, we examine the issues involved in assessing the long-run or indirect effects of paid search. This requires a broader perspective on the impact of paid search advertising and typically requires the analyst to conduct additional analysis and/or to gather additional information on customer behaviors and link that to the paid search campaign. The indirect effects of paid search include such factors as spillover to future searches, spillover to organic search, spillover to bookmarking and direct type-in of URLs, and spillover to future purchasing. Our chapter closes with a look at emerging developments in the paid search analysis and then concludes.

Figure 1: Direct vs. Indirect Effects of Paid Search



SEO versus Paid Search

Since this chapter will focus on paid search advertising, we provide just a brief overview of SEO before returning to paid search. SEO is based on understanding the ranking algorithms used by search engines and on adapting a firm's website so that it is classified as relevant and therefore listed high up on the organic portion of the SERP. SEO can be conceptualized as similar to a firm's brand equity, but taken from the perspective of the search engines versus end consumers (as with traditional brand equity). Firms with high "search equity" are deemed relevant for certain queries by the search engines and will be displayed high in the organic rankings. However, as is true with a firm's brand equity, managing SEO is a long-run process and search engines actively discourage short-term gamesmanship. These so called "black-hat" strategies can lead to the (at least temporary) removal of the offender's website from search engine listings. For example, both BMW and J.C. Penney suffered from this after they attempted to bolster organic rankings using link farms. (Link farms artificially inflate the number of sites which link to the firm's site, thereby making the site appear more relevant to the search engine.) From our perspective, managing SEO is not a topic yet suited for database marketing analysis. Nonetheless, we note that efforts devoted to improving a firm's organic search results can be seen as a complementary endeavor to paid search and one which often competes with it for scarce marketing resources.

Paid search advertising allows firms to appear on SERPs without the need to first be found relevant by a search algorithm. This means that firms with undeveloped SEO (e.g., firms new to the Web) or unsuccessful SEO to date can still obtain listings and exposure in response to certain user search queries, enabling them to reach this audience. Paid search ads are not displayed in the organic listings but in specifically marked sections on the SERP, e.g., Google marks this section as "Ads." Placement and labeling have changed over recent years, e.g., Google formerly called its paid search section

"Sponsored Results" and used to place paid search ads only to the right of the organic ads.

A Brief History

We start our discussion of how paid search works with a short history (without any claim to completeness). In 2000, Google started its Adwords program, although smaller, now defunct search engines, offered paid search programs as early as 1995. Initially, advertisers were billed by impressions, i.e., the number of searches. Next, the more appealing pay-per-click concept was introduced, where advertisers paid only for clicks on their search ads. In traditional advertising, available advertising space is limited and the value of the space can be easily assessed, for example the cost of a one-page ad in the New York Times. In paid search, each search provides the opportunity to place advertising, so the effective ad space is unbounded. This makes it hard from the search engine's perspective to correctly price a click based on the search query alone – some searches are much more valuable from an advertiser's perspective than others. The problem was solved with the use of real-time second-price auctions to determine the cost-per-click (CPC). In a second price auction, a firm pays slightly above the bid of the firm below and firms bid their willingness-to-pay.

For search engines, however, the second price auction by itself did not provide the optimal way to price its online real estate. The most valuable spot a search engine has to sell is the first position in the paid listings. A second price auction only ensures that the ad displayed in the first position has the highest CPC. Assuming different bidder valuations, in a second price auction the winner pays the second highest bid plus a small delta. In the case of paid search, this is one cent more than the second highest bid. For the search engine, revenue also depends on the number of clicks a paid search ad generates. One of Google's most successful business decisions was to abandon the pure second price auction mechanism and introduce an enhanced auction that takes into account the past performance of the paid search ad to determine its future ranking. Google uses a so-called quality score to represent the past performance and other ad related differences, such as campaign performance and fit of the ad to the search (for more details please see Google's own description of quality score). In this enhanced auction, the first position goes to the ad which will, in expectation, provide Google with the most revenue based on the number of clicks and the CPC. All major search engines followed Google's lead and switched to using an enhanced auction mechanism to determine the rankings of paid search ads. Unfortunately for advertisers these mechanisms have become a black box, as search engines keep their auction algorithms secret and provide descriptions in vague terms. (Based on our conversations with search engine managers, it seems highly unlikely that more information on these proprietary algorithms will be provided to advertisers anytime soon.) This limited information poses a challenge not only to managers but also to researchers who are seeking to understand the nature and effectiveness of paid search advertising (Yao and Mela 2009).

Though advertisers are not privy to the "black box" which determines their ad placement and exact CPC, the search engines have nonetheless made it relatively easy to set up a paid search advertising campaign. First, the advertiser selects a set of relevant keywords, i.e., search terms, which seem likely to be used by consumers interested in or shopping for the firm's product or service. Search engines also provide tools to aid advertisers in developing a set of keywords or what is also called a campaign. Second, bids (or the willingness to pay for clicks from searches involving a given keyword) are set, along with budgets for daily and monthly spending ceilings. Third, text ads – a headline, two lines of body, and a URL – are created. Fourth, the landing pages that are to be linked with the text ad via the URL are created or specified. Paid search allows keywords to be linked with customized ad copy and a customized landing page. However, most firms use similar ad copy and/or landing pages for groups of keywords. Naturally, all these decisions should be driven by profit implications. While it is easy to set up a campaign, it is not as straightforward to determine which keywords will perform well, how much to bid, what ad copy to use and what landing page design to implement. Fortunately, these questions can be addressed to a large extent through database analysis.

What Data are Available for Analysis?

Google and the other major search engines provide an extensive amount of data to their advertisers. These data are standardized and aggregated to the keyword-ad level. For each keyword-ad combination, the search engines provide daily information on the number of impressions (i.e., number of searches), the number of clicks, the average position of the ad and the cost incurred. (Advertisers can also obtain more detailed hourly information in some cases if desired.) Additionally, an advertiser can track, by itself or by using a third party provider, whether a paid search visitor to its site goes on to make a purchase on the site or take some other desired action. Thus, data available on keywords are aggregated across customers, but allow targeting customers by keyword and ad copy.

On the other hand, advertisers are not able to easily (if at all) collect additional data on the process that occurs before a searcher arrives at their site. Indeed, Google discourages backtracking of searches and efforts to scrape its SERP by threatening to suspend the offending advertiser's paid search campaign. Citing privacy concerns, search engines also do not provide searcher-level data to advertisers. And, unlike in consumer packaged goods where marketing research firms like Nielsen collect and share competitive data across firms, search engines do not provide information on competitive advertising. For example, firms cannot even obtain information on which other firms are listed on the SERP along with their own, nor do they see competitor bids or rankings.

2. A Short-term Perspective – Paid Search as a Direct Marketing Tool

Taking a short-term perspective, paid search can be conceptualized as a direct marketing tool. Traditionally, direct marketing focuses on eliciting a direct response using a targeted marketing action. The goal of direct marketing is to determine which marketing offers generate response and who should be targeted, i.e., which consumers are more likely to respond. An example of a traditional direct marketing campaign would be a special offer sent to a subset of customers. First, the firm has to determine what the offer should be, e.g., a special holiday item. Second, the firm has to determine to which of its existing customers the offer should be extended. In general, extending an offer is costly, e.g., the firm incurs mailing costs to send the offer, and as such the offer should only be extended to consumers who are likely to buy. Whether a consumer might respond to the offer is generally evaluated based on a test mailing, the customers' demographics or the customers' past response to other, similar, offers. Applying this direct marketing perspective to paid search, the performance of a paid search campaign can be evaluated based on click-through and conversion performance. Profit generated directly by paid search can be derived based on:

$$\text{Profit} = \text{Impressions} \times \text{CTR} \times (\text{CVR} \times \text{M} - \text{CPC})$$

where CTR is click-through rate, CVR is conversion rate, M is margin and CPC is cost-per-click.

In the short-term (or direct) view the focus is on whether a consumer who has clicked on a paid search ad buys directly following this click. Recall that paid search is a pay-per-click model: the firm has to pay for each click on its paid search ads and thus incurs costs. A direct view asks whether the resulting revenues that can be directly tied to these clicks, i.e., occurring in the session that was started by the click, outweigh the costs incurred for these clicks. Thus, measuring this direct performance is relatively simple given the data provided by the search engines. Potentially, a paid search click could be a starting point of a buying process and a consumer might come back several times after a paid search click before buying or buy through another channel, e.g., via phone. As such, the direct view only provides a partial perspective on the success of a paid search campaign and can function as a lower bound. As we will discuss in the next section, taking an indirect (or long-run) perspective requires a different modeling approach and additional information on behavior (i.e., additional data need to be collected) or insights into the firm's business model. Models to help evaluate the direct effect can be seen as a one-size-fits-all approach that can be easily implemented across different firms and industries. Models to investigate indirect effects, on the other hand, are far more customized and cannot as easily transferred across domains.

Direct evaluation can be done either at a campaign-level or at more granular units such as individual keywords. To illustrate the development of a direct approach we first introduce a real world paid search dataset for an anonymous national U.S. hotel chain. For discussion, we use data from April 2004 and we note that Google still reports data to advertisers using the same format. The campaign used 301 keywords, spent $5,106.74 on paid search and generated 14,302 clicks and 518 sales (in the case of the hotel chain, reservations for rooms). Taking a direct view at the campaign-level, the click-through rate (CTR) is 0.6% and the conversion rate 3.6% (click-to-sales ratio). From a direct marketing perspective, a conversion rate of 3.6% is very good. The average cost-per-sale is $9.86, which is well below expected revenue. Thus, from a campaign perspective, paid search appears to be profitable. However, the chain is spending money on 301 different keywords. Using an overall campaign evaluation, i.e., average cost-per-sale of $9.86, we do not know whether performance is also profitable for all of the keywords.

Table 1: Sample Statistics for a Hotel Campaign on Google

April 2004 – Google	
Keywords (in campaign)	301
Position (daily average)	6.0
Impressions (total)	2,281,023
Clicks (total)	14,302
Reservations (total)	518
Cost (total)	$5,106.74
Click-through-rate	0.6%
Conversion Rate	3.6%
Cost/Impression	$0.002
Cost-per-Click	$0.36
Cost-per-Reservation	$9.86

Because paid search data are reported at a keyword-level, it should possible to repeat this basic data analysis, i.e., calculating performance ratios such as CTR and conversion rate for specific keywords. However, for most campaigns, this is not trivial to do. While some keywords generate many searches, clicks and sales, most keywords do not. In the case of the hotel chain, 217 keywords (out of 301) did not generate any sales in April 2004, but did generate clicks for which the firm was charged by Google. If we were to rely on the same simple ratios as above, we would conclude that these 217 keywords are money losers – costs for clicks are incurred, but no sales have been generated. Additionally, many of the 84 keywords that generated sales had very few searches and very few clicks. From a statistical perspective, making reliable statements with regards to their performance is problematic due to the sparse number of observations. For example, keyword #181 (we cannot reveal the actual keywords) generated 8 clicks and 0 sales, while keyword #221 generated 8 clicks and 1 sale. Relying on simple ratios, one would erroneously conclude that keyword #221 is a very good keyword with a conversion rate of 12.5%, while keyword #181 is unprofitable. From a statistical perspective, a reliable estimate of conversion rate for these two keywords cannot be calculated. It is also not possible to conclude that the conversion rate of keywords #181 and #221 is significantly different. Firms need a way to handle the data problems driven by the inherent sparseness present in paid search data on a keyword-level. In addition, advertisers might also be interested in learning what drives differences in keyword performance. This is difficult to assess using the simple ratios of CTR or conversion rate alone.

Taking a modeling approach to the analysis of paid search data permits analysts to address the sparseness problem and to assess the drivers of keyword-level performance. One potential driver of performance is the position in which the paid search ad is served. For example, ads in higher positions could be signaling a better fit of the firm's offering to the searcher's need. Also, the nature of the keyword itself could play a role. Shorter, more general search terms such as "hotels CA" may be more likely to be used earlier in the search process while longer, narrower keywords such as "hotels CA Anaheim Disneyworld" might be used later in the search. Keywords used later in the search might convert at higher rates because consumers are closer to the end of the purchase funnel. We now briefly describe a simple modeling approach which can be used to assess the drivers of keyword performance, such as position and keyword characteristics, while also providing better estimates of conversion in the presence of sparse data. The latter, of course, could be particularly helpful to firms pursuing a "long-tail" strategy in their paid search campaigns. Additionally, if ad-level covariates are available, they can be included in the analysis. Recall that search engines use a keyword-ad level to set up the data. We will discuss the use of ad-level information and the resulting complexities later.

One way to model paid search performance is to begin by casting consumer decisions here in a binary (i.e., 0/1) framework. Following a search and the serving of ad impressions, we might assume that a consumer clicks on a firm's paid search ad if the utility provided by the ad exceeds a certain threshold. Based on standard paid search data, a keyword-level click-thru model can be formulated as:

$$u_{kt}^{click} = X_{kt}^{click}\beta_k + \varepsilon_{kt}, \tag{1}$$

where u_{kt}^{click} represents the utility for a consumer using keyword *k* at time *t*, X_{kt}^{click} are keyword-specific covariates such as position or broad vs. narrow for keyword *k* at time *t*, β_k is the response sensitivity of consumers using keyword *k* and ε_{kt} is a logit error. Based on the assumption of an extreme value error, the probability of a click p_{kt}^{click} conditional on a search using keyword *k* at time *t* is given by

$$p_{kt}^{click} = \frac{exp\left(u_{kt}^{click}\right)}{1+exp\left(u_{kt}^{click}\right)}. \tag{2}$$

The model in equation (2) is keyword-centric and assumes that consumer response is homogeneous within keywords, i.e., all consumers using a certain keyword in their search will respond equally to, for example, the position of the ad. While a consumer-level model would be more appealing, paid search data does not contain consumer-level information to build such a model. Similar to the click-model given by (1) and (2), a model for the probability of conversion $p_{kt}^{conversion}$ conditional on a click on keyword *k* at time t can be specified. More details on how to estimate an integrated click and conversion model is provided, for example, in Ghose and Yang (2009) or in Rutz et al. (2012). In paid search, care has to be taken with respect to position as a covariate, as managerial feedback, missing competitive information and measurement error (search engines only report aggregate data) can bias the estimation of the effect of position. Ghose and Yang (2009), Rutz et al. (2012), and Narayanan and Kalyanam (2011) provide different approaches to address this potential bias.

A key finding from the estimation of keyword-level models on various data sets is that CTR and conversion rates do differ significantly across keywords. This means that advertisers who manage paid search at the campaign level are likely to miss substantial opportunities to improve performance. For example, Rutz et al. (2012) show how estimated conversion rates can be used to determine precisely which keywords are profitable or not. Based on the ability to calculate profitability at the keyword-level, they show in a hold-out sample that a new selection of keywords determined by the model can outperform a campaign-level evaluation. The model also performs better than using simple analytical strategies based on observed ratios and CTRs.

A second key finding from models is that keyword-level measures such as position and semantic keyword characteristics improve the assessment of CTR and conversion rates. Position is a particularly important and interesting measure. An ongoing debate among paid search practitioners involves the extent to which conversion rates vary by keyword position, if they vary at all. In a 2009 blog post, Google chief economist Hal Varian argued that conversion rates do not vary much by position (Friedman 2009). While some evidence from the practitioner community suggests higher positions raise conversion rates (Brooks 2004), other evidence supports Varian's contention (Ballard 2011; van Wagner 2010). The empirical evidence is also mixed from the academic marketing literature, based on papers that have measured the effect of position on conversion. Ghose and Yang (2009) and Rutz et al. (2012) find that higher position yields higher conversion rates. Narayanan and Kalyanam (2011) report that the first position usually performs better than others, but their findings are mixed for lower positions. The argument for higher position to yield higher conversion rates stems from a potential association between position and "quality" or "trust" perceptions (Ghose and Yang 2009).

Agarwal et al. (2011) take a different approach to the study of position effects and use data from a field experiment with randomized bidding. They focus on the top seven positions and find conversion rates *increase* at lower positions. Their argument for higher position to yield lower conversion rates stems from the notion that buyers higher up the purchase funnel often click on ads in higher positions without purchasing while buyers lower down the purchase funnel are more likely to visit lower positions. Another perspective on top positions is given by Jerath et al. (2011). Taking a game-theoretic approach, the authors find that a position paradox can exist where a superior firm may bid lower than an inferior firm and will obtain a position below, but still obtain more clicks. This is a profitable strategy for the superior firm if it can receive only slightly fewer clicks at the lower ranked position, but at a greatly reduced cost. Thus, position is an important element of paid search performance, but at this point it remains unclear whether a generalizable effect exists or whether the effect of position will depend on the product category, competition, and/or other factors.

Lastly, modeling results also show that keyword semantics are important. For example, inclusion of a brand name (either the firm's own brand or brands of products the firms sells) increases both click-through and conversion rate. Inclusion of other keywords semantics, e.g., a certain phrase, say "hotel", or other measures related to keywords, e.g., keyword length or keyword count, helps to improve estimates of click-through and conversion rates on a keyword-level. However, the effect of these semantic characteristics is highly situational and based on the currently available research no empirical generalizations have yet emerged. For example, including the firm's brand could potentially lead to cannibalization if the firm's SEO presence is very strong.

3. A Long-term Perspective – Indirect Effects of Paid Search

Using the direct or short-term perspective discussed above, every click that does not result in a sale for the firm would be categorized as a loss. A direct marketing analogy is that a catalog which fails to generate an order in a certain timeframe after it was mailed is seen as ineffective. This approach to measuring the performance of mail-order catalogs ignores the potential advertising value of the catalog itself. Thus, it might lead to suboptimal targeting if this indirect effect is not considered (Simester et al. 2006). A similar problem arises in paid search. First, a consumer could very well decide not to purchase during the visit triggered by the paid search click, but return to the site at a later point in time. On the return visit(s), opportunities exist to generate revenue for the firm. These revenues could include sales, but could also involve other sources such as advertising revenue from displaying banner ads on the site. Second, a consumer could decide to purchase using another channel, e.g., via phone or visiting a brick-and-mortar store. A critical question is how to attribute this indirect revenue back to the paid search click. As noted above, a standard model for the direct effect can be set up, but evaluating the indirect effect of paid search is a highly situational task which from our perspective requires a customized approach. We will discuss three case studies in how to define an indirect effect of interest to the advertiser and how this indirect effect can be captured by using a model-based approach.

Spillover from Generic Keywords to Branded Keywords

The first case study brings us back to the US hotel chain we discussed above. A prominent feature of the data is that certain categories of keywords show very different performance. One categorization scheme that shows a striking difference is branded versus generic keywords. A branded keyword includes the brand name of the advertiser, while a generic keyword does not. From a performance perspective, branded keywords seem to perform significantly better than generic keywords on all dimensions: click-through rates (13.68% vs. 0.26%), conversion rates (6.03% vs. 1.05%), CPC ($0.18 vs. $0.55) and cost-per-reservation ($2.94 vs. $55).

The performance of branded keywords seems “too good to be true.” Indeed, some consumers may use paid search ads as a short cut to get to the web site – as “white pages” so to speak. Thus, these consumers would not have been acquired by paid search in a direct marketing sense. To our knowledge, this issue has not yet been researched and may require survey-based data in conjunction with paid search data to investigate.

On the other hand, is the performance of the generic keywords truly as bad as it appears? Potentially, a generic search could generate awareness for the brand and lead to subsequent branded search. If this is the case, then some branded search volume might be attributable to generic search activity. Looking at the data over time one can observe patterns of modest spikes in generic activity followed by increases in branded search. Industry studies also have advanced the notion of "generic first, branded second." According to one study, 70% of searches begin with a generic keyword and, as the search process continues, it tends to become increasingly specific (Search Engine Watch 2006). For example, consider a consumer searching for a cruise vacation (Enquiro 2006). A user starts his search using the keyword "cruise," a generic keyword that returns a broad number of results. His next search narrows down the destination, as he searches for "Caribbean cruise." In the early stage, as the consumer can explore the space of options available to him, conversion rates are low. As the searches become more narrow, the consumer researches options in more detail. This particular consumer narrowed his search by reading third-party reviews on Panama cruises. After reading the reviews, his final search included a brand name, "Princess Panama cruise." This type of highly-targeted search had a much higher conversion rate, possibly 30%-40% according to the study.

Typical paid search data do not include the search history of individual consumers. This makes it impossible to examine spillover from generic to branded or effects relating to the search funnel at the consumer level. Fortunately, a test for spillover at the aggregate level can be performed, as demonstrated by Rutz and Bucklin (2011). In their approach, awareness of the relevance of the brand for the search purpose is conceptualized as an unmeasured, latent construct following the leaky-bucket (or ad stock) formulation of Nerlove and Arrow (1962). This construct permits past generic search activity to be linked with current branded search activity in a regression-type model. Exposure to brand related information due to generic search raises awareness and, in turn, greater awareness raises subsequent branded search activity. From the model estimates of this process, adjustments to the value of generic search can be made (for details on the model, estimation, and findings, please see Rutz and Bucklin 2011).

Spillover to Future Site Visits

In our second example, we focus on a source of firm revenue that does not stem from selling products, but from hosting advertising on its website. Next to selling products and services, many firms also host banner advertising on their sites. Unlike the pay-per-click model in paid search, banner advertising revenue is typically based on impressions, i.e., the number of times the advertisement is viewed. To increase the number of impressions served, firms can drive new users to the site and/or increase the number of page views for each visitor. In this situation, the productivity of paid search could be affected by the extent to which those users acquired through search return to the site for future visits and the extent to which they browse the site.

The company providing the data for this study is in the automotive business, selling new cars and trucks by linking buyers with local car dealerships. The firm also generates substantial advertising revenue by displaying ads from major car manufacturers on its website. Due in part to the long search process for new cars and the large number of keywords in the company's list (more than 15,000), trying to determine which keywords attract visitors who will ultimately buy was prone to significant error. On the other hand, because buying a car is a lengthy process, the company knew that some consumers visit the site many times following their initial visit via paid search. If there are differences across keywords in the propensity to attract such repeat visitors, this could be factored into keyword selection.

Visitors can access a website either by typing in the URL or using a saved bookmark (direct) or by clicking on either an organic or a paid search result (indirect). (Access via click on a banner ad is also possible; the firm in question was not using banner ads on other sites). The firm tracked daily counts of how many visitors were sourced directly versus indirectly (through either paid or organic search). This enables one to set up a model in which current direct visits are modeled as a function of past direct and indirect visits. The idea is to use the model to gauge the proportion of direct traffic that is related to past paid search activity, thus determining aggregate spillover. But the firm also has visits recorded at the keyword level. Can the model be extended to link subsequent direct visits to paid search visits by keyword? Doing so poses the challenge of the so-called “small n, large p” problem: a limited time series (in this study example, 60 days), but many thousands of predictors (keywords).

Rutz, Trusov and Bucklin (2011) show how to address this modeling issue and report a series of findings from the automotive website study. The results show that paid search visitors do return to the site, generating additional advertising revenue over and above that from the initial paid search visit. They also find that keywords differ in their ability to generate such return traffic. For the keywords which are significant in generating return visits (599 out of 3186 examined), the average number of return visits is 3.3 per click. Moreover, the authors also find that semantic characteristics of the keywords can be used to shed light on these differences. The best keywords for generating repeat visitors include the firm’s brand name, car brand names, general terms relating to search (e.g., “search”, “information”, comparison”) and general terms related to web use (e.g., “online”, “web”). Keywords that are very specific (“BMW 325i sports package”), or include general terms related to price and general terms related to used cars, have a lower propensity to generate return visits. In sum, the findings indicate that broad keywords appear to be better than narrow keywords for capturing consumers who will come back to the site.

Paid Search and Customer Lifetime Value

Our third example takes another view on paid search – how does paid search affect customer lifetime value (CLV)? So far, we have looked at the direct effect and specific indirect effects in scenarios where consumers purchase from the firm on a one-shot basis. For firms that strive to achieve a relationship with their customers over time it is of interest to evaluate the performance of paid search from a CLV perspective. Chan et al. (2011) investigate this for a small U.S. B2B firm in the biomedical and chemical lab supplies business. The firm sells both online and through offline channels. While traditional word-of-mouth was historically used to reach new customers, the firm had recently started to use paid search to acquire new customers. The authors assembled a dataset that allows lifetime tracking of customers that were acquired by paid search, i.e., the customers’ initial visits to the firm’s website came via paid search. The key question is whether lifetime customer profits exceed the initial acquisition cost. Note that the acquisition cost in paid search is not only the CPC that was spent on the customer’s click but also money spent on paid search that did not result in sales. For example, out of 100 paid search visitors (100 clicks), only one visitor becomes a customer by purchasing. If CPC is $0.50, acquisition cost for this customer is $50, not $0.50 as one could erroneously conclude.

In their study, Chan et al. find that customers acquired through Google have a higher lifetime value than customers acquired through other channels (CLV $1,332 vs. CLV $1,028). They also find that it is important to take into account offline purchases as customers who purchased offline after finding the company through paid search have a higher CLV ($1,637 vs. 1,226) when compared to customers who only bought online after a paid search click. In other words, for this firm, the “Google” customer dominates the “non-Google” customer from a CLV perspective. The authors also look at the value of customer acquisition, i.e., how does CLV compare when adjusted by customer acquisition cost? Correctly accounting for offline purchases as well as taking the customer lifetime perspective, each customer acquired from Google nets the company $1,280 on a lifetime basis. Thus, using paid search as a customer acquisition tool is highly profitable. In a final analysis, the authors consider the recent rise in CPC for paid search. They calculate a break-even CPC of $13.56, significantly greater than the current CPC of $0.80.

4. Beyond Keywords

Up to this point our focus has been on approaches that use keyword-level covariates, such as position or keyword semantic characteristics, to investigate differences in keyword-level performance, either from a direct or an indirect perspective. However, while a consumer searches by keyword, he or she ultimately clicks on the firm’s small text advertisement. These ads, unlike traditional ads used in most marketing campaigns, are entirely text – a paid search ad consists of a headline, two lines of text and a URL. Because the ads consist only of text, it may be possible to mine it to develop new predictors that could help understand paid search performance. These could be included in simple models such as the direct model given in equations (1) and (2).

We discuss text mining in paid search following the approach developed by Rutz and Trusov (2011) for a dataset in the ringtone space. The approach begins by defining linguistic characteristics and design elements used in paid search advertising by firms competing in the ringtone space. A particular advertiser may have specific creative skills or favor certain approaches to ad design which could differ from others. Thus, they propose to analyze ads created by many different firms in the ringtone space. The question of how to collect these ads is addressed using a novel procedure which expands the list of possible competitors, and therefore improves the representativeness of the ads sampled. The procedure uses the Google API for Adwords Sandbox tool: for each keyword from the firm's campaign, e.g., "ringtones", the Adwords Sandbox tool is used to produce a list of a certain number (in their example 100) of related terms which include, for example, "free ringtones", "poly ringtones", "sony ericsson ringtones" and "ringtones no subscription." Based on the resulting list of relevant keywords, Google is used to search for each of the identified keywords and collect the displayed paid search ads. The resulting unique ads (11,356 in this case) had 5,843 unique headers and 7,211 unique bodies. The longest ad contained 117 characters, the longest header was 35 characters long, and the longest body was 84 characters long. The longest body in their sample had 16 words: *"Send System Of A Down Ringtones To Your Cell & Get 10 Bonus Tones Now!"* and the shortest one had 3: *"Ultrasonic Ringtones Complimentary."* The longest headers contained five words, e.g., *"Find Hot Hip Hop Ringers"*, while the shortest headers contained only one, e.g., *"Mosquitotone."*

Although this descriptive exploration was done in a fully automated fashion, their proposed ad content analysis mainly relied on human intelligence. Based on unique headers and bodies, 11 distinct features could be identified. Some of the features occur relatively infrequently in the corpus. For example, *Price* and *Discount/Promotion* appear in less than one percent of headers (44 instances). On the upper end, approximately 64% (3726 instances) of headers contain the *Tone Identifier* feature or list a specific song or artist. A key difference between header and body is that it is typical to have just one, or, in some cases, two features listed in the ad header while the ad body typically includes multiple features. For example, the following ad body includes three features – Call for Action, Artist and Promotion: *"Send Metallica Ringtones To Your Cell & Get 10 Bonus Tones Now!"*

Next, an *Attention* $\rightarrow$ *Desire* $\rightarrow$ *Action* framework was used to investigate how the different textual features can be combined to create a successful ad. When it comes to generating *attention* two types of stimuli on the SERP may attract the consumer's gaze. The first one is low level stimuli, such as the ad location and visual characteristics (e.g., Itti and Koch 2000; Pan et al. 2007; Van der Lans et al. 2008). The second one is high level stimuli which include "higher order scene structure, semantics, context or task-related factors" (e.g., Cerf et al. 2008). Applying this framework to textual ads, candidates for measuring differences in low stimuli across ads are the density of an ad (number of words/characters in the rectangular area occupied by the ad) and the brightness of the ad (number of words/characters rendered using "Bold" font). Also, since a larger font is used for ad headers, the length of the header (measured in number of words or number characters) may have an additional effect on attracting attention.

While the main purpose of a headline is to capture attention and generate interest, the main function of the ad body is to create *desire* for the product and "to create real conviction in a product's superiority to competitors" (Vestergaard and Schroder 1985). Google's recommendation is to "tell that audience exactly what you have to offer." After investigating thousands of ad copies, virtually no cases of superiority claims could be found in the ringtone space. Thus, it appears that most advertisers do follow Google's advice to stay very specific and to list product/service/phone/media-format features in the hopes that their ad will be seen by a customer with matching interests. As with the header, it can be expected that feature(s) match with a reader's interest revealed through a search query translates into higher likelihood of perceiving an ad as relevant to the reader.

With respect to action, Vestergaard and Schroder (1985) state that as much as "'Buy X' is the most direct exhortation one can think of, [but] it is rare." Indeed, out of 7,211 unique ad bodies, only five invite a reader to *buy* a product or a subscription. The rest of the advertisers use different expressions to *get action* from a potential customer. Overall, in a ringtone corpus a call-for-action appears only in 8.2% of cases. This seems particularly surprising given that Google explicitly recommends to "include a call-to-action in your copy that tells users what you expect them to do after clicking your ad." Given the very competitive market of ringtone advertisers, the low occurrence of a call for action in the ad body is unlikely to be an oversight, but perhaps reflects specificity of the product.

The authors use the textual feature in a model of click-through to test their predictions with regards to performance of different features. While many of the textual features have the predicted effects and help to explain differences in CTR across keywords, some features are not significant in explaining click-through behavior (for details, see Rutz and Trusov 2011). However, the basic text mining framework proposed by Rutz and Trusov (2011) can be easily applied and is valid across categories.

5. Emerging Topics

In sections 1-4 we have covered the basics of paid search and set forth how to use modeling to investigate a paid search campaign's performance. In this final section we discuss emerging topics for which research is still in its infancy. We expect that more attention will be paid to these topics as the management of paid search advertising becomes better understood.

The Long Tail in Paid Search

The idea of a long tail in paid search is straightforward. Frequently searched keywords are generally expensive, as many firms bid competitively in the auction, increasing CPC and thus lowering profits from a sale. But literally millions of keywords exist which are hardly searched at all. CPC on these long-tail keywords is lower due to less competition. If keywords can be identified which are cheap, but will result in clicks and conversions, they could be "diamonds in the rough." To investigate this, we have examined multiple paid search datasets. We find that across these datasets, about 20% of keywords account for roughly 80% of all impressions and clicks, while 20% of keywords account for 80-90% of the conversions. Thus, as discussed above, it is crucial to be able to manage the major keywords (the top 20%) correctly using data- and model-driven approaches.

Given this empirical regularity, is it worthwhile for firms to expend the effort needed to manage a long-tail in a keyword list? Using the same three data sets, we have examined the relationship between sales revenue per ad dollar spent and the popularity of the keyword in the list. The analysis indicates that long-tail keywords actually show lower direct profitability than the popular keywords. We have also investigated how well long-tail keywords generate future website visits when compared to popular keywords and find that long-tail keywords underperform major keywords in this respect. Our findings are not conclusive (and could be driven by poor selection of long-tail keywords and/or not weeding out the ineffective ones). Nonetheless, caution should be used when focusing on the long-tail, as early indications are that "diamonds in the rough" appear to be hard to find. Even when they can be found, it is important to note that they will be very low volume keywords. Thus, managers will need to weigh the potential gains from high ROI but low volume keywords against the costs of managing much larger keyword lists.

Paid Search vs. Organic Search

The SERP displays two types of results – organic search results and paid search advertising. As discussed in the introduction, different factors drive the position of listings in the organic vs. the paid search section of the results page. Because each organic listing is "free," it is of interest to consider whether organic search aids or hinders paid search and vice versa. Yang and Ghose (2010) investigate this relationship by modeling the effect of an organic listing on a paid listing. More specifically, they investigate whether the presence of an organic listing for the firm has a positive, negative or neutral effect on a paid search ad that is shown on the same SERP. To do this, they use an integrated modeling approach that is again similar to equations (1) and (2), i.e., search volume, CTR, conversion rates, CPC and position (see Yang and Ghose 2010 for details). They find that higher organic click-troughs lead to higher paid click-troughs and vice versa. However, the effect is asymmetric, with organic to paid about 3-4 times stronger than paid to organic. Based on their empirical findings, they estimate that the interdependence increases firm profits by roughly five percent. They also validate this finding in a controlled field experiment. These findings suggest that the commonly held perspective that paid search cannibalizes the free organic results might be incorrect. Rather than cannibalization, the evidence suggests that there are positive synergies.

6. Conclusion

The effective management of paid search advertising provides both new opportunities and poses new challenges for database marketers. The data provided by search engines to advertisers is extensive and detailed. In this chapter, we began by discussing how this permits paid search campaigns to be assessed and managed at a granular level by taking a direct or short-term perspective. For example, individual keywords within an advertiser's campaign can be assessed by their performance in producing clicks and conversions in a cost effective manner. In cases where keyword performance data are sparse (e.g., for keywords in the long tail of a campaign's list), model-based analyses enable analysts to extend their evaluation to the entire list of keywords and improve results when compared to model-free approaches.

The effects of paid search advertising also extend beyond the direct effects that can be linked to a particular click through. Data-based analyses can be developed to assess some of these indirect or longer term effects. For example, if generic keyword searches spillover to influence subsequent branded keyword searches, adjustments to the returns from both types of keywords can be made by regression-type modeling approaches. Paid search can also produce additional site visits in the future, over and above those due to the initial click through as well as be responsible for higher levels of lifetime customer value. Model-based approaches can be used to estimate these effects, though they will typically need to be tailored to the company, category, and industry involved. Looking more broadly ahead, search advertising can also be tied to significant effects on offline sales channels, in addition to its strong effect on e-commerce sales (e.g., Dinner et al. 2011).

In addition to evaluating the profitability of individual keywords in both a direct and indirect manner, a series of additional opportunities await ambitious database marketers. For example, advertisers also can use paid search data and models to assess the effects of different text ad content. This capability enables advertisers to further optimize the effectiveness of the spending on paid search ad campaigns. Research is also ongoing into emerging topics such as the productivity of the long-tail in paid search and the effect of paid search listings on the performance of organic listings. In sum, working in paid search provides a rich set of possibilities, some straightforward while others more nuanced, for database marketers to contribute to the productivity of online advertising in the years ahead.

References

Agarwal, A., K. Hosanagar, and M.D. Smith (2011), "Location, Location, Location: An Analysis of Profitability in Position in Online Advertising Markets," *Journal of Marketing Research*, forthcoming.

Ballard, M. (2011). Paid Search Conversion by Position.

Brooks, N. (2004). The Atlas Rank Report II: How Search Engine Rank Impacts Conversions. Atlas Institute, NY.

Cerf, M., E. Frady and C. Koch (2008), Using Semantic Content as Cues for Better Scanpath Prediction, *Proceedings of the symposium on Eye tracking Research & applications*.

Chan, T., C. Wu and Y. Xie (2011), "Measuring the Lifetime Value of Customers Acquired from Google Search Advertising," *Marketing Science*, 30(5), 837-850.

Dhar, V. and A. Ghose (2010), "Sponsored Search and Market Efficiency," *Information Systems Research*, 21(4), 760-772.

Dinner, I., H. van Heerde, and S. Neslin (2011), "Driving Online and Offline Sales: The Cross-Channel Effects of Digital versus Traditional Advertising," working paper, Tuck School of Business, Dartmouth College, Hanover, NH.

Enquiro (2006). Inside the Searcher's Mind, White Paper *at www.enquiro.com*.

Friedman, D. (2009). Conversion Rates Don't Vary Much by Ad Position. http://adwords.blogspot.com/2009/08/conversion-rates-dont-vary-much-with-ad.html.

Ghose, A. and S. Yang (2009), "An Empirical Analysis of Sponsored Search in Online Advertising," *Management Science*, 55(10), 1605-1622.

Goldfarb, A. and C. Tucker (2011), "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science*, 57(3), 458-470.

Itti, L., and C. Koch. (2000), A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention., *Vision Research*, 40 (10-12) 1489-1506.

Jerath, K., L. Ma, Y. Park and K. Srinivasan (2011), "A 'Position Paradox' in Sponsored Search Auctions," *Marketing Science*, 30(4), 612-628.

Narayanan, S. and K. Kalyanam (2011), “Measuring Position Effects in Search Advertising: A Regression Discontinuity Approach,” Working Paper, Stanford Graduate School of Business.

Nerlove, N. and K. Arrow (1962), “Optimal Advertising Policy Under Dynamic Conditions,” *Economica*, 29 (114), 129-142.

Pan, B., H. Hembrooke, T. Joachims, L. Lorigo, G. Gay and L. Granka (2007), In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *J. Computer-Mediated Communication*. 12(3) http://jcmc.indiana.edu/vol12/issue3/pan.html.

PricewaterhouseCoopers (2012), IAB Internet Advertising Revenue Report, http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_FY_2011.pdf.

Rutz, O. J. and R. E. Bucklin (2011), “From Generic to Branded: A Model of Spillover Dynamics in Paid Search Advertising,” *Journal of Marketing Research*, 48(1), 87-102.

Rutz, O. J. and M. Trusov (2011), “Zooming In on Paid Search Ads – An Individual-level Model Calibrated on Aggregated Data,” *Marketing Science*, 30(5), 789-800.

Rutz, O. J, M. Trusov and R. E. Bucklin (2011), “Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?,” *Marketing Science*, *30*(4), 646-665.

Rutz, O. J., R. E. Bucklin and G. P. Sonnier (2012), “A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising,” *Journal of Marketing Research*, 49(3), 306-319.

Search Engine Watch (2006). Delving Deep Inside the Searcher's Mind, *A Special Report from the Search Engine Strategies Conference*, August 2-5, 2004, San Jose, CA.

Simester, D. I., P. Sun and J. N. Tsitsiklis (2006), “Dynamic Catalog Mailing Policies,” *Management Science*, 52(5), 683-696.

Van der Lans, R., R. Pieters and M. Wedel (2008), Competitive Brand Salience, *Marketing Science*, 27(5), 922-931.

Van Wagner, M. (2010). PPC Mad Scientists Prove Google Right and Wrong. http://searchengineland.com/ppc-mad-scientists-prove-google-right%E2%80%A6-and-wrong-45236.

Vestergaard, T. and K. Schroder (1985), The Language of Advertising, Basil Blackwell Ltd.

Yang, S. and A. Ghose (2010), “Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?,” *Marketing Science*, 29(4), 585-601.

Yao, S. and C. Mela (2009), “Sponsored Search Auctions: Research Opportunities in Marketing,” in *Foundations and Trends in Marketing* 3(2), 75-126.

Yao, S. and C. Mela (2011), “A Dynamic Model of Sponsored Search Advertising,” *Marketing Science*, 30(3), 447-468.