The Good, the Bad, and the Random: An Eye-Tracking Study of Ad Quality in Web Search

Georg Buscher
DFKI
Knowledge Management Dept.
Kaiserslautern, 67663, Germany
georg.buscher@dfki.de

Susan Dumais
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
sdumais@microsoft.com

Edward Cutrell
Microsoft Research India
196/36 2nd Main, Sadashivnagar
Bangalore, 560 080, India
cutrell@microsoft.com

ABSTRACT

We investigate how people interact with Web search engine result pages using eye-tracking. While previous research has focused on the visual attention devoted to the 10 organic search results, this paper examines other components of contemporary search engines, such as ads and related searches. We systematically varied the type of task (informational or navigational), the quality of the ads (relevant or irrelevant to the query), and the sequence in which ads of different quality were presented. We measured the effects of these variables on the distribution of visual attention and on task performance. Our results show significant effects of each variable. The amount of visual attention that people devote to organic results depends on both task type and ad quality. The amount of visual attention that people devote to ads depends on their quality, but not the type of task. Interestingly, the sequence and predictability of ad quality is also an important factor in determining how much people attend to ads. When the quality of ads varied randomly from task to task, people paid little attention to the ads, even when they were good. These results further our understanding of how attention devoted to search results is influenced by other page elements, and how previous search experiences influence how people attend to the current page.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**] User/Machine Systems – *Human information processing, Human factors* .

General Terms

Design, Experimentation, Human Factors, Measurement.

Keywords

Gaze tracking, user study, search engine results pages

1. INTRODUCTION

In designing effective search systems, it is important to understand how people search and interact with the information presented on search engine result pages (SERPs). In this paper we use an eye-tracking study to increase our understanding of the processes that people use in examining result pages, and of variables that influence these processes.

Previous studies have used eye-tracking to understand how people attend to and interact with different elements of SERPs. This work has developed well-known terms to describe typical gaze distributions on SERPs, such as the "golden triangle" [12] or the "F-shaped pattern" [18]. Figure 1 shows an example of a characteristic heat map for a SERP. These studies tend to be fairly high-level, with qualitative descriptions aggregated across many different pages or tasks. Other researchers have taken a more controlled experimental approach and reported quantitative summaries of eye movements on SERPs, often explicitly controlling users' tasks. These studies characterized how visual attention is distributed on the 10 organic results, e.g., [6], [9], [10], [16]. However, all of today's major commercial search engines include additional elements on a SERP such as sponsored links at the top and on the right rail, related searches, graphical elements such as maps, illustrations, or other content. In this study we seek to understand how the visual attention devoted to organic results is influenced by these other page elements.

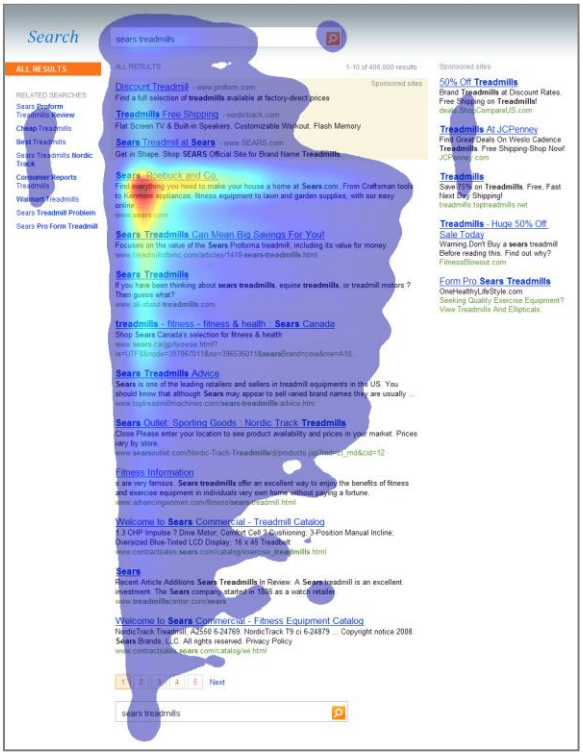

Figure 1: Gaze heat map on a search engine results page.

Sponsored links are an especially important component of the SERP since they form the main source of income for search engines. Depending on the search intent of the user, ads may provide valuable information and lead searchers directly to their

SIGIR'10, July 19–23, 2010, Geneva, Switzerland.

goal. In contrast, if ads are off-topic or simply not relevant to the immediate goal, they run the risk of annoying or distracting users, perhaps even impeding completion of their search.

The main goal of this paper is to study factors that might influence how users distribute their visual attention on different components on a SERP during Web search tasks. While we examine visual attention on most components typically present on a SERP, we are especially interested in sponsored links. Applying eye-tracking techniques, we determine basic differences in gaze distribution due to ad quality and task type. In addition we examine the effects of the sequence in which ads of good or poor quality are presented. Sequence effects reflect how prior search experiences influence behavior on the current search task.

After presenting an overview of related research, we describe the experimental design for our eye-tracking study. We then provide an analysis and a discussion of factors influencing the visual attention to SERP components, with special attention to sponsored links. We conclude with a summary of the implications of the results and some directions for future research.

2. RELATED WORK

2.1 Web Search Behavior in General

Several factors including the quality of the results and their presentation, the type of search task, and individual differences have been shown to influence search behaviors and success.

Search interactions are influenced by the quality of the search results, although the relationships are often weak when measured using total time or overall search success [22]. Similarly, Smith & Kantor [21] find that the search success is the same for both good and degraded systems, but that users alter their strategies depending on the quality of the results.

In the context of Web search, Broder [2] and Rose & Levinson [20] describe three general classes of user goals informational, navigational and resource or transactional. These different search goals lead to different search success, with users being faster and more successful for navigational tasks in general [6], [9]. The influence of query frequency and the popularity of search goals have also been analyzed by Downey et al. [7]. They find that searchers are more successful for common queries and common goals, and also if the frequency of the query matches the frequency of the user's information need. In addition, caption features such as the occurrence of query terms in the title of a result entry significantly influence whether searchers choose to view a result [5].

There is a large body of work on individual differences in search behavior. White and colleagues summarize this work and report findings from large-scale log studies in which they find that search experts [24] and domain experts [23] are more successful and employ different search strategies than novices.

2.2 Eye Tracking on SERPs

Previous research has used eye-tracking studies to understand in detail how searchers examine search results.

Joachims et al. [16] show that the way in which searchers examine a SERP is influenced by the position and relevance of the results. Searchers have a strong bias towards result entries at higher positions on the SERP. Pan et al. [19] and Guan & Cutrell [10] have also reported similar findings.

Cutrell & Guan [6] look in more detail at how eye movements are influenced by the snippets for search results. They find that longer snippets lead to better search performance for informational tasks, but degrade performance for navigational tasks.

Aula et al. [1] find two different types of searchers exhaustive and economic searchers. Exhaustive searchers examine a SERP thoroughly and look up and down the SERP several times before choosing a result entry to click on. In contrast, economic users sequentially look from the top to the bottom and click on the first relevant result entry they notice (see also [9]).

2.3 The Influence of Ads

The previously mentioned eye-tracking studies focus on eye movement behavior on the 10 organic results. However, SERPs typically contain many additional elements including sponsored links, spell suggestions, related queries, rich snippets, etc. Yet there is very little research about the influence of these components on the search behavior and success.

Most of the available research work focuses on sponsored links, which can account for 10% to 23% of all links presented on a SERP [11], depending on the search engine and query. Fallows [8] reports that in 2005 only 38% of searchers were aware of the concept of sponsored links at all, and only 12% could reliably differentiate between sponsored links and organic results.

Jansen et al. [14] analyze factors relating to clicks on sponsored links. They conducted a study in which participants had to interact with SERPs that had 10 organic results and some textual ads on the right side. They find that 51% of the users only look at organic results and completely ignore the ads. (However, this was determined with think-aloud techniques rather than eye tracking.) In addition, users typically examine ads if they do not find an answer to their task on the initially viewed organic results. In general, they report a considerable bias against ads, even when controlling for their quality.

Interestingly, Jansen [13] finds that summary snippets of ads presented by commercial Web search engines are usually as relevant as summaries of organic results. Further, Jansen & Spink [15] report that seamlessly integrating ads with the organic results (i.e., making a differentiation between them impossible) does not increase their click-through rate. Finally, as shown by Yan et al. [25] behavioral targeting of ads can drastically improve click-through rates.

In summary, search behavior is influenced by individual user characteristics, the type of search task at hand, and the relevance of the search results. We extend this work by analyzing visual attention to the full range of elements in contemporary SERPs. We also systematically vary and examine the effects of ad quality. Finally, we study the dynamics of user attention and behavior by varying the order in which ads of different quality are presented.

3. METHODS

We use eye-tracking as an instrument to provide detailed information about the user's visual attention. It is common for eye-tracking studies to take gaze position as a proxy for visual attention. Thus, gaze tracking can provide data leading to valuable insights about search strategies and processes.

3.1 Experimental Design and Procedure

We designed an eye-tracking experiment in which participants had to complete a number of given search tasks using a Web search engine. We were interested in effects of:

- task type (i.e., informational or navigational),
- elements on search engine results pages (SERPs), most importantly the quality of the ads, and
- the order in which SERPs containing ads of good or bad quality were presented to a participant.

Table 1: Examples of task descriptions and initial queries used for the study.

Task Description	Initial Task Query	Task Type
How much optical zoom does the compact digital camera Sony Cyber-Shot W230 have?	sony cyber shot W230	Info
Find the special offers page for Southwest Airlines.	southwest special offers	Nav
Find the official Web site of the Venetian casino in Las Vegas.	las vegas casino venetian	Nav
How many guest rooms does the Bellagio hotel in Las Vegas have?	bellagio las vegas rooms	Info
What are some side-effects of Ibuprofen?	ibuprofen side effects	Info
Go to NikeStore on the official Nike homepage.	nike shoes	Nav

Tasks

Every participant had to solve the same set of 32 search tasks. Half of the tasks were navigational (i.e., they had to find a specific Web page) and half were informational (i.e., they had to find factual information). All of the tasks were of a commercial nature so that ads would be a realistic component of the SERPs.

Each task had a description telling the participants what they should look for. In order to make the initial SERP comparable across participants, we provided them with an initial query for each task. Some examples of task descriptions and the corresponding initial task queries are given in Table 1.

We cached results for each initial query. This allowed us to have a consistent initial set of results for each task which we knew included a solution to the task in a fixed position. For 24 (75%) of the tasks, the static first SERP contained a solution within the top 3 organic results, for 6 tasks (19%), a solution could be found in positions 4-6, and for 2 tasks (6%), a solution was after position 6.

After the initial SERP was presented, participants were free to proceed as they wished. They could click links, view the next page of results, or re-query. The combination of an initial fixed SERP and full search functionality provides a good balance between experimental control and search realism for a laboratory study.

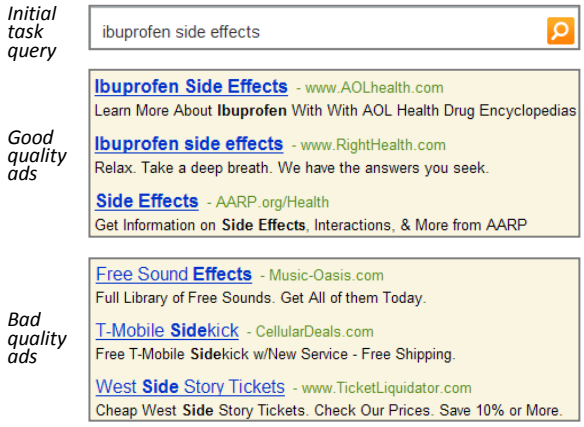

Figure 3: Example of good and bad quality ads for the same initial task query.

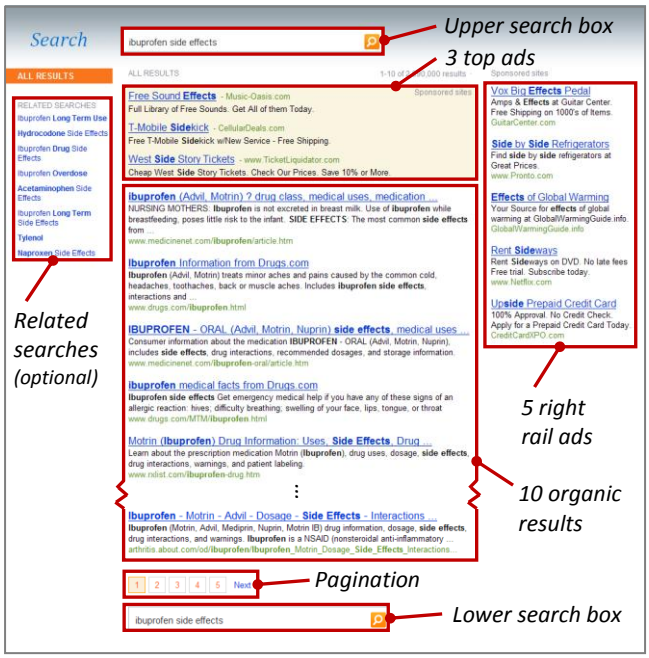

Figure 2: SERP layout

Ad Quality

In the study, each SERP contained 3 ads at the top and 5 ads on the right rail (see Figure 2). For each SERP, all of the displayed ads were either of good or of bad quality. Figure 3 shows an example of 3 good quality and 3 bad quality ads for the query "Ibuprofen side effects". Across participants, each task was shown equally often with good quality ads or bad quality ads.

The good ads were selected from the ads shown by commercial Web search engines such as Bing, Google, and Yahoo! in response to the initial task queries. The bad ads were selected from the same commercial Web search engines by generating queries using a subset of the terms occurring in the initial task queries. This matching method is responsible for some types of bad matches observed in practice. Since the bad ads were generated from query terms, they contained highlighted terms making them visually similar to the good ads. For the determination of good and bad ads we only consider topicality, but no other factors such as the reputation of the sponsor, etc.

SERP Elements and SERP Generation

The layout of the SERPs was modeled after a commercial Web search engine. As depicted in Figure 2, a SERP contained the following important elements

- an upper and lower search box,
- 10 organic results (not containing any special elements like maps, videos, images, or deep links),
- 3 top ads,
- 5 right rail ads, and
- related searches on the left rail for queries for which they were available (20 of the 32 initial queries contained related searches).

To generate the SERP for a query, we implemented our own search interface shown in Figure 2. For the initial task query the interface showed a locally cached version of the first SERP for the query. For any other user-generated query, the interface queried a commercial Web search engine in the background, took the

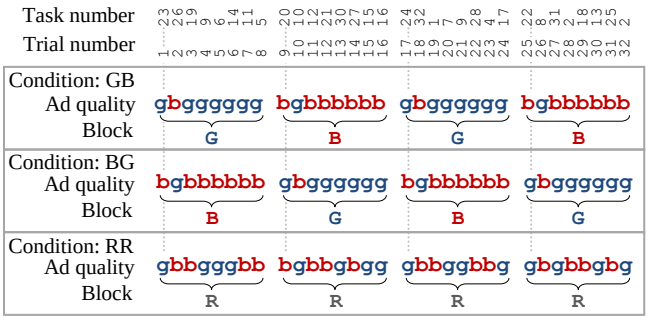

Figure 4: Experimental variables. Each sequence of randomly assigned tasks is performed in 1 of 3 conditions (BG, GB, RR). The sequence conditions determine when the SERPs contain good (g) or bad (b) quality ads.

organic results and the related searches (if any), inserted ads, and displayed them using our modified interface layout.

For each task, we had a pool of good quality and of bad quality ads. The static first SERPs for the initial task queries always contained the same ads from the appropriate pool. For subsequent queries, ads from the appropriate pool were randomly selected and integrated into the SERP at runtime.

Trial Sequences

To study sequence effects, we controlled the order in which SERPs with good or bad ad quality appeared within the sequence of 32 tasks. In the following, we introduce some terminology for describing how the task sequences were created (see Figure 4).

A *trial* is one unit of the experiment starting from reading the task description until completing the task. There were 32 trials in an experiment, one trial for each task. For each trial, we specified which task to solve and whether to show only good or only bad quality ads on the SERPs for that task.

The experiment was divided into 4 *blocks*, of eight consecutive trails. There are three types of blocks - Good (*G*), Bad (*B*) or Random (*R*). A good quality block (*G)* contains 8 trials with mostly good ad quality, whereas a bad quality block (*B)* contains 8 trials with mostly bad ad quality. To make the blocking effect less obvious to the participants, the ad quality in the second trial of each *G* or *B* block is reversed. In Figure 4, a lower case g or b is used to delineate a trial containing only good (g) or only bad (b) ads. Random blocks (*R)* consist of half good and half bad ad quality trials (randomly distributed within an *R* block). The only constraint on the random selection was that, across all participants, each task should be performed using good quality ads around the same number of times as using bad quality ads. For all trials in all conditions, all of the ads displayed on a SERP were either of good (g) or of bad (b) quality.

Each participant was assigned to one of 3 *conditions GB*, *BG*, or *RR*. Each condition contains 4 blocks of trials GBGB, BGBG, and RRRR, respectively (see Figure 4). Thus every participant performed 16 tasks with SERPs showing good quality ads and 16 tasks with SERPs showing bad quality ads.

The order of the tasks in a 32 trials sequence was randomly assigned. Each unique task sequence was performed in all 3 conditions by 3 different participants. It is important to note that the participants saw 32 trials without any special delineation of the block structure or the quality of the ads. The blocks and presentation conditions are for analysis purposes.

Summary of Independent Variables

To summarize, the main independent variables for each trial were

- Task type (informational/navigational)
- Quality of the ads (good/bad) shown on the SERPs
- Block (*G*/*B*/*R*) the trial belongs to
- Condition (*GB*/*BG*/*RR*) the participant was assigned to

Procedure

After a short introduction to the study, the eye tracker was calibrated using a 5 point calibration. Then, the participants started with one practice task to illustrate the procedure and continued in the same way for the remaining 32 tasks.

For each task, we provided the participants with a written task description and the corresponding initial query. After reading the description and the query aloud, the participants pressed a search button to begin searching using the initial query. The first SERP was always the locally stored, static page containing ads of the appropriate quality for that trial. From here on, participants were free to interact with search results. To solve the task, they had to navigate to an appropriate Web page and point out the solution on it to the experimenter. After finding a solution, they had to answer the question "How good was the search engine for this task?" (5-point Likert scale).

After completing the example task and all 32 main tasks like this, the participants had to fill in a study questionnaire asking about their Web search experience and practices during the study and in general. The experiment took about one hour per participant.

3.2 Apparatus

The experiment was performed on a 17" LCD monitor (96 dpi) at a screen resolution of 1280x1024 pixels. We used the browser Internet Explorer 7 with a window size of 1040x996 pixels. With this setting, the page fold was usually between the organic results at positions 6 and 7. For gaze tracking, we applied a Tobii x50 eye tracker which has a tracking frequency of 50 Hz and an accuracy of 0.5° of visual angle. Logging of click and gaze data was done by the software Tobii Studio.

3.3 Participants

Thirty-eight participants produced valid eye-tracking data (out of 41). Participants were recruited from a user study pool. They ranged in age between 26 and 60 years (mean = 45.5, σ = 8.2), and had a wide variety of backgrounds and professions. 21 participants were female and 17 were male.

For the 38 participants, we generated 13 unique task sequences. 13 participants were assigned to the GB condition, 13 to BG, and 12 to RR. Overall, we got valid eye-tracking data for 1210 trials.

3.4 Measures

For our analysis, we wanted to know how visual attention and clicks are distributed among different elements of the SERP. Therefore, we assigned gaze and click data to areas of interest (AOIs) on the SERPs.

AOIs

Since all SERPs presented during the study had the same kinds of elements, we created common areas of interest. All regions labeled in Figure 2 were defined as AOIs. For the top ads, the right rail ads, and the organic results, we introduced further AOIs matching the different result entries (i.e., separate AOIs for each of the 10 organic results, for the 3 top ads and for the 5 right rail ads). In addition, each of those result entries contained 3 AOIs matching the title, the summary text snippet, and the URL.

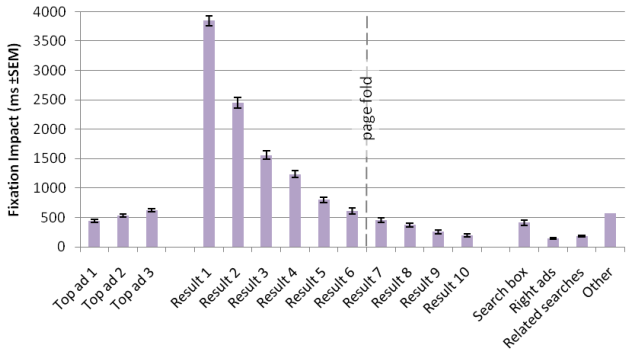

Figure 5: Mean fixation impact on SERP elements in milliseconds (including standard errors of the mean).

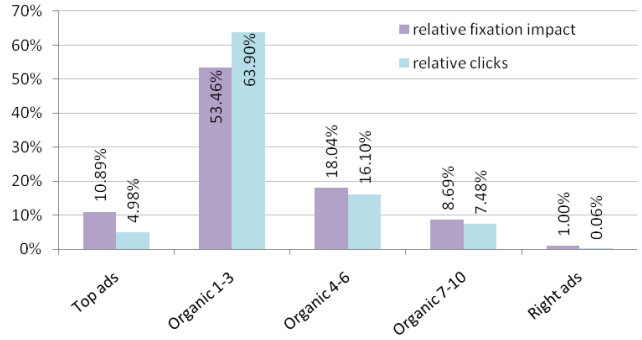

Figure 6: Percentage of visual attention and of clicks attracted by different AOIs.

Fixation Impact *fi(A)*

Fixations were detected using built-in algorithms of Tobii Studio. The algorithms generate a fixation if recorded gaze locations of at least 100ms are close to each other (radius 35 pixels).

We used the measure fixation impact *fi(A)* to determine the amount of gaze an AOI *A* received. This measure was introduced by Buscher et al. [4] and is a modified version of simple fixation duration. Fixation duration assigns the entire duration to the AOI(s) that contain the center point of the fixation, but the fixation impact measure spreads the duration to all AOI(s) close to the fixation center using a Gaussian distribution. Thus, fixation impact prorates the duration of a fixation to all AOIs that are projected on the foveal area of the eyes.

Clicks *c(A)*

We also count the number of clicks on any links on the SERPs (e.g., organic results, top and right rail ads, related searches, etc.). *c(A)* specifies the number of clicks aggregated for the AOI *A*. For example, for the AOI *top_ads* spanning all 3 top ads, *c(top_ads)* would be the sum of clicks on any of the 3 top ads.

Time on SERP *t*

Finally, for each participant and task, time on SERP *t* measures the time the participant spent on the first static SERP for a given task. This time includes all views of the first static SERP, not only the time to first click.

4. RESULTS

For this analysis we focus on several aspects of gaze on SERPs. First, we want to determine *differences in visual attention (on AOIs) with respect to task type*. Second, we are interested in the *difference in visual attention on good and on bad quality ads* on SERPs. Third, we focus on *sequence effects of presenting good or bad ads* in different orders.

We concentrate our gaze-based analysis on the static first SERPs for the initial queries we provided to the participants. Since the same static first SERPs are viewed by each participant (either with good or with bad quality ads displayed), we are confident that all participants are looking at exactly the same information, and comparability is ensured. Gaze on the first SERP represents 88% of the total gaze on all SERPs, and 32% of total task time (with the remainder of the time spent on reading Web pages).

4.1 General Gaze Distribution on SERPs

We start our analysis with a general overview of the distribution of visual attention on SERP components. Figure 1 shows a gaze heat map depicting the distribution of visual attention, averaged over all 38 participants and all 32 trials. (It should be noted that the specific SERP in the background of the figure is just an example to show gaze relative to AOIs.) The figure shows the well-known gaze distribution referred to as "golden triangle" [12] or "F-shaped pattern" [18], describing where people allocate their visual attention on SERPs in the aggregate.

Figure 5 shows mean fixation impact on the different SERP components averaged across participants and across trials. Not surprisingly, most visual attention was devoted to the top few organic results. Interestingly, however, the top ads received as much attention as results around the fold (at positions 6-7). In Figure 6 we show the percentage of visual attention AOIs received from the participants along with the percentage of clicks on the respective AOIs. Gaze and click patterns are in general agreement, but there are some interesting differences. The organic result entries at position 1, 2, and 3 together attracted only 53.46% of visual attention on the SERPs, but 63.90% of all clicks. In addition, the top ads received more than 10% of the visual attention but fewer than 5% of the clicks.

Discussion. Overall, our findings are in line with previous research concerning the distribution of visual attention to organic search results on SERPs [12], [18]. We further see that the relative distribution of clicks does not always reflect the relative distribution of visual attention (Figure 6). We find that there are proportionally more clicks that attention on top results, which is consistent with the previously reported bias towards selecting one of the top organic results [16], [17].

Conversely, we find that top and right rail ads receive a higher fraction of visual attention than of clicks. This extends previous subjective reports of a bias against sponsored links [14]. In addition, although each of the top ads got approximately as much visual attention as organic results at the page fold (see Figure 5), the top ads got considerably fewer clicks than organic result entries even below the fold.

4.2 Effects of Task Type

Figure 7 (left side) shows average fixation impact for SERP elements, broken down separately for informational and navigational tasks. There are several large differences of the general gaze distribution with respect to task type.

First, the participants spent significantly more time on SERPs for informational tasks than for navigational ones (mean=16.5 and 12.9s respectively, $t(1208)=3.8$, $p < 0.01$).

Second, most of the additional time during informational tasks was spent on the organic results and on the upper search box. Almost every single position in the organic results received more visual attention for informational than for navigational tasks. This

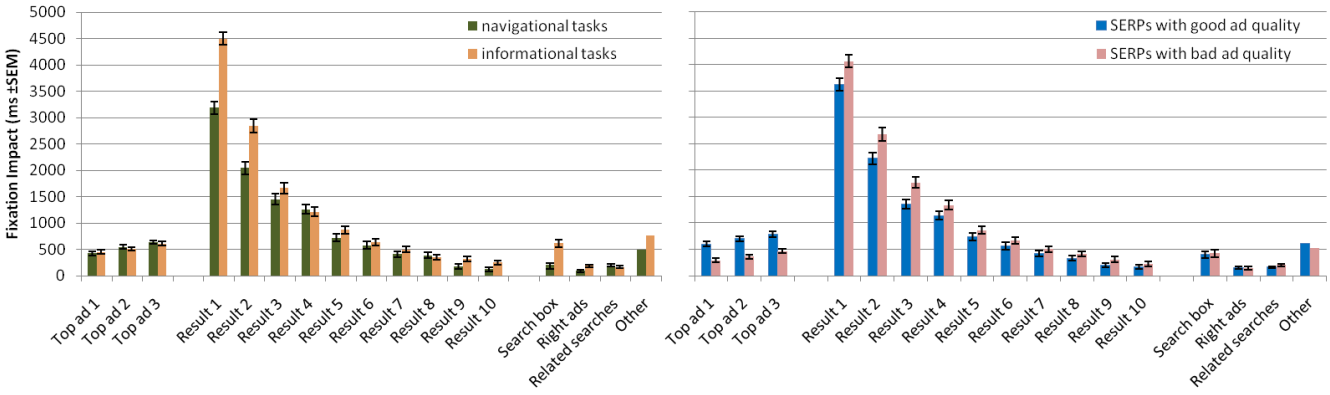

Figure 7: Comparison of mean fixation impact on SERP elements for navigational and informational tasks (left) and for SERPs displaying good or bad ads (right).

is especially evident for the organic results at positions 1 and 2 ($t(1208)=7.4$, $p < 0.01$).

Interestingly, there was virtually no difference in the distribution of gaze on the top 3 ads. For the right rail ads we observe slightly more visual attention during informational tasks, although the absolute amount of attention is very low (mean informational=189ms, mean navigational=104ms, $t(1208)=2.7$, $p < 0.01$).

Discussion. It is striking that none of the extra time for informational tasks was spent on the top ads. Users did not distribute their additional time evenly on the elements of the SERP but seemed to concentrate their attention on the top 2 organic results. This suggests that for informational tasks where users typically focus more on text snippets, the bias for the top organic results is even stronger.

Furthermore, there is a noticeable difference in visual attention on the upper search box which is more than twice as high for informational tasks. This reflects that the fact users requeried more during informational tasks (1.20 queries) than navigational tasks (1.05 queries). Interestingly, even when participants were not able to find the solution on the first static SERP, they did not divert their attention much towards other components of the SERP such as ads or related searches, instead they requeried.

4.3 Effects of Ad Quality

Figure 7 (right side) shows average fixation impact for SERP elements, broken down separately for SERPs containing good and bad ads. There are several large differences of the general gaze distribution with respect to ad quality.

Overall, participants spent somewhat less time on SERPs when good quality ads were displayed (mean time on SERP 14.2s, σ=16.5s for good quality ads vs. 15.2s, σ=16.0s for bad quality ads), however, the difference is not statistically significant.

There are, however, interesting differences in the gaze distribution on different components of the SERPs. Participants devoted about twice as much visual attention to top ads when the ads were of good quality (mean=2.1s and 1.1s for good and bad quality, respectively, $t(1208)=6.8$, $p < 0.01$). In contrast, participants paid consistently less attention to the organic results when good quality ads were displayed (mean=10.8s and 12.8s for good and bad quality, respectively, $t(1208)=2.6$, $p < 0.01$). There were no reliable effects of ad quality on the remaining SERP components such as the search box, right rail ads, and related searches.

We further analyzed the participants' search engine judgments for each trial with respect to ad quality. When good quality ads were displayed, participants rated the search engine slightly better than when bad quality ads were presented (mean of 4.55 vs. 4.49 on a 5-point Likert scale), however this difference is not significant. In addition, the total time to complete a task was about 10% shorter when good quality ads were shown (mean 50.4s, σ=53.1) than when bad quality ads were shown (mean 54.4s, σ=62.2), but this difference is not significant.

Discussion. The quality of ads on a SERP directly influenced participants' attention and performance. Top ads of good quality attracted twice as much visual attention as those of bad quality. In addition, the amount of attention devoted to organic results was influenced by the quality of the ads, with less attention to organic results when the ads were good.

The effect of ad quality on visual attention was not evident for right rail ads. Right rail ads seem to be largely ignored, and when participants looked there, they did not do so differentially as a function of ad quality.

4.4 Sequence Effects

Every sequence of 32 tasks was performed in three different conditions, *GB*, *BG*, and *RR* (see Figure 4). The condition determined the order in which good or bad quality ads were displayed on the SERPs for the different tasks. In this section, we concentrate on effects observed in those different conditions. Figures 8 and 9 show results (for fixation impact and clicks) for these three conditions, broken down separately for SERPs containing good and bad ads.

In Figure 8, we see that mean fixation impact on the top ads from participants in either the *BG* or the *GB* condition was around 1.8 times larger than from participants in the *RR* condition ($t(1208)=5.3$, $p < 0.01$). Good ads generated higher fixation impact for all conditions than bad ads. In the blocked conditions (*BG* and *GB*), good quality top ads received twice as much visual attention as in the random condition (*RR*). Further, good quality top ads in the random condition received only as much gaze as bad quality top ads in the blocked conditions.

In Figure 9, we see even larger differences for clicks as a function of condition and ad quality. Not surprisingly, the quality of the ads had a large effect on click rate – there were no clicks on bad ads and a click rate of about 13% for good ads. Condition also had a large effect on click behavior. Participants in the *BG* or *GB* conditions clicked on the top ads 2 to 3 times more often than

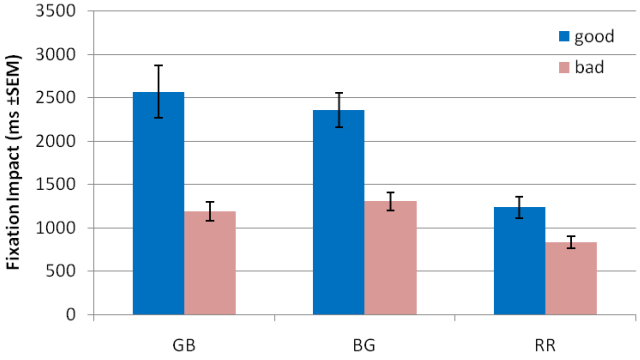

Figure 8: Mean fixation impact on the top ads *fi(top_ads)* split by sequence of blocks (GB, BG, RR) and the quality of the displayed ads on the SERP (good / bad).

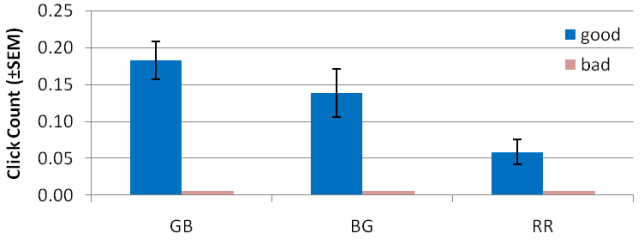

Figure 9: Mean number of clicks on the top ads split by sequence of blocks for good ads (there were no clicks on bad ads).

participants in the *RR* condition (average click rate of 16% for BG and GB vs. 6% for RR, t(1208)=3.0, p < 0.01).

Discussion. The differences in fixation impact and number of clicks suggest that the order in which the participants see SERPs with either good or bad ad quality strongly affected their search behavior. When SERPs with good or bad ad quality were presented in random order, participants tended to ignore the ads more, even when they were of good quality. On the contrary, when SERPs contained ads of consistently good quality, then participants were more likely to pay attention to and click them.

This observation implies that predictability is an important factor influencing how users attend to different regions on the SERP. If the quality of ads is unpredictable, then users seem to get "ad blind" so that even good ads receive less attention. This finding is of direct importance to search engine providers, who might be able to increase revenue from sponsored links by showing few ads of consistently high quality. If ads are of *predictably* good quality, then general ad blindness might be reversed.

4.5 Blocking Effects

Since we designed the trial sequences to contain four blocks of SERPs with different ad quality, we expected to see behavioral changes within each trial sequence, especially with respect to the way the participants attend to the top ads.

As expected, we observed that participants completed tasks more quickly during the course of the experiment. The total time they needed to complete a task dropped from the first half to the second half of the experiment (1st half mean=55.7s, 2nd half mean=49.1s, t(1208)=2.0, p < 0.05). Also, the time the participants needed to evaluate the first SERP for a given task query decreased from the first to the second half (1st half mean=15.9s, 2nd half mean=13.5s, t(1208)=2.5, p < 0.05).

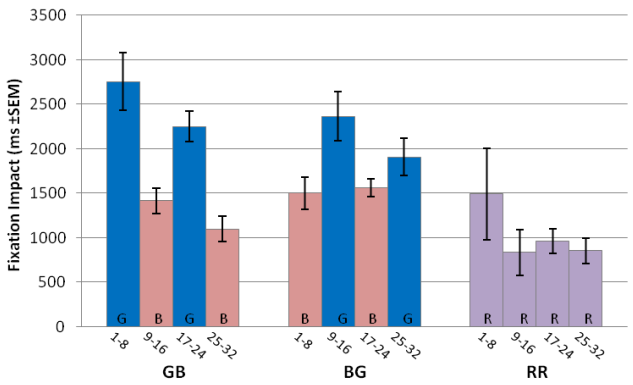

Figure 10: Mean fixation impact on the top ads *fi(top_ads)* split by sequence of blocks (*GB*, *BG*, *RR*) and the block type (good, bad or random) of each block of 8 trials (see Figure 4).

Figure 10 shows the average fixation impact on top ads for the four blocks in the experiment, broken down by condition. As expected, during blocks containing SERPs with good ad quality ("*G*"), more visual attention was directed to the top ads than during bad ad quality blocks ("*B*"). For example, in the *GB* condition, users spent more than twice as much time looking at top ads in the first and third blocks which contain good quality ads. Within the random condition, the highest fixation impact on the top ads occurs during the first block, with a more than 30% drop for subsequent blocks (1st block mean=1.5s, σ=1.7, 2nd-4th block mean=0.9s, σ=1.2, t(377)=3.8, p < 0.01).

Discussion. These findings show that if ad quality is predictably good, then users do notice and pay more attention to the top ads. Even after having experienced a block containing SERPs with bad ads, participants start to pay more attention to the ads when their quality changes. Interestingly, this was also the case for the second block in the *BG* condition in which participants started with a block of bad ad quality and then switched to a block of good ad quality. This shows that amount of attention devoted to ads depends on the context of previous experience.

Interestingly, the amount of visual attention on the top ads during the first block of the random condition *RR* (which consists of 50% SERPs with good quality ads) was the same as during the first block of the *BG* condition (which consists almost only of SERPS with bad quality ads). This suggests that predictability concerning the quality of the ads is important from the very beginning. If quality is unpredictable, then users quickly start to devote less attention to the ads.

5. CONCLUSION

In this paper, we presented the results of an eye-tracking study to characterize how factors such as task type and ad quality influence how users allocate their visual attention to different components of search engine results pages (SERPs). We found significant effects of task type (informational/navigational), ad quality (good/bad) and the sequence in which ads of different quality were shown.

Consistent with previous research, we found a considerable bias of users' visual attention towards the top few organic result entries which is even stronger for informational than for navigational tasks. Furthermore, we found a strong bias against sponsored links in general. Even for informational tasks, in which participants

generally had a harder time finding a solution, the ads did not receive any more attention from the participants.

The quality of ads had a significant influence on the amount of visual attention that participants devoted to both the top ads and the organic results. When good quality ads were displayed, participants paid twice as much attention to these ads and less attention to the organic results. This is strong evidence that how people attend to search results depends on the quality and content of other page elements.

In addition, gaze patterns were strongly related to the order in which good and bad ads were presented. When the quality of ads varied randomly across trials, participants attended to them less than half as long as when the quality varied more predictably by blocks. Strikingly, when ad quality varied randomly, participants attended to the top ads no more than they did for trials in which only bad ads were shown, even though the ads were good on half of the trials. These results are relevant for search engine providers in understanding how ads are perceived under different conditions, and more generally in understanding how prior search experiences influence the current allocation of attention.

This research represents a first step in understanding how task, ad quality and sequence influence search interaction. In our study, we focused on a specific static SERP composition which always consisted of 10 textual organic results with top and right rail ads. One next step for our research is to look at richer variations of SERP composition, e.g., including snippets that contain images, maps or deep links, and using a broader range of queries for which ads may or may not be present. In addition, we would like to explore how the quality of ads interacts with the quality of organic results. For example, how does the presence of a bad ad (or result) affect the perception of the other good quality ads (or results)? Finally, we would like to extend our understanding of temporal dynamics to enable us to develop richer models of search processes and strategies that go beyond individual SERPs to include session behavior as well as longer-term effects.

6. ACKNOWLEDGMENTS

We would like to thank our colleagues in adCenter, Bing and MSR for their valuable comments and great support. We are also grateful for the reviewers' thoughtful comments, and to our participants for their time and insights.

7. REFERENCES

[1] Aula, A., Majaranta, P. & Räihä, K. Eye-tracking reveals the personal styles for search result evaluation. In *Proceedings INTERACT 2005*, 1058-1061.

[2] Broder, A. A taxonomy of web search. *SIGIR Forum*, 2002, vol. *36*, 3-10.

[3] Beymer, D., Russell, D. & Orton, P. Z. An eye tracking study of how font size, font type, and pictures influence online reading. In *Proceedings INTERACT 2007*, 456-460.

[4] Buscher, G., Cutrell, E. & Morris, M. R. What do you see when you're surfing? Using eye tracking to predict salient regions of web pages. In *Proceedings CHI 2009*, 21-30.

[5] Clarke, C. L. A.,, Agichtein, E.., Dumais, S. & White, R. W. The influence of caption features on clickthrough patterns in web search. In *Proceedings SIGIR 2007*, 135-142.

[6] Cutrell, E. & Guan, Z. What are you looking for? An eye-tracking study of information usage in web search. In *Proceedings CHI 2007*, 407-416.

[7] Downey, D., Dumais, S., Liebling, D. & Horvitz, E. Understanding the relationship between searchers' queries and information goals. In *Proceedings CIKM 2008*, 449-458.

[8] Fallows, D. Search engine users. Pew Research Center, 2005. Retrieved January 18, 2010 from http//www. pewinternet.org/Reports/2005/Search-Engine-Users.aspx.

[9] Granka, L. A., Joachims, T. & Gay, G. Eye-tracking analysis of user behavior in WWW search. In *Proceedings SIGIR 2004*, 478-479.

[10] Guan, Z. & Cutrell, E. An eye tracking study of the effect of target rank on web search. In *Proceedings CHI 2007*, 417-420.

[11] Höchstötter, N. & Lewandowski, D. What users see – Structures in search engine results pages. *Information Sciences*, 2009, vol. *179*, 1796-1812.

[12] Hotchkiss, G., Alston, S. & Edwards, G. Eye tracking study, 2006. Retrieved January 18, 2010 from http//www.enquiro.com/eyetrackingreport.asp.

[13] Jansen, B. J. The comparative effectiveness of sponsored and nonsponsored links for Web e-commerce queries. *ACM Transactions on the Web*, 2007, vol. *1*, article 3.

[14] Jansen, B. J., Brown, A. & Resnick, M. Factors relating to the decision to click on a sponsored link. *Decision Support Systems*, 2007, vol. *44*, 46-59.

[15] Jansen, B. J. & Spink, A. Investigating customer click through behaviour with integrated sponsored and nonsponsored results. *International Journal of Internet Marketing and Advertising*, 2009, vol. *5*, 74-94.

[16] Joachims, T., Granka, L., Pan, B., Hembrooke, H. & Gay, G. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings SIGIR 2005*, 154-161.

[17] Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., Granka, L., Pellacini, F. & Pan, B. Eye tracking and online search Lessons learned and challenges ahead. *JASIST*, 2008, vol. *59*, 1041-1052.

[18] Nielsen, J. F-Shaped pattern for reading Web content, 2006. Retrieved January 18, 2010 from http//www.useit.com/alertbox/reading_pattern.html.

[19] Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G. & Granka, L. In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 2007, vol. *12*, 801-823.

[20] Rose, D. E. & Levinson, D. Understanding user goals in web search. In *Proceedings WWW 2004*, 13-19.

[21] Smith, C. L. & Kantor, P. B. User adaptation: Good results from poor systems. In *Proceedings SIGIR 2008*, 147-154.

[22] Turpin, A. & Schöler, F. User performance versus precision measures for simple search tasks. *In Proceedings of SIGIR 2006*, 11-18.

[23] White, R. W., Dumais, S.T. & Teevan, J. Characterizing the influence of domain expertise on Web search. In *Proceedings of WSDM 2009*, 132-141.

[24] White, R. W. & Morris, D. Investigating the querying and browsing behavior of advanced search engine users. In *Proceedings of SIGIR 2007*, 255-262.

[25] Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y. & Chen, Z. How much can behavioral targeting help online advertising? In *Proceedings WWW 2009*, 261-270.