Whole Page Optimization: How Page Elements Interact with the Position Auction

PAVEL METRIKOV, Northeastern University
FERNANDO DIAZ, Microsoft Research
SÉBASTIEN LAHAIE, Microsoft Research
JUSTIN RAO, Microsoft Research

We study the trade-off between layout elements of the search results page and revenue in the real-time sponsored search auction. Using data from a randomized experiment on a major search engine, we find that having images present among the search results tends to simultaneously raise the ad click-through rate and flatten the ad click curve, reducing the premium for occupying the top slot and thus impacting bidding incentives. Theoretical analysis shows that this type of change creates an ambiguous impact on revenue in equilibrium: a steeper curve with lower total click-through rate is preferable only if the expected revenue distribution is skewed enough towards the top bidder. Empirically, we show that this is a relatively rare phenomenon, and we also find that whole page satisfaction causally raises the click-through rate of the ad block. This means search engines have a short-run incentive to boost search result quality, not just a long-run incentive based on competition between providers.

Categories and Subject Descriptors: H.3.3 [**Information Search and Retrieval**]: Selection Process; J.4 [**Social and Behavioral Sciences**]: Economics

Additional Key Words and Phrases: whole page optimization; sponsored search; search ranking; page layout; page elements

1. INTRODUCTION

Web search is monetized through a real-time auction for advertising slots, a practice known as "sponsored search." There is a substantial literature on the generalized second price auction (GSP) used for slot allocation and pricing; see for instance [Edelman et al. 2007; Lahaie 2006; Varian 2007]. A standard assumption in the literature is that the incentives and payoffs present in the auction are not impacted by other elements of the search engine results page (SERP) on which the ads appear. This assumption is reasonable for relatively simple page layouts consisting of "ten blue links," which was the standard when the seminal GSP papers were published, but since then the SERP has become increasingly less standard because of the integration of novel page elements such as images, maps, shopping results, etc. [Arguello et al. 2009; Navalpakkam et al. 2013] and these page elements impact user attention [Diaz et al. 2013]. This interplay introduces strategic considerations for bidders and raises the specter of results manipulation by search engines.

Behavioral interactions between elements of the SERP and the ad unit impact the auction primarily through the advertising slots' click-through rate, collectively re-

Author's addresses: Pavel Metrikov, College of Computer and Information Science, Northeastern University, metpavel@ccs.neu.edu. Fernando Diaz, Sébastien Lahaie, and Justin Rao, Microsoft Research, New York City, {fdiaz,slahaie,justin.rao}@microsoft.com

EC'14, June 8–12, 2014, Stanford, CA, USA. ACM 978-1-4503-2565-3/12/06 ...$15.00.

http://dx.doi.org/10.1145/2600057.2602871

ferred to as the "click curve." The slope of the click curve determines the degree to which higher slots are superior goods and the level gives the fraction of clicks going to ads. Since a larger bid increases both the chance of getting a higher slot and the expected payment conditional on winning, the degree to which bidders should shade bids down from their true value depends on the slope of the click curve. All else equal, the flatter the curve the more one should shade because the slots are closer substitutes [Gomes and Sweeney 2012]. If the impact of new page elements on user behavior translates to a meaningful change in the click curve, and therefore bidding and revenue, then this has important implications for the competitive landscape because the auctioneer and the publisher are typically the same economic entity (namely the firm owning the search engine). The possibility of this sort of manipulation lies at the center of recent anti-trust litigation surrounding search monopoly in Europe [Miller and Scott 2014].

In this paper we theoretically model the auctioneer's optimization problem assuming that the click curve can be impacted by changing the algorithmic results, and empirically calibrate the trade-off between algorithmic result features and ad revenue using data from a ranking experiment on a major search engine. The experiment took a large set of queries for which images were thought to be relevant and, instead of using normal production settings, randomized (for a fraction of traffic) the location of the image unit into one of the first five slots in the web results section of the page or off the page entirely. This exogenous variation allows us to estimate the true click curve for each regime and thus the causal impact of image location.

We empirically establish that the click curve is indeed manipulable and doing so can have a large impact on revenue. The presence of images on the SERP tends to flatten the click curve—in particular, the top ad slot does not get as high a click-through rate (CTR) premium as it does in the ten blue link setting. One might be tempted to conclude that revenue maximization would involve widespread removal of images to increase differentiation and competition for ad slots, creating a strong tension between user experience and revenue. In fact, we find that while images reduce the importance of the top slot, the other slots benefit from images so much that the overall CTR of the ad-unit generally goes up, with the largest impact coming when the image is in slots two through five (that is, not in the first slot directly below the ads). This is consistent with eye-tracking studies and the cascade model of user attention, which point to overall page quality as a driver of clicks on all units of the page [Joachims et al. 2005; Kempe and Mahdian 2008]. However, removing images entirely—potentially at a cost to user experience—is optimal when most of the revenue comes from the top bidder. We use supplementary data from a mouse-tracking experiment and confirm the hypothesis that the initial point of attention is strongly influenced by the presence and location of images on the SERP.

The experimental data, which comes from a fraction of traffic over a short period of time, are insufficient on their own to understand the long-run impact of the presence of images, or units with similar visual features, because advertisers would react if a new policy was rolled out to 100% of traffic. We address the gap with theoretical analysis. We show that the revenue impact of a click curve can be decomposed according to how 'steep' the curve is and its overall CTR. To formally assess whether a click curve is steeper than another, we draw on the concept of *majorization*, traditionally used to compare income distributions. In a novel application of the concept, we use it here to compare how CTR is distributed across slots. Our equilibrium model shows that if the total CTR of the ad block is held fixed, then the steeper the click curve, the higher the revenue. A steep click curve shifts clicks towards the top slots, increasing the incentive to bid aggressively and funneling clicks to higher bidding advertisers. However, if the auctioneer induces a steeper click curve while also lowering the overall CTR, as we empirically observe for image removal, then the impact depends on the distribution of expected revenue across the ad slots.

We take these insights to the data and show that revenue maximization reduces to a choice between pushing an image out of the top slot (down the page) or suppressing entirely. It turns out that suppression is rarely an optimal policy because the premium on the top slot is only worth protecting if the expected revenue distribution is heavily skewed towards the top bidder. For most monetizable queries, this is not a likely circumstance as they tend to appeal to many advertisers. Brand queries, such as “nike” or “us airways” sometimes have the requisite skew but images are usually not relevant for such queries. Suppressing images when the expected revenue distribution is not sufficiently skewed significantly harms total returns and we show it also degrades user experience metrics. Pushing the image out of the top slot to a placement down the page can increase revenue by a considerable amount and this can often be done with minimal or zero loss to user experience.

So while the image placement choice can (and does) create a tension between SERP quality and ad performance, in practice the trade-off is nuanced, with ad performance and user experience often moving in the same direction. In terms of the competitive landscape, our findings imply that search engines do not in general have the incentive to degrade algorithmic results to boost revenue, even in the short-run, but this tension is typically present in at least some parts of the decision space. The positive interaction between user satisfaction and ad CTR means search engines have a short-run incentive to boost algorithmic quality in many cases. Competition, such as the current status quo of two major providers, would presumably limit manipulation in the parts of the decision space where user metrics and revenue move in opposite directions. Considered from this angle, our results highlight the importance of competition as well.

2. BACKGROUND

Work on whole page optimization naturally draws from both the sponsored search and web search literatures, which have traditionally existed as relatively separate communities; see Kempe and Mahdian [2008] for a notable exception. In web search, considerable effort has gone into modeling how a user goes about achieving their objectives. In sponsored search, effort has generally centered around mechanism design of the auction. Whole page optimization studies this interplay so we review the related literature from both fields here.

2.1. Web Search

Early click models of algorithmic search results factored CTRs into an attractiveness effect and a position effect [Dupret and Piwowarski 2008]. To go beyond the restrictive assumptions of this model, researchers introduced models that incorporated the attractiveness of other page elements [Carterette and Jones 2007]. The *cascade model*, named to reflect the idea that a user starts at the top of the page and scrolls down, assumes that the click-through rate of a document at a given position is dependent on documents in higher positions [Craswell et al. 2008]. The Dynamic Bayes Net model extends the cascade model by explicitly modeling relevance (as opposed to clicks) and allowing the user to select multiple relevant documents [Chapelle and Zhang 2009].

Past work motivates our hypothesis that novel page elements interact with click dynamics on the ad unit. The Partially Observable Markov model demonstrates that non-web (everything except the ten blue links) elements such as query suggestions affect user interaction on the SERP [Wang et al. 2010]. Analysis of mouse-tracking data also demonstrates that user interactions with the SERP are dependent on non-web verticals [Diaz et al. 2013]. Furthermore, controlled experiments have demonstrated the impact of the quality and relevance of non-web results on user satisfaction and task performance [Arguello and Capra 2012; McCay-Peet et al. 2012; Navalpakkam et al. 2013].

2.2. Sponsored Search

When a user enters a query into a search engine, an auction is run to select and display ads on the page alongside the search results. Let m be the number of available ad slots and let $n > m$ be the number of candidate ads. The sponsored search auction used by major search engines is known as the *generalized second-price auction* (GSP) [Edelman et al. 2007]. Its specification consists of ranking and payment rules.

Each ad i is associated with a bid b_i together with a quality score w_i determined by the search engine. Ads are ranked in decreasing order of their weighted bid $w_i b_i$. Throughout, we re-index the ads so that $w_1 b_1 \geq w_2 b_2 \geq \cdots \geq w_n b_n$. Therefore ad i is allocated to slot i, unless $i > m$ in which case the ad is not shown. In sponsored search an advertiser is charged only when a click is received, and the GSP uses a second-price payment rule: the ad is charged the lowest bid that would maintain its position. This implies advertiser i's bid must satisfy $w_i b_i \geq w_{i+1} b_{i+1}$ to maintain position i, which leads to a cost per click (CPC) of $w_{i+1} b_{i+1} / w_i$. In practice search engines also impose reserve prices and various relevance filters on the ads, and must also select an ad layout besides the ranking (e.g., how many ads to show at the top of the page versus the side).

Let c_{is} denote the expected click-through rate (CTR) of ad i in slot s. Following the sponsored search literature, we assume that CTRs factor into an *ad effect* e_i and a *position effect* x_s, so that $c_{is} = e_i x_s$. Search engines place substantial resources into estimating ad effects because they are key inputs into the ad weights used for ranking [Graepel et al. 2010; Richardson et al. 2007]. In this work the focus is on the position effects and their impact on revenue. The vector of position effects $x = (x_1, \ldots, x_m)$ is called a *click curve* and we assume it is monotonically decreasing: $x_1 \geq x_2 \geq \cdots \geq x_m$.

It is well known that the GSP is not a truthful auction, meaning that it is not optimal for advertisers to bid their true value per click (i.e, willingness to pay), denoted v_i for ad i. When there is more than one ad slot, bids are shaded down from actual values [Lahaie 2006]. The advertiser's bid choice depends on v_i, the relative position effects of the different slots, and the opponents' bids. The next section examines in detail the potential revenue impact of distorting a query's click curve taking into account potential advertiser reactions via their bids.

3. THEORETICAL MODEL

Before discussing our experimental results let us introduce our theoretical framework. The framework will guide the interpretation of the experimental results, allowing us to understand the impact of click curve distortion not just for the short-run, but also once bids have reached an equilibrium (i.e., steady state). The standard notion of equilibrium for the GSP is *symmetric equilibrium*, also known as *envy-free equilibrium* [Edelman et al. 2007; Varian 2007]. In a symmetric equilibrium, every ad's allocated slot maximizes the advertiser's utility (expected value minus price) holding the opponent bids fixed. Furthermore, weighted bids $w_i b_i$ are increasing in weighted values $w_i v_i$, so ads are in fact ranked by weighted value even though bids do not equal values. We refer to Varian [2007] for an exact definition and a full treatment of the equilibrium's various properties, and here will only mention the relevant ones for the purpose of revenue analysis.

3.1. Equilibrium Revenue

Symmetric equilibrium is a useful concept to reason about revenue because the set of symmetric equilibria form a lattice. In particular, for fixed advertiser values there are minimum and maximum equilibria, which thus provide lower and upper bounds on possible equilibrium revenue. We will consider the lower bound since it is the conservative estimate, but the results below are easy to adapt to the maximum equilibrium and the general insights are the same.

For simplicity we assume that each ad has a weight and ad effect of 1. Under the ranking and pricing rules of sponsored search, the revenue as a function of the click curve is then

$$R(x) = \sum_{s=1}^{m} x_s b_{s+1}.$$

Since the bids depend on the click curve itself, we instead derive an expression in terms of exogenous values. The lowest symmetric equilibrium bids are in fact given by a closed-form formula:[1]

$$x_s b_{s+1} = \sum_{t=s+1}^{m+1} v_t (x_{t-1} - x_t). \tag{1}$$

To evaluate the expected revenue we treat the advertiser values as random variables and assume they are drawn i.i.d. from a common distribution. Let V_t be the t-th highest value among n draws from the distribution. We have the following characterization of expected equilibrium revenue.

PROPOSITION 3.1. *The expected revenue in lowest symmetric equilibrium is given by*

$$\mathbf{E}[R(x)] = \sum_{t=1}^{m} x_t \mathbf{E}\left[tV_{t+1} - (t-1)V_t\right], \tag{2}$$

where the expectation is taken with respect to the advertisers' common value distribution.

PROOF. Summing (1) over all slots, we obtain

$$\begin{aligned} R(x) &= \sum_{s=1}^{m} \sum_{t=s+1}^{m+1} v_t (x_{t-1} - x_t) \\ &= \sum_{t=1}^{m+1} \sum_{s=1}^{t-1} v_t (x_{t-1} - x_t) \\ &= \sum_{t=1}^{m+1} (t-1) v_t (x_{t-1} - x_t) \\ &= \sum_{t=1}^{m} x_t (t v_{t+1} - (t-1) v_t). \end{aligned}$$

Treating values as random variables and taking the expectation of both sides completes the result. □

[1] For readers familiar with auction theory, an ad's lowest symmetric equilibrium bid is the VCG payment of the advertiser one slot above [Edelman et al. 2007; Varian 2007].

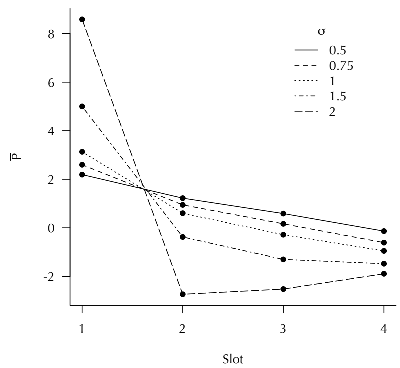
According to (2) we see that the expected revenue as a function of the click curve will depend on the monotonicity properties of the random variable

$$P_t = tV_{t+1} - (t-1)V_t. \tag{3}$$

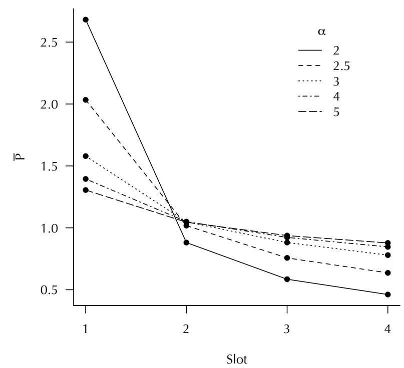
The expectation of P_t, denoted $\bar{P}_t$, can be easily evaluated numerically or even analytically for certain value distributions. The following result establishes monotonicity for a large class of distributions. The *hazard rate* of a distribution F is defined as $f(v)/(1-F(v))$, where f is the corresponding density for F. The hazard rate is a concept that often arises in auction theory and reliability theory.

PROPOSITION 3.2. *The expectation $\bar{P}_t$ is monotonically increasing with higher slots if the value distribution has an increasing hazard rate.*

PROOF. Let $D_t = t(V_t - V_{t+1})$, which is known as the t-th normalized *spacing*. Note that $P_t = V_t - D_t$. Barlow and Proschan [1966, Cor. 5.2] have shown that if the underlying distribution has an increasing hazard rate, then D_{t+1} stochastically dominates D_t. In particular, this implies that $\mathbf{E}[D_t] \leq \mathbf{E}[D_{t+1}]$. This combined with the fact that $\mathbf{E}[V_t] \geq \mathbf{E}[V_{t+1}]$ by definition yields the result. □



(a) Lognormal



(b) Pareto

Fig. 1. Monotonicity of $\bar{P}_t$ for the lognormal and Pareto distributions with four slots and five bidders. For the lognormal, the mean log is set to 0.5 and the standard deviation parameter σ is varied. For the Pareto, the location is set to 1 and the shape parameter α is varied. Each point is an average over 10,000 draws.

Many common distributions exhibit increasing hazard rate including the uniform, normal, exponential, and Weibull distributions. However, the lognormal and Pareto distributions, two natural distributions for advertiser value, do not satisfy this property. Note that Proposition 3.2 only provides a sufficient condition, and it is apparent from the proof that $\bar{P}_t$ may be increasing even with decreasing hazard rate.[2]

Figure 1 plots estimates of $\bar{P}_t$ for four slots and five bidders under the lognormal and Pareto distributions. The range of parameters was chosen taking into account estimates for value and bid distributions from the literature. For instance, Lahaie and Pennock [2007] report a lognormal value distribution with parameters $\mu = 0.35$ and $\sigma = 0.71$, which leads to a curve very similar to the one for $\sigma = 0.75$ in the figure. Both

[2]One plausible conjecture is that $\bar{P}_t$ is monotone increasing if the value distribution is *regular* in the sense of Myerson [1981]. This condition is more general than increasing hazard rate, but it would still not cover the practical case of the lognormal distribution.

the lognormal and Pareto plots show a pattern of steeper curves as tails get heavier (higher σ and lower α). The only non-monotone curve occurs for the lognormal at $\sigma = 2$, which has a very heavy tail. These plots suggest that non-monotonicity should not be common in practice.

3.2. Revenue Variation

We have found that an informative way to compare click curves is to consider how 'unequally' total CTR is allocated across slots in the curves, drawing on the concept of *majorization*.[3] See Marshall and Olkin [1979] for an in-depth treatment of the mathematical theory of majorization.

To understand the relevance of the concept, consider the click curves $(1, 0, 0)$ and $(1/3, 1/3, 1/3)$. Both have the same aggregate position effect, and in the first case there is effectively just one slot. The first click curve is intuitively more unequal than the second. Now suppose there are three ads. When there is just one effective slot (the first case), the GSP reduces to the classic second-price auction, which induces truthful bidding. Therefore the revenue will be the second-highest value per click. When all slots have equal position effects an advertiser receives the same utility no matter what its bid, and so it is an equilibrium for all to bid zero, generating zero revenue. Thus moving to a more 'equal' click curve in this extreme example eliminates all revenue.

For a click curve x, let its *norm* be $|x| = \sum_s x_s$, namely the aggregate CTR in the curve. A click curve y *majorizes* x, written $x \prec y$, if $|x| = |y|$ and y can be obtained from x by a series of transfers of CTR from lower slots to higher slots, provided that each transfer preserves the monotonicity of position effects. That is, we consider transfers of the form

$$(x_1, \ldots, x_i + \delta, \ldots, x_j - \delta, \ldots, x_m)$$

where $\delta > 0$, for sufficiently small δ to keep the decreasing order intact. An equivalent definition that is more convenient to verify is that $x \prec y$ if

$$\sum_{s=1}^{k} x_s \leq \sum_{s=1}^{k} y_s \qquad (k = 1, \ldots, m) \tag{4}$$

with the case of $k = m$ holding with equality (i.e., $|x| = |y|$). The majorization order can only compare two click curves if their norms are equal. It captures how 'unequally' CTR is allocated among slots, so for a proper comparison click curves must be normalized to a common norm. Also, majorization is just a partial order, so click curves x and y may be incomparable even if $|x| = |y|$.

Recall formula (2) for the expected revenue in equilibrium given a click curve. To assess the change in revenue when moving from curve x to x', we separate the change into two parts:

$$\begin{aligned} &\sum_{s=1}^{m} x'_s \bar{P}_s - \sum_{s=1}^{m} x_s \bar{P}_s \\ &= \underbrace{\sum_{s=1}^{m} x'_s \bar{P}_s - \frac{|x'|}{|x|} \sum_{s=1}^{m} x_s \bar{P}_s}_{\text{equivalent variation}} \quad + \quad \underbrace{\frac{|x'| - |x|}{|x|} \sum_{s=1}^{m} x_s \bar{P}_s.}_{\text{compensating variation}} \end{aligned}$$

[3]Majorization is very closely related to the concept of the Lorenz curve, originally developed to quantify inequality in income distribution [Lorenz 1905].

The *equivalent variation* captures the change in revenue due to shifting CTR between slots by comparing click curves x' and $(|x'|/|x|)\,x$, which have the same norm. The *compensating variation* captures the change in revenue due to the percent increase (or decrease) in aggregate CTR. With a slight abuse of terminology, we say that the equivalent variation is "increasing in the majorization order" if it is non-negative whenever $x \prec x'$ (after normalizing for comparison), namely the new click curve majorizes the old one.

THEOREM 3.3. *The equivalent variation is increasing in the majorization order if and only if $\bar{P}_s$ is monotone increasing with higher slots.*

PROOF. Let $\tilde{x} = (|x'|/|x|)\,x$. Assume w.l.o.g. that $\tilde{x} \prec x'$, so that x' can be obtained from $\tilde{x}$ by a series of transfers of CTR from lower to higher slots. If $\bar{P}_s$ is increasing with slots, then each transfer increases the equivalent variation. Conversely, suppose there is a slot t where $\bar{P}_t < \bar{P}_{t+1}$. Take two click curves x and x' that differ only in a small transfer of clicks from $t+1$ to t to obtain x' from x. We have that $x \prec x'$ by construction, but by assumption the transfer decreases the equivalent variation. $\square$

The simulations reported in Figure 1 establish that the equivalent variation should be increasing in the majorization order for lognormal or Pareto distributions of bidder values, except for implausible parameters leading to very heavy tails. In light of Proposition 3.2 we also immediately obtain the following.

COROLLARY 3.4. *The equivalent variation is increasing in the majorization order if the bidder's value distribution has an increasing hazard rate.*

To summarize, there are two possibly opposite impacts on revenue when moving from one click curve to another, taking into account changes in the advertisers' bids as they re-equilibrate. A 'flatter' or more 'equal' click curve will have less revenue than a 'steeper' curve that differentiates more between slots, assuming the two curves have the same aggregate CTR—this is the equivalent variation. However, this may be offset if the flatter curve has a higher aggregate CTR—this is the compensating variation. As we report in Section 4, in practice a flatter curve may have substantially more aggregate CTR, so the best choice of click curve from a revenue standpoint becomes an empirical question.

4. EMPIRICAL RESULTS

Our data comes from a controlled experiment on a major commercial search engine during the first quarter of 2013. The goal of the experiment was to explore different image locations for queries with features that made it likely images would be relevant to the intent of the user. From this set of queries, we only look at those with high commercial intent by restricting to query-instances with four ads present above the algorithmic results. It is widely known that a majority of search engine revenue comes from queries for which many ads are shown, which is not surprising as the presence of many ads indicates high demand from advertisers.

We ensure that each query-instance in our data occupied all three image location regimes we are most interested in: no image, image in the first slot, and images in slots 2–5. Each bucket thus has the same queries and the *exogenous* assignment to bucket—based on the randomization step in the experiment—allows us to safely infer that the differences in click curves we observe reflect a causal relationship between image position and click patterns, and are not due to a spurious correlation with features of the query. As an example of spurious correlation, it could be the case that for the set of queries with high commercial intent, the subset with images present tend to have higher overall CTR than those without, not due to the impact of the images but rather because these queries differ in other factors that directly influence CTR. (Indeed we will see this is the case.)

There are 352,628 query-instances in our primary experimental analysis, with 2605 unique queries. In additional analysis, we loosen the sampling restrictions, which expands the data to 634,547 query-instances in the experiment, with 8706 unique queries. Finally, when we look at production data to understand what would go wrong if we did not use a randomized experiment, there are over 30,000,000 query-instances.

4.1. Click Curve Estimates

We estimate the position effect curves for each experimental condition using a maximum likelihood approach. Suppose that we work with a sample from search engine logs containing N records, $i = 1, ..., N$, each one indicating whether a query q^i with an advertisement a^i displayed at mainline position s^i was clicked by a user ($c^i = 1$) or not ($c^i = 0$). As previously mentioned and consistent with our theoretical model, a common technique to model the probability of a click is to decompose it into the product of two factors: *quality effect* and *position effect* [Varian 2007]. Let $e_{a,q}$ be a measure of the quality of advertisement a with respect to the query q, and x_s be the position effect for any advertisement placed in slot s. Then the probability of click is modeled as:

$$\mathbf{Pr}\left(c = 1 \mid e_{a,q}, x_s\right) = e_{a,q} \cdot x_s. \tag{5}$$

We are interested in estimating the unknown sets of parameters, $\{e_{a,q}\}$ and particularly $\{x_s\}$, from the search engine logs. Their likelihood, or equivalently the probability of observed clicks for all records $i = 1, ..., N$ (assuming independence) given these parameters is:

$$L\left(\{e_{a,q}\}, \{x_s\}\right) = \prod_{i=1}^{N} \begin{cases} e_{a^i,q^i} \cdot x_{s^i} & \text{if } c^i = 1 \\ 1 - e_{a^i,q^i} \cdot x_{s^i} & \text{if } c^i = 0 \end{cases} \tag{6}$$

Then, in the parameter space, we apply coordinate ascent method with Newton-Raphson step size (see Appendix A for details) in order to obtain maximum likelihood estimates (MLE) for all position and quality effects simultaneously:

$$\{e^*_{a,q}\}, \{x^*_s\} = \underset{\{e_{a,q}\},\{x_s\}}{\arg\max}\, L\left(\{e_{a,q}\}, \{x_s\}\right). \tag{7}$$

To estimate the variance of position effects estimates $\{x^*_s\}$ we approximate the main diagonal values of inverted Fisher matrix:

$$\mathrm{Var}[x^*_s] \approx \left(-\frac{\partial^2 \ln L\left(\{e_{a,q}\}, \{x_s\}\right)}{\partial x_s^2}\right)^{-1}\Bigg|_{\{e^*_{a,q}\},\{x^*_s\}} \tag{8}$$

A half-width of reported 95% confidence intervals corresponds to $1.96\sqrt{\mathrm{Var}[x^*_s]}$.

Table I shows the estimated click curves and their partial sums after normalizing each to norm 1. We place the experimental conditions into three groups: *Image at 1*, *Image below 1* and *No Image*. The second group includes more than one image placement because we did not observe significant differences across these positions in the randomized experiment.[4] The partial sums show that the curves may be compared according to the majorization order:

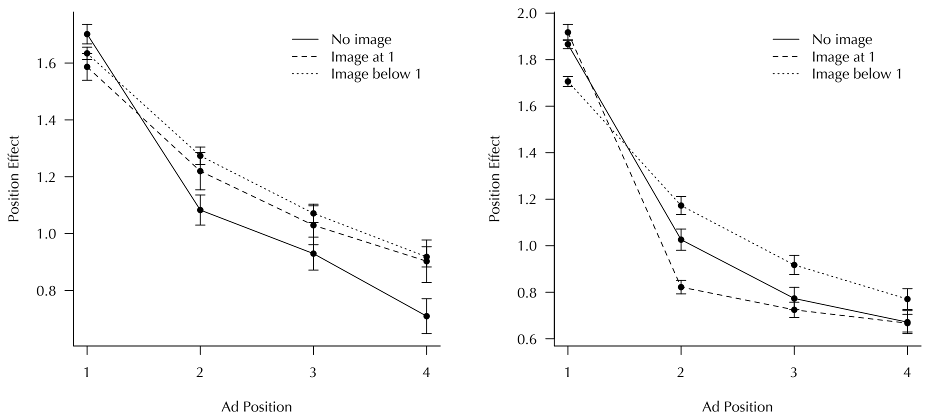
$$\text{Image at 1} \prec \text{Image below 1} \prec \text{No image.}$$

[4]The impact probably depends on screen resolution, which would be an important factor to consider in production serving.

Table I. Normalized click curves and their partial sums (in bold). Aggregate CTRs have also been normalized for easier comparison.

	1	2	3	4	Aggregate
No image	0.385 **0.385**	0.245 **0.629**	0.210 **0.840**	0.160 **1.000**	1.00
Image at 1	0.335 **0.335**	0.257 **0.592**	0.217 **0.809**	0.191 **1.000**	1.071
Image below 1	0.334 **0.334**	0.260 **0.594**	0.219 **0.813**	0.187 **1.000**	1.107

The ordering reflects how the curves relate to each other in terms of revenue when considering only the equivalent variation. On the other hand, the aggregate column shows that *Image below 1* has 10.7% higher norm than *No image*, while *Image at 1* has 7.1% higher norm. Therefore *Image below 1* dominates *Image at 1* in terms of both equivalent and compensating variations, so should lead to higher revenue in both the short- and long-term. However, there is a trade-off between the two effects when comparing SERPs with images removed versus pushed down below the first position.



(a) Randomized controlled trial

(b) Production data

Fig. 2. Position effect curve by image location. Panel A: Estimates taken from MLE estimation using the image-explore *experiment* restricted to queries appearing in all three experimental conditions; Panel B: Estimates taken from MLE estimation over *production data*

The click curves are presented graphically in Panel A of Figure 2 (for numerical values refer to Appendix B). The y-axis uses an arbitrary normalization factor. As in Table I, we restrict to queries that appear in each experimental condition to ensure that each bucket has the same queries ($n = 352,628$). The solid line shows the experimental condition where images are randomized off the page. In this case the average slope is steeper, with ad position 1 getting more clicks and ad positions 2–4 getting fewer clicks as compared to both curves with images present. The increased steepness is most pronounced between ad positions 1 and 2. Comparing *Image at 1* to *Image below 1*, it is easy to see the former is dominated and that these curves have similar slopes. The error bars give 95% confidence intervals—in aggregate the differences we quote are statistically significant beyond all conventional levels.

To improve statistical power, we relaxed the restriction that a query must occupy all three experimental conditions and instead required that a query occupy at least two experimental conditions. This doubles our sample size to 634,547. The new click curve estimates were nearly identical to those shown in Figure 2 Panel A (which is exactly what one would expect given the randomization) and the precision of the estimates improves considerably, with point-wise statistical significance now achieved for nearly all relevant comparisons. For the sake of brevity we do not include this figure in the paper but it is available from the authors.

In order to highlight the importance of randomized experiments to infer the causal effects, we plot click curves by image location using production data in Panel B of Figure 2. In Panel B each plotted line, instead of indicating an experimental condition, represents the production serving decision for one week of all US-located queries. It turns out that less than a quarter of queries with four ads at the top have images anywhere on the page; accordingly, these queries dominate the *No image* cluster. This makes sense, as many queries with commercial intent, such as "car insurance," have no relevant images to display. Since different queries constitute the data for each curve in Panel B, we would expect different click patterns due to differing characteristics of the queries outside of the direct impact of images. In fact this is precisely what we observe: *Image at 1* has the steepest curve and the highest CTR for the first ad slot. This is the precise opposite of the pattern we found using the randomized flight. The explanation is that Panel B shows correlations. Since one should only intervene into a system based on true causal effects (Panel A), the estimates shown in Panel B would produce a flawed plan of action.[5]

4.2. Revenue Impact

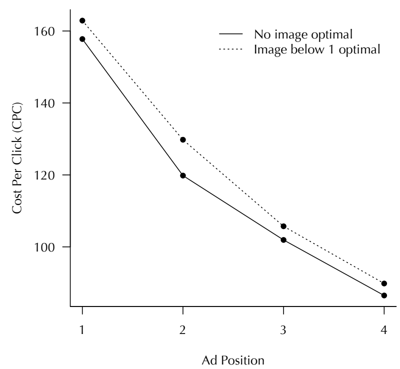
Returning to our theoretical analysis, calibrating the model with the click curve given in Figure 2 Panel A establishes that serving images at web position 1 is never optimal revenue-wise, highlighting a potential tension between user experience and revenue—we will study this in detail further on. The calibrated model also posits that comparing the steeper *No image* click curve and the flatter but higher overall CTR *Image below 1* curve depends on the distribution of advertiser values. Obtaining advertiser valuations directly is not possible because the GSP does not induce truthful bidding. Additionally, the experimental changes occurred on a small fraction of traffic for a limited period of time, meaning we would not expect an advertiser response. As such, we cannot assess full equilibrium behavior in our empirical analysis. We can, however, simulate an online decision of how to arrange page elements based on the standing bid distributions for that query-instance. If the changes to the page layout persisted, then bids would eventually respond and our theoretical model gives the expected direction of change. We note this where appropriate.

We simulate an online decision with a straightforward procedure: for a given query-instance, take the advertiser quality effects and CPCs (information that is actually available before serving the page) and determine which image placement is predicted to have higher revenue. Recall that according to our analysis in Section 4, showing an image will be optimal if the overall lift in clicks (compensating variation) overcomes the decrease in revenue that arises from flattening the click curve (equivalent variation).

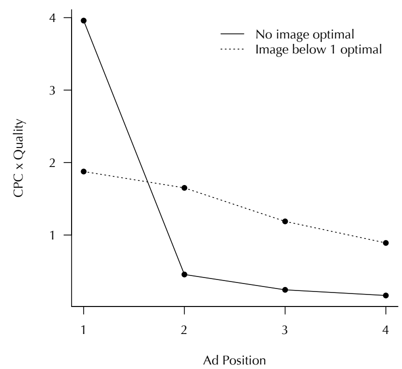
In the following figures each plot corresponds to the optimal serving decision using this procedure. The y-axes in all figures have been obscured by multiplying by a random scalar to protect business interests. The decision to suppress vs. push down is driven by both the CPC and CTR. Figure 3 Panel A shows that CPC distributions look

[5]As a note, the relevant question we address is, "For the queries that do have images, what is the impact of manipulating their location?" One could ask the question, "For queries on the borderline between including images or not, should one introduce images to improve ad performance?" Our experiment is unable to address this interesting extension.

quite similar across query-instances where suppression is revenue optimal and those where pushing down the page is optimal. The intuition is that it is not high CPC in the first slot that motivates us to suppress images, rather it is a combination of CPC and CTR. We refer to the quantity CPC $\times$ Quality as "expected revenue", normalized for position, as quality is the prediction of the ad's innate tendency to draw clicks as estimated via the maximum likelihood procedure. The CPC $\times$ Quality distributions given in Figure 3 Panel B show the required skew in expected revenue necessary to justify suppression—we only suppress when a huge fraction of revenue is expected to come from the top slot.



(a) CPC



(b) Expected revenue

Fig. 3. CPC distribution (Panel A) and expected revenue distribution (CPC *weighted by ad quality*, Panel B) for recommendations *No image* and *Image below 1*.

The decision to suppress requires a rare set of circumstances—without a big skew, optimizing for revenue keeps images on the page. It stands to reason that there is a set of queries where we would not suppress images in the short-run, but since a steeper click curve generally induces more aggressive bidding in equilibrium, we would want to in the long-run. Our decision to suppress is thus conservative and the resulting set is smaller than implied by long-run equilibrium. However, as we discuss below, suppression can negatively impact user experience, pushing back on the revenue incentive.

Table II shows the revenue impact by comparing our predicted optimal choice to the other image positions. We can do this because although we did not control placement directly in the experiment, by chance the optimal layout was chosen quite frequently due to the randomization at play. We have scaled the numbers to protect business interests by multiplying by a common scalar. The main diagonal (in bold) exceeds the sub-optimal placements by a margin of 7–10%, powerful evidence that optimizing page layout can have important revenue consequences. For both policies, we would expect bids to increase in the long-run—in the *No Image* case due to the steepness of the curve, in the *Image below 1* case due to the level shift in CTR—thus the short-run analysis tends to underestimate the long-run revenue impact.

Table II. Revenue Impact by Optimal Image Prediction.

Actual position	Predict No Image	Predict Slots 2–6
No image	**298.6**	236.4
1	249.8	255.5
2–6	261.4	**275.7**

Table II also shows that in many cases taking images off the page hurts revenue. In the second column we recommend putting the image in web slots 2–6. In this case, suppressing the image entirely would significantly damage revenue as compared to even putting it at the first slot below the ads—damaging the quality of algorithmic results hurts the tendency of the ad unit to draw clicks. Further on we will examine mouse-tracking data that indicates this is due to lower attention given to the relevant area of the ad unit; past work has found similar effects with eye-tracking data [Granka et al. 2004]. It turns out that empirically the interplay between algorithmic results and ads is nuanced; user experience and revenue metrics can move in the same or opposite directions.

One might be concerned that, once this sort of policy was rolled out, advertisers would have the incentive to manipulate the serving decision with their bids. We first note that our serving decision is based on expected revenue—given the second price bidding rule, this will not depend on one's bid (conditional on position) and thus limits the ability for this sort of manipulation. Moreover, the envy-free condition we typically use limits the ability of bidders to impact the CPC of their competitors without an immediate response.

4.3. User Experience

We now look at how user experience metrics varied across the revenue-optimal image location. The experiment was conducted with randomization at the query-instance level, meaning that within a search session, a user could be in both treatment and control. This means we can only look at "page-level" metrics as opposed to metrics recorded for a "search session." An important page metric is the satisfied click rate, namely the fraction of SERPs that lead to a click with a dwell time longer than thirty seconds [Wang et al. 2009]. We present this analysis in Table III (with values multiplied by a random scalar to protect business interests). When we recommend pushing the image down the page, we find that when the image is instead suppressed entirely, satisfied click rate significantly falls, as does the satisfied click rate on the ad unit. This is a lose-lose proposition—again showing that suppression is by no means a generally revenue enhancing strategy.

Table III. Satisfied CTR Impact by Optimal Image Prediction. Units multiplied by a random scalar to protect business interests.

(a) Web Satisfied CTR

Actual position	Predict No Image	Predict Slots 2–6
No image	6.957	8.831
1	7.915	9.188
2–6	7.076	9.461

(b) Ad Satisfied CTR

Actual position	Predict No Image	Predict Slots 2–6
No image	21.49	16.04
1	19.47	18.60
2–6	21.34	18.91

For suppression-optimal query instances, the web satisfied click rate is lower when images are suppressed or moved down the page rather than placed at the top spot, with the largest damage coming from suppression. This highlights the tension between user experience metrics and revenue in this case. Furthermore, pushing images down the page significantly lowers the likelihood they are clicked, consistent with past work on "position bias" [Craswell et al. 2008]. Obviously if the image is suppressed, it cannot be clicked. Since both revenue-optimal policies result in lower CTR on the image-unit, user experience consequences that are not captured in satisfied click rate should also be considered for the final serving decision.

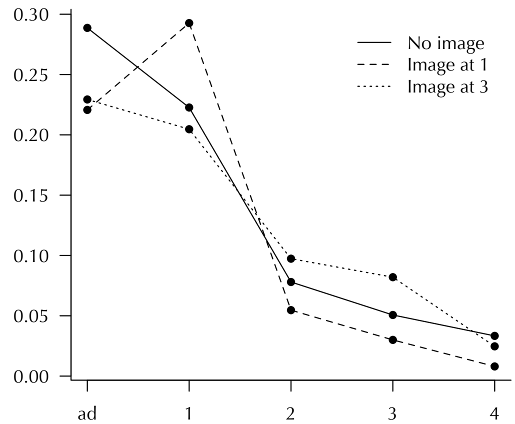


Fig. 4. Initial mouse fixation for three arrangements which have an ad at the top of the SERP. Distribution values do not include non-mainline probabilities and have been multiplied by a random scalar to protect business interests.

The user experience metrics show that there are instances where user experience and ad revenue move in the same direction. This is seen most clearly in our finding that suppressing images is often bad for both. Eye-tracking studies have suggested this effect in the past [Granka et al. 2004]. One potential reason for this is that, although cascading down the page is common, most people inspect results below the link they eventually click [Joachims et al. 2005], which helps explain why elements down the page can impact the CTR of the ad-unit at the top of the page. In the case of image suppression, there are fewer clicks on the ad unit overall, but this is recovered by the revenue gained from the top slot. User metrics, however, tend to drop in this case. Long-run optimization involves striking a balance between these competing incentives.

4.4. Understanding the User Attention Mechanism

Past work using eye-tracking technology suggests that the impact of page elements on the ad click curve occurs through altered visual attention. We believe that measures of visual salience may thus provide insight into the mechanisms driving our experimental results. We investigate this hypothesis with data gathered from user mouse movement for a fraction of non-randomized production traffic [Diaz et al. 2013]. Mouse movement has been shown to correlate with visual attention on SERPs [Rodden and Fu 2007]. Because these data were not collected as part of a randomized experiment, the behavioral patterns will be biased by user intent. However, before a page is rendered, a user is (very likely) unaware of the geometric arrangement of units on the page. Our strategy is therefore to measure the user's mouse fixation in the first few milliseconds after the SERP renders. This procedure should produce an unbiased measure of the initial point of focus.[6]

We present the cursor distribution over positions for three arrangements which include an ad-unit at the top of the page in Figure 4. Users are 20% more likely to initially focus at the top of the page when there is no image present. When an image is presented at the first position, immediately after the ad unit, users' visual attention

[6]Some bias may still be introduced if we believe the user's intent affects her initial mouse fixation.

initially fixates on this position, drawing attention from both higher and lower units. This effect is present when the image is at lower positions—attention is more likely to start down the page—but the aggregate impact is diminished for these arrangements. In fact, the ad unit engagement is slightly higher when the image is down the page as compared to when the image is at the first position.

As we discussed earlier, without a randomized experiment, the worry is unobserved factors can contaminate the analysis. In comparing Panels A and B of Figure 2 the culprit was query composition. Query composition and varying user intent are a concern here as well. Total click volume and other patterns will be biased by these factors, but our maintained assumption is that initial attention focus is not. If this assumption holds, at least approximately as is our view, then the mouse-tracking data provide strong support for the hypothesis that visual attention mediates that interplay between SERP features and the ad unit.

4.5. Query Clusters

We examined the top 100 queries (by revenue) in each policy decision. Queries in the suppress group tended to be brand terms, such as “nike shoes,” whereas queries in the down-the-page group tended to be broader product descriptions, such as “basketball shoes.” The intuition is that the highly skewed expected revenue distribution necessary to make suppression the optimal decision is unlikely to occur on broader queries because these queries are typically attractive to a wide range of advertisers. We saw the CPC distributions were similar across the optimal serving decision, which means in the suppression case the top bidder must very clickable relative to bidders 2–4. Since he is also paying a relatively high CPC, this means there must be a very high second bid, likely a competitor trying to scoop a valuable brand term. We also note that the differing user intent for brand queries is a likely reason that removing images on brand queries only has a muted negative impact on user experience, whereas removing images on more “categorical” queries significantly hurts user experience.

5. DISCUSSION AND CONCLUSION

Whole page optimization studies the interplay between SERP features and the incentives and behavior in the sponsored search auction. We have studied the case of images, but similar analysis could be applied to maps, shopping verticals, local results and so on. From an applied game theory perspective this interplay implies that bidding optimally is a complex task requiring knowledge of SERP features, such as the location of images and other visual elements. Standard analytics might not be sufficient to update one’s bid and we would thus expect market participants to profit from the use of screen scrapes and other novel data to help optimize a bidding strategy.

We theoretically show that if the search engine can manipulate the click curve with features of the algorithmic search results page, then it can alter equilibrium bidding behavior and expected revenue in the ad auction. The relevant questions are to what degree is this sort of manipulation possible and what are the returns to such manipulation. If manipulation is easy and profitable, then we would worry that search engines, especially a monopolist, might degrade results quality to boost revenue.

We empirically establish that the click curve is indeed manipulable and doing so can have a large impact on revenue. However, at least for the case of images, to increase clicks on the top slot the search engine must sacrifice the total click-through rate of the ad unit. This is consistent with past work that degrading an element of the search page can reduce the CTR and attention on other page elements. Theoretical analysis shows that this sort of change creates an ambiguous impact on revenue in equilibrium: a steeper curve with lower total CTR is preferable only if the expected revenue distribution is highly skewed to the top bidder. Empirically we find that the skew necessary to suppress entirely is a relatively rare phenomenon. Moreover, removing images on queries which lack this large skew significantly hurts revenue and is worse than even placing the image in the top slot. The implication is that search engines do not have a dominant incentive in this domain to degrade algorithmic results in order to boost revenue, even in the short-run. However, we also find that pushing images out of the top web results slot increases revenue almost universally, which can create a user experience versus revenue trade-off, although this is query dependent.

We can see now how the results inform the debate on competition in the search engine market. Principles from the economics literature assert that the degree to which we should guard against monopoly depends on how exerting market power impacts consumers. In this light, our findings paint a nuanced picture. On one hand, search quality and ad CTR often move in the same direction, meaning a monopolist would have an incentive to use page elements to boost algorithmic result quality even in the absence of competitive pressure. On the other hand, the choice becomes a true trade-off in other parts of the decision space. In the current marketplace with two main competitors, competition presumably restrains the degree to which search engines can (or should) degrade user experience to maximize revenue in cases where this tension exists.

ACKNOWLEDGMENTS

We would like to thank Nathan Chalmers, Liwei Chen, Jerel Frauenheim, Rukmini Iyer, Anand Oka, Nathan Powell, Luke Simon, Ravi Kiran Holur Vijay, and Daniel Waldinger for feedback and support.

A. NEWTON-RAPHSON METHOD FOR CLICK CURVE ESTIMATES

We perform Newton-Raphson approach to solve equation (7) by iteratively re-estimating parameters of interest until they converge:

$$\{e_{a,q}\},\{x_s\} := \{e_{a,q}\},\{x_s\} - H^{-1}\nabla_{\{e_{a,q}\},\{x_s\}} \ln L\left(\{e_{a,q}\},\{x_s\}\right)$$

Here H is a Hessian, i.e. a square matrix of second derivatives of $\ln L$ with respect to $\{e_{a,q}\},\{x_s\}$. In order to make computation tractable (H is huge in our experiments), we compute only the main diagonal elements of H, assuming zeroes everywhere else. The first and second derivatives of the logarithm of likelihood function can be expressed in a closed form and can be efficiently recomputed on each iteration of the coordinate accent:

$$\frac{\partial \ln L\left(\{e_{a,q}\},\{x_s\}\right)}{\partial x_s} = \sum_{\substack{i:c^i=1,\\ s^i=s}} \frac{1}{x_s} - \sum_{\substack{i:c^i=0,\\ s^i=s}} \frac{e_{a^i,q^i}}{1-e_{a^i,q^i}\cdot x_s}$$

$$\frac{\partial \ln L\left(\{e_{a,q}\},\{x_s\}\right)}{\partial e_{a,q}} = \sum_{\substack{i:c^i=1,\\ a^i=a,\\ q^i=q}} \frac{1}{e_{a,q}} - \sum_{\substack{i:c^i=0,\\ a^i=a,\\ q^i=q}} \frac{x_{s^i}}{1-e_{a,q}\cdot x_{s^i}}$$

$$\frac{\partial^2 \ln L\left(\{e_{a,q}\},\{x_s\}\right)}{\partial x_s^2} = -\sum_{\substack{i:c^i=1,\\ s^i=s}} \frac{1}{x_s^2} - \sum_{\substack{i:c^i=0,\\ s^i=s}} \left(\frac{e_{a^i,q^i}}{1-e_{a^i,q^i}\cdot x_s}\right)^2$$

$$\frac{\partial^2 \ln L\left(\{e_{a,q}\},\{x_s\}\right)}{\partial e_{a,q}^2} = -\sum_{\substack{i:c^i=1,\\ a^i=a,\\ q^i=q}} \frac{1}{e_{a,q}^2} - \sum_{\substack{i:c^i=0,\\ a^i=a,\\ q^i=q}} \left(\frac{x_{s^i}}{1-e_{a,q}\cdot x_{s^i}}\right)^2$$

B. POSITION EFFECT CURVES: NUMERICAL VALUES

Table IV. Position effect curve estimates by image location along with 95% confidence intervals.

(a) Randomized controlled trial

Ad Position	No Image	Image at Slot 1	Image at Slots 2–6
1	1.70 ± 0.03	1.59 ± 0.05	1.63 ± 0.02
2	1.08 ± 0.05	1.22 ± 0.07	1.27 ± 0.03
3	0.93 ± 0.06	1.03 ± 0.07	1.07 ± 0.03
4	0.71 ± 0.06	0.90 ± 0.07	0.92 ± 0.04

(b) Production data

Ad Position	No Image	Image at Slot 1	Image at Slots 2–6
1	1.87 ± 0.02	1.92 ± 0.03	1.71 ± 0.02
2	1.03 ± 0.05	0.82 ± 0.03	1.17 ± 0.04
3	0.77 ± 0.05	0.72 ± 0.03	0.92 ± 0.04
4	0.67 ± 0.05	0.67 ± 0.04	0.77 ± 0.04

REFERENCES

Jaime Arguello and Robert Capra. 2012. The effect of aggregated search coherence on search behavior. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*. 1293–1302.

Jaime Arguello, Fernando Diaz, Jamie Callan, and Jean-François Crespo. 2009. Sources of Evidence for Vertical Selection. In *Proceedings of the 32nd International ACM Conference on Research and Development in Information Retrieval (SIGIR)*. 315–322.

Richard E. Barlow and Frank Proschan. 1966. Inequalities for linear combinations of order statistics from restricted families. *Annals of Mathematical Statistics* 37 (1966), 1574–1592.

Ben Carterette and Rosie Jones. 2007. Evaluating Search Engines by Modeling the Relationship Between Relevance and Clicks. In *Neural Information Processing Systems (NIPS)*.

Olivier Chapelle and Ya Zhang. 2009. A dynamic Bayesian network click model for Web search ranking. In *Proceedings of the 18th International Conference on the World Wide Web (WWW)*.

Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *Proceedings of the International ACM Conference on Web Search and Data Mining (WSDM)*. 87–94.

Fernando Diaz, Ryen W. White, Georg Buscher, and Dan Liebling. 2013. Robust Models of Mouse Movement on Dynamic Web Search Results Pages. In *Proceedings of the 22nd ACM conference on Information and Knowledge Management (CIKM)*.

Georges E Dupret and Benjamin Piwowarski. 2008. A user browsing model to predict search engine click data from past observations.. In *Proceedings of the 31st International ACM Conference on Research and Development in Information Retrieval (SIGIR)*. 331–338.

Benjamin Edelman, 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 (2007), 242–259.

Renato Gomes and Kane Sweeney. 2012. Bayes-Nash equilibria of the generalized second-price auction. *Games and Economic Behavior* (2012).

Thore Graepel, Joaquin Quiñonero Candela, Thomas Borchert, and Ralf Herbrich. 2010. Web-Scale Bayesian Click-Through rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*. 13–20.

Laura A Granka, Thorsten Joachims, and Geri Gay. 2004. Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th International ACM Conference on Research and Development in Information Retrieval (SIGIR)*. 478–479.

Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th International ACM Conference on Research and Development in Information Retrieval (SIGIR)*. 154–161.

David Kempe and Mohammad Mahdian. 2008. A cascade model for externalities in sponsored search. In *Workshop on Internet and Network Economics (WINE)*. Springer, 585–596.

Sébastien Lahaie. 2006. An analysis of alternative slot auction designs for sponsored search. In *Proceedings of the 7th ACM conference on Electronic commerce*. 218–227.

Sébastien Lahaie and David M. Pennock. 2007. Revenue analysis of a family of ranking rules for keyword auctions. In *Proceedings of the 8th ACM conference on Electronic commerce*. 50–56.

M.O. Lorenz. 1905. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association* 9, 70 (June 1905), 209–219.

Albert W. Marshall and Ingram Olkin. 1979. *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York.

Lori McCay-Peet, Mounia Lalmas, and Vidhya Navalpakkam. 2012. On saliency, affect and focused attention. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (SIGCHI)*. 541–550.

Cain Miller and Mark Scott. 2014. Google Settles Its European Antitrust Case; Critics Remain. `www.nytimes.com/2014/02/06/technology/google-reaches-tentative-antitrust-settlement-with-european-union.html`. (2014). Accessed: 2014-04-10.

Roger B. Myerson. 1981. Optimal Auction Design. *Mathematics of Operations Research* 6, 1 (February 1981).

Vidhya Navalpakkam, LaDawn Jentzsch, Rory Sayres, Sujith Ravi, Amr Ahmed, and Alex Smola. 2013. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proceedings of the 22nd international conference on World Wide Web (WWW)*. 953–964.

Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting Clicks: Estimating the click-through rate for new ads. In *Proceedings of the 16th International World Wide Web Conference (WWW)*. 521–530.

Kerry Rodden and Xin Fu. 2007. Exploring How Mouse Movements Relate to Eye Movements on Web Search Results Pages. In *Proceedings of the SIGIR 2007 Workshop on Web Information Seeking and Interaction*, Kerry Rodden, Ian Ruthven, and Ryen W. White (Eds.).

Hal R. Varian. 2007. Position Auctions. *International Journal of Industrial Organization* 25 (2007), 1163–1178.

Kuansan Wang, Nikolas Gloy, and Xiaolong Li. 2010. Inferring search behaviors using partially observable Markov (POM) model. In *Proceedings of the third ACM International Conference on Web Search and Data Mining (WSDM)*. 211–220.

Kuansan Wang, Toby Walker, and Zijian Zheng. 2009. PSkip: estimating relevance ranking quality from Web search clickthrough data. In *Proceedings of the 15th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*.