Contents | Author index | Subject index | Search | Home

We still trust in Google, but less than 10 years ago: an eye-tracking study

Sebastian Schultheiß, **Sebastian Sünkler**, and **Dirk Lewandowski**

Introduction. *The purpose of this study is to replicate a study from 2007, which found that student users trust in Google's ability to rank results more than in their own relevance judgements.*
Method. *In a between-subjects experiment using eye-tracking methodology, participants (n=25) worked on search tasks where the results ranking on search engine results pages had been manipulated.*
Analysis. *The data were analysed using descriptive statistics, analysis of variance (ANOVA), t-tests and mixed model analyses with the statistical program SPSS.*
Results. *This study confirms major results from the study we replicated, but finds one important difference: although the viewing behaviour was influenced more by the position than by the relevance of a snippet, the crucial factor for a result to be clicked was the relevance and not its position on the results page.*
Conclusions. *The subjects of our experiment displayed an emancipated search behaviour while choosing relevant snippets even in lower positions. This means that, despite the fact that students were influenced by the position of a result, they made choices on the basis of relevance.*

Introduction

In Web search, the ordering of search results is of great importance: users expect the *best* results to be presented on the first positions, and they trust the search engines' ability to find the best results and rank them accordingly (Purcell, Brenner, and Rainie, 2012). Based on actor-network theory (Latour, 2005), Röhle (2010) found that four actors are present on the search engine market: the search engine(s), the information providers, the search engine optimisation professionals, and the users. These all influence a search engine's ranking. While it is obvious that a search engine vendor has influence on its own results ranking, it is also important to consider the other actors. Information providers are interested in achieving the best possible rankings for their content in search engines like Google, as they are to a large degree dependent on visitors they get through search engines. Often, they hire search engine optimisation professionals, whose aim is to increase the visibility of Websites in search engines through the *optimisation* of on-the-page factors, and off-the-page factors, as well. While there is a vast body of practitioner-oriented work on search engine optimisation, we still lack a clear understanding of the effects it has on what users in search engines actually get to see. Finally, users influence the results rankings. Search engines not only use links but also clicks to automatically adjust their rankings through machine learning. Therefore, we can say that every click counts when it comes to search engine results pages.

It is challenging for search engine users to filter large quantities of results. Therefore, users need to trust the search engine's ability to rank results by relevance to the query. Given that no search engine has full control over its results, and that users have no other means than looking at the results they see to determine the quality of a search engine's results in general, the central questions is whether the strong trust users express in search engines (see, e.g., Purcell *et al.*, 2012) is indeed justified.

The role that trust plays in selecting search results and its relationship to users' actual relevance judgements has been examined by Pan *et al.* (2007) in their study 'In Google we trust: users' decisions on rank, position, and relevance'. Their eye-tracking experiment revealed that college students trusted *'Google's positioning more than their own rational judgements based on the evaluation of different alternatives'* (2007, p. 816). Three participant groups had to solve ten search tasks each with the aid of partly manipulated search engine results pages. In the *normal* condition, the results were in their original order. In the *swapped* condition, the positions of the first two snippets were exchanged. In the *reversed* condition, the whole results page had been turned around (the first-ranked snippet was swapped with rank ten, the second-ranked snippet with rank nine, and so on). The major finding was that the testing persons were strongly biased in their decisions towards links in high positions even if the results themselves were not relevant for solving a task.

The aim of the present study is to replicate Pan *et al.'s* (2007) study. This is worthwhile, as (1) the original study is now ten years old, and it is interesting to see whether the results still hold true today; (2) the original study used college students from the U.S. as testing persons. It is also interesting to see whether the results are applicable to other groups of users; and (3) in a larger context, the replication of research results is important, as it strengthens the external validity when the same or comparable results are reached, and weakens or even questions the external validity when further studies achieve different or even contradictory results. Replication and reproducibility, respectively, are now recognised as being important in the context of information retrieval evaluation (Ferro *et al.,* 2016). Beyond simply replicating the original study, however, we elaborate on some issues regarding how the relevance judgements used are collected (see methods section below). The rest of this paper is structured as follows: first, we give an overview of related research on eye-tracking in the context of Web search, on results ordering and its effects, and on trust in search engines. Then, we state our research questions and hypotheses. After that, we describe the methods used, with a focus on where our methods differ from the methods used in the original study. This is followed by the presentation of the results, a discussion, and our conclusions.

Related research

The following section is organized into three parts. We begin with an overview of relevant eye-tracking research focusing on Web search. Next, we describe the effects of the ordering of search results, followed by work on users' trust in search engines.

Eye-tracking research focusing on Web search

The viewing behaviour on several types of Websites, e.g., on Google and Yahoo results pages, has been the subject of research for Pan *et al.* (2004). Their results showed that amongst others the sex of a person and the type of Website are factors influencing viewing behaviour.

One of the first studies exploring information retrieval utilizing eye-tracking was conducted by Granka, Joachims, and Gay (2004). On the one hand, they provided proof of already known behaviour such as the preference of high ranked results. On the other hand, their results demonstrated correlations between viewing and clicking frequency: the time spent on snippets one and two was found to be nearly equal. However, the first one is being selected clearly more often. The work of Granka *et al.* (2004) also delivered groundwork for the study of Pan *et al.* (2007) that we replicated.

Hotchkiss, Alston, and Edwards (2005) focused on the search behaviour of search engine users. Through an eye-tracking study, the authors introduced the term of the *golden triangle* which describes an area of intense eye scan activity on search engine results pages. The concept of the golden triangle, however, only applies to uninterrupted results pages, i.e., simple ranked lists of results. Since Google's Universal Search, users' attention distributes distinctly more because of the integrated results from vertical search engines (such as news, pictures and videos; see Lewandowski, 2015; Liu, Liu, Zhou, Zhang, and Ma, 2015). The importance of the golden triangle for the present study is due to the reason of the study we replicated. In that study, researchers used Google results pages only consisting of ten organic results presented in ranked lists Pan *et al.* (2007). For comparative purposes, we followed the same procedure.

Results ordering and its effects

Results ranking is an integral part of search engines, as their function lies in filtering the vast amount of information available on the Web in response to users' queries. Regarding users' querying and selection behaviour, a large number of transaction-log and click-through studies have been conducted. These studies are descriptive and often limited to time-related and thematic factors. Major findings are that users mostly enter only short queries and select results from the first hits shown (Goel *et al.,* 2010; Höchstötter and Koch, 2009; Jansen and Spink, 2006). This behaviour was found to be consistent over different search engines (Jansen and Spink, 2006). Users most often choose only from the first results page, and they prefer the first few results listed (Joachims *et al.,* 2007). Petrescu (2014) reports that more than two-thirds of all clicks go to the first five positions, and the result ranked first alone accounts for 31% of all clicks. In a large-scale study analysing millions of queries from the Yahoo search engine, Goel *et al.* (2010) found that only 10,000 different Websites account for approx. 80% of clicks on the results pages.

This does, however, not necessarily mean that users get the best results on the first few positions. Schaer *et al.* (2016) compared top-ranked results with so-called *long tail* results, i.e., results shown on lower results positions. They found that the top results are judged as

being only slightly more relevant than the long tail results. They concluded that the long tail provides a rich resource, as it provides the user with different, although still relevant results. Hariri ([2011](#)) came to a similar conclusion when comparing Google's top results with results shown on lower ranks.

Users are generally satisfied with the first few results, even when the results positions are mixed, and therefore, less relevant results are shown on the first position(s) ([Keane, O'Brien, and Smyth, 2008](#); [Pan *et al.*, 2007](#)). Keane *et al.* ([2008](#)) investigated to what extent search engine users are being influenced by the order of results. Parallel to the study we replicated, their results showed that the users largely preferred top-ranked results, even when they were not relevant to the given search tasks. Bar-Ilan *et al.* ([2009](#)) examined the user preferences for different orderings of search results. One had been the original ordering, the other a synthetic ordering consisting of the same snippets. Their results indicated that there is only a slight preference for the original orderings. The authors concluded that the most important factor for clicking on a result is its position, not its relevance.

Trust in search engines

In a representative survey of U.S. Internet users, Purcell *et al.* ([2012](#)) found that users generally regard search results as accurate and trustworthy. Search engine rankings are even seen as an indicator for credibility ([Westerwick, 2013](#)), and users do not reflect on the results judgements given by search engines ([Tremel, 2010](#)), although search engines model credibility predominantly based on popularity data ([Lewandowski, 2012](#)). One of the most striking results from the Pan *et al.* ([2007](#)) study was that users trust Google's rankings even more than their own relevance judgements. This leads to the question relating to the responsibilities of search engine providers regarding their results ([Grimmelmann, 2010](#); [Lewandowski, 2017](#)).

Research question and hypotheses

As this research is a replication of a prior study, our research question is the following: *'Can the results by Pan et al. ([2007](#)) be replicated, despite temporal and geographical differences?'*

The analysis was guided by four hypotheses, detailed below. As we did not find evidence for a changed behaviour in the literature, H1-H3 were retained from Pan *et al.* ([2007, p. 811](#)). On the page level, we analysed viewing and clicking behaviour on all search engine results pages. On the snippet level, we evaluated the determinants of whether a snippet was viewed or clicked. The newly-added Hypothesis H4 and the relevance level focused on the assumed preference for highly ranked results regardless of their relevance.

> **H1:** *At the page level of analysis, ocular data would differ among the three conditions.*
> **H2:** *At the abstract level of analysis, the eye data from participants in the reversed condition would indicate explicit trust for Google's ranking, as evidenced by a lack of significant difference among the three conditions in the number of fixations per abstract on the top two positioned abstracts. Furthermore, subjects would look at the last two positioned abstracts (the number one and two Google ranked abstracts) more than in the other two conditions, indicating an implicit awareness of their significance, either from confusion or interest.*
> **H3:** *Participants in both the swapped and reversed conditions would still choose abstracts of actual lower rank more often than subjects in the control condition (those who viewed Google results in their actual ranked order).*

The authors used 'abstract' as a synonym for 'snippet'. In hypothesis H2, they also mentioned the collection of participants' pupil dilation data. As our eye tracker did not provide that opportunity, this data is not used in our study.

> **H4:** *Participants in all three conditions would choose snippets higher in position even if they were not relevant to the given search tasks.*

Methods

Similar to Pan *et al.* ([2007](#)) we conducted a between-subjects eye-tracking experiment. Participants were randomly divided into three groups; each participant had to solve ten search tasks with the aid of partly manipulated search engine results pages. In the *normal* condition, the results were in their original order. In the *swapped* condition, the positions of the first two snippets were exchanged. In the *reversed* condition, the whole results list was shown in reverse order (the first-ranked snippet was swapped with rank ten, the second-ranked snippet with rank nine and so on). The subjects were randomly assigned to one condition and were not aware of the manipulated results pages. While solving the tasks, the participants' eye movements were recorded. After each task, the subject had to make relevance judgements for the previously seen snippets (i.e., the results descriptions shown on the results pages) and Webpages (i.e., the results documents).

As our study is based on Pan *et al.* (2007), we will primarily point out differences to their methodology in the following, and give only a brief description of methods where they are the same as in the original paper. The reader is referred to Pan *et al.* (2007) for a more detailed description of the methods.

Experimental procedure

The procedure of our experiment corresponded substantially to the study by Pan *et al.* (2007). We made individual appointments with all participants for their participation in April 2016. Before the actual experiment, we gave general instructions regarding the procedure. The subjects were told to view the Webpages and search as they would under normal conditions. After the participant had signed several consent forms (e.g., the privacy agreement), the eye tracker was calibrated using a five-point calibration procedure. The main part of the experiment was to solve the ten search tasks appearing in a random order, shown through the Google interface. Due to technical reasons we used prescribed search engine results pages, so participants were unable to enter their own queries. The maximum time for completing each task was restricted to three minutes for the purpose of maintaining comparability. After solving a task or reaching the time limit, relevance judgements had to be made. Thus, the relevance of the previously seen snippets and Websites (2 * 10 ratings for each task) had to be assessed. After completing all tasks, the experiment was finished.

The subjects

The participants were undergraduate students from a large university in Germany with diverse majors, mainly Library and Information Management, or related. We recruited twenty-eight subjects and obtained twenty-five complete data sets, including fifteen females and ten males. The average age of our participants was twenty-five years and six months. They were randomly assigned to only one of the three conditions (*normal*, *swapped*, or *reversed*).

Search tasks

Pan *et al.* (2007) presented the participants with ten search tasks, half of them navigational, half transactional queries (for the distinction between query types, see (Broder, 2002) and (Lewandowski, Drechsler, and Von Mach, 2012). While we went for similar tasks and a similar distribution between query types, we did not adopt the exact search tasks because of their regional context. Instead, we developed new search tasks relating to Hamburg and Germany, respectively. The aim was to provide a similar relation to our subjects. The complete list of search tasks used in our study can be found in Table 1.

Task type	Task	Correct answer
Navigational	Find the homepage of 'Schulz von Thun Institut für Kommunikation'	http://www.schulz-von-thun.de/
	Find the page displaying an overview of various providers of Hamburg harbour tours	http://www.hamburg.de/hafenrundfahrt/868760/hafenrundfahrten-hamburg/
	Find the homepage of 'Tierpark Hagenbeck'	http://www.hagenbeck.de/startseite.html
	Find the page with an overview of the student residence in Hammerbrook, Hamburg	http://www.studierendenwerk-hamburg.de/studierendenwerk/de/wohnen/wohnanlagen/detail/?id=766
	Find the homepage of Christian Rach - the chef who has several television cooking programmes	http://www.christianrach.de/
	Where is the	

	office of the oldest German shoe factory and in which year this company was founded?	Pirmasens, Germany; in 1838
Informational	Where does Norbert Hackbusch, left-wing politician and member of the Hamburg parliament, work?	Gruner + Jahr
	Who are the television presenters of 'Hamburg 1'?	List of names: https://de.wikipedia.org/wiki/Liste_der_Hamburg-1-Moderatoren
	In which year was 'Frauenklinik Finkenau' founded?	1914
	Which manager of the German Bundesliga is doing his job for the longest successive time?	Markus Weinzierl

Table 1: Navigational and informational search tasks used in this study

Manipulation of the search engine results pages

To evaluate the influence of position and relevance on clicking and viewing behaviour, the search engine results pages were modified in two ways. First, we reduced all results pages to a list of ten organic results. This was done to avoid distraction. We removed elements like sponsored results that otherwise could have influenced participants' behaviour (Buscher, Dumais, and Cutrell, 2010), and Universal Search results. Secondly, the results pages were manipulated for each group of participants in one of three ways. In the *normal* condition, the ordering complied the original Google ranking. In the *swapped* condition, the positions of the two first ranked snippets were swapped, keeping the rest intact. In the *reversed* condition, the whole results page had been turned around (first ranked snippet was swapped with rank ten, second ranked snippet with rank nine, and so on).

Explicit relevance judgements

As stated above, we gathered relevance assessments for all snippets and results (Webpages). In addition to the factors *position* and *condition*, the relevance value is the third factor for which we determined whether it was significant for viewing or clicking a search result, respectively. The collection of judgements was done in two steps after every individual search task. Firstly, we displayed screenshots of the ten previously seen snippets in a random order.

Figure 1: Snippet with slider for relevance judgement

As can be seen from Figure 1, below each of the ten snippets was a continuously adjustable slider. This allowed the participants to judge the relevance of a respective snippet on a scale

ranging from zero to 100. Secondly, we did the same with the Webpages the snippets represented, also in the form of screenshots and sliders. Overall, we collected 5,000 ratings (25 participants * 10 tasks * 20 ratings for each task).

Technical implementation

Identical to Pan *et al.* (2007), we used eye-tracking as the methodology for our study. An eye-tracking device can record the eye movements of subjects. This allowed us to determine which parts of a search engine results page are being considered for how long, how often and in which order (Blake, 2013).

The subjects' eye movements were recorded using a Tobii T60 eye tracker. It employs pupil centre corneal reflection, which is the most commonly used technique for eye-tracking. In doing so, both eyes are being illuminated by a light source with a camera capturing images of the resulting reflections. Following, employing the angle between the reflections of cornea and pupil, the gaze direction can be calculated (Tobii Technology, 2010).

As software, Tobii Studio in version 3.1.6 was used. Therein, we defined the experimental procedure, which was implemented by a tool (described below), and the data to be collected. Typically, several measurements are recorded in eye-tracking experiments. One important metric is *fixation*, which is a moment the eyes are relatively stationary, taking in or encoding information (Poole and Ball, 2005). For comparative purposes, we set the minimum duration for a fixation as 50 milliseconds, as Pan *et al.* (2007) did. Besides the number (*Fixation Count*) and the duration (*Fixation Duration*) of a visual contact, we chose *Mouse Click Count* as metric, as well. Also in Tobii Studio, we set a five-point calibration procedure to be passed before the beginning of an experiment, as suggested in the Tobii T60 eye tracker manual (Tobii AB, 2016). The calibration was successfully completed for all participants. As mentioned above, we developed a tool that predefined the experimental procedure. It was accessible online and enabled us to let the participants click through from task to task. All integrated search engine results pages, as well as the snippets and Websites for the relevance assessments were screenshots. To provide a faithful copy of Google, the results pages were image maps created with the *Online Image Map Editor*. Thus, the results pages became clickable, and every mouse click on a snippet led to the respective Website.

Defining the areas of interest

To gather data for the desired parts of the search engine results pages like the individual results, we defined areas of interest (AOIs), as shown in Figure 2. Areas of interest are defined parts of a display or interface that allow for analysing only the eye movements that fall within each respective area (Poole and Ball, 2005). To aggregate the data of all subjects per task, we grouped the eleven areas of interest of each participant (one results page and ten snippets).
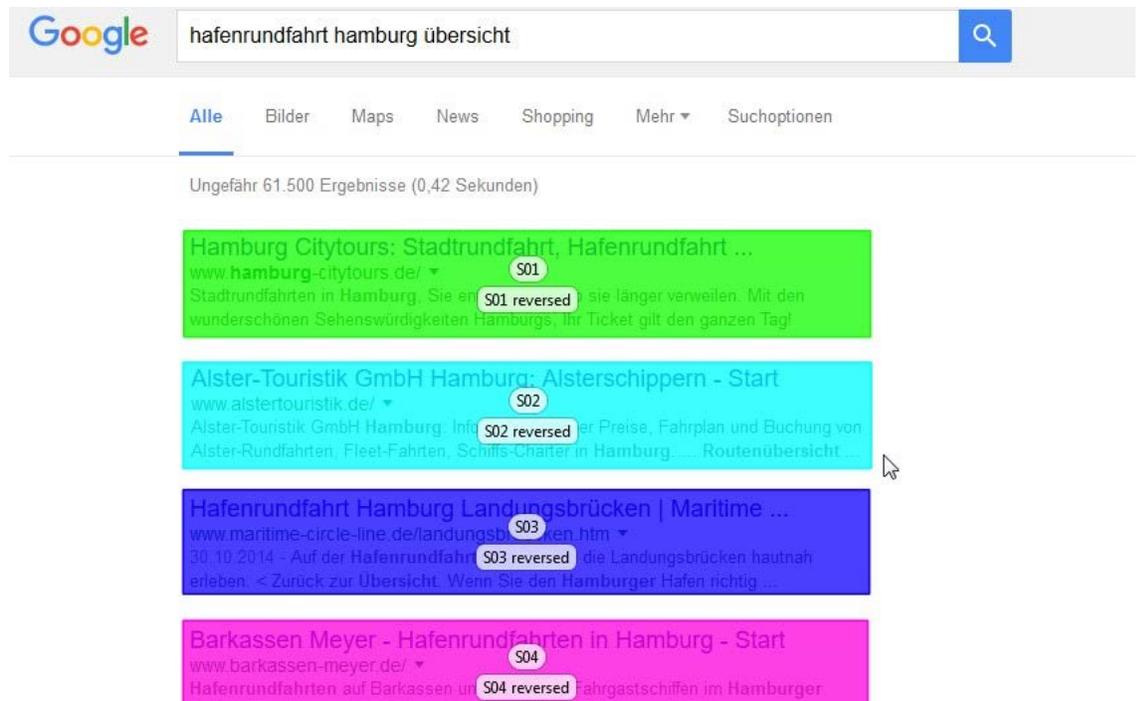

Figure 2: Reversed search engine results page with areas of interest in Tobii

Results

Page level analysis

First, we compared the differences among the three groups concerning fixation duration, number of fixations, and number of mouse clicks on snippets on the search engine results pages.

Participants in the *reversed* condition spent significantly more time viewing the results page during a search task than the subjects in the other two conditions (37.6 seconds vs. 16.5 seconds in the *normal* condition and 23.4 seconds in the *swapped* condition, $p < 0.02$). The subjects in the *reversed* condition also made the most fixations while looking at a results page on average (140.4 fixations vs. 56.9 in the *normal* condition and 79.6 in the *swapped* condition, $p < 0.03$). Regarding the number of mouse clicks on snippets on the search engine results pages, there are no significant differences between the three conditions (2.1 snippets in the *reversed* vs. 1.4 both in the *normal* and in the *swapped* condition). We can confirm our first hypothesis (H1). Although the differences in clicking behaviour are not significant, the participants in the *reversed* condition spent the most time and made the most fixations while checking a results page.

Snippet level analysis

Next, we took a closer look at the ten snippets on each search engine results page. We compare viewing and clicking behaviour among the conditions through diagrams, and the fixations are analysed statistically. Figures 3-5 illustrate the average number of fixations until the first click and the ratio of each snippet on all first clicks in each condition. Subjects in the *normal* condition viewed the top ranked snippet with the highest frequency and clicked on it up to 80% of the total clicks. The snippet on position two also received a high number of fixations but was clicked considerably less often. None of the participants in the *normal* condition chose one of the last five results.
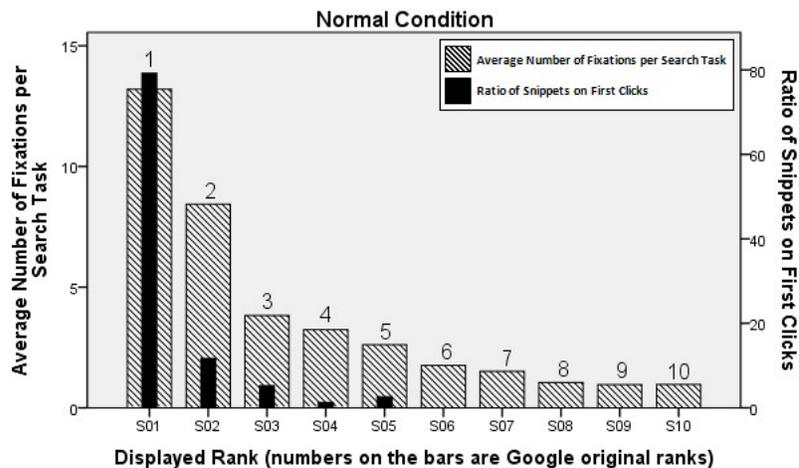

Figure 3: Views and clicks under normal condition

When looking at Figure 4 (showing the *swapped* condition), a clear shift is noticeable. Here, the first two snippets received a nearly equal number of fixations. In contrast to the results of the *normal* condition, snippet number two was now clicked on most often (which is the first result on the *normal* results page). Nevertheless, even if the second result in the *swapped* condition is the same as the first snippet in the *normal* condition, it was chosen substantially less (about 55% compared to the 80% in the *normal* group).
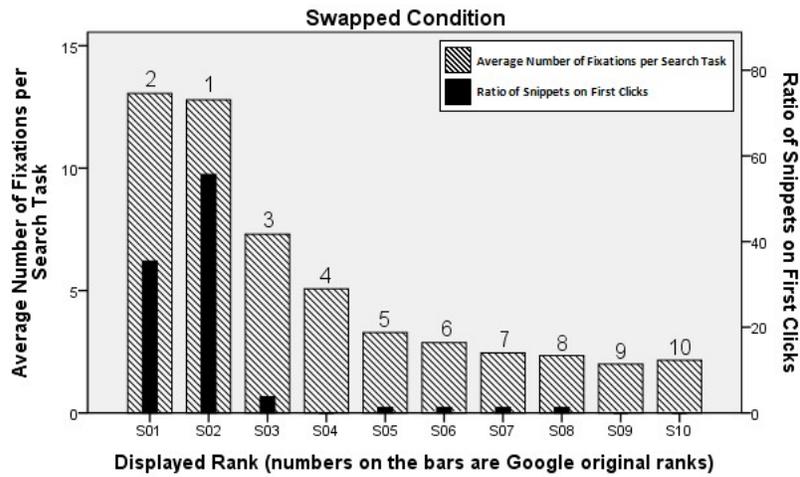
Figure 4: Views and clicks under swapped condition

In the *reversed* condition (Figure 5) the fixations are rather evenly distributed over the ten results of the search engine results page, including two peaks at the top and the bottom of the results list. The snippet on the last position, which is number one in the *normal* condition, was clicked by every second participant (about 50%). Moreover, clicks were distributed over a larger number of different results. For instance, the snippets on the first five positions were clicked by up to 10% in the *reversed* condition. In contrast, the subjects in the *normal* condition did not choose any of these five snippets, which were listed as results six to ten in their lists.
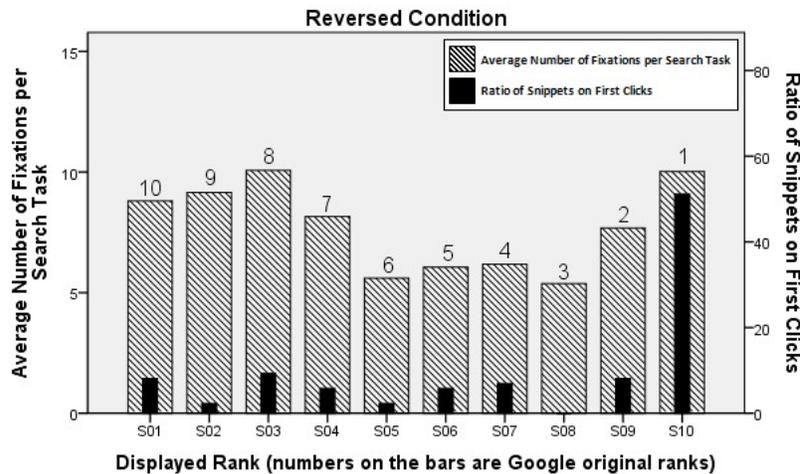

Figure 5: Views and clicks under reversed condition

The last part of the snippet level analysis is the statistical evaluation of the fixations on the snippets. In a first step, we compared the average number of fixations on the first and second snippet, respectively. We did not find significant differences among the three conditions in fixation density between the two first ranked snippets. Hence, even though the participants in the *reversed* condition were actually viewing snippets nine and ten on the first two positions, these snippets were evaluated with equal attention compared to the first two snippets in the *normal* condition.

Next, we compared the three groups regarding the average number of fixations on the ninth and tenth ranked snippet. There were significant differences among at least two of our three conditions ($F(2, 22) = 6.32$, $p < 0.01$). Post-hoc comparisons indicated that the differences are located between the groups *normal* and *reversed* ($p < 0.05$).

Thirdly, we compared the three groups again on the average number of fixations on the snippet ranked first by Google. This snippet corresponds to result number one in the *normal*, result number two in the *swapped* and result number ten in the *reversed* condition. We did not find significant differences in this case.

We can fully confirm our second hypothesis (H2). There are no significant differences among the three groups regarding the number of fixations on the first two snippets. However, there is a significant difference when looking at the snippets nine and ten between the *normal* and

reversed conditions.

We can also confirm our third hypothesis (H3). As can be seen from Figures 3-5, the participants in the *swapped* condition clicked the first ranked snippet considerably more often than the subjects in the *normal* condition chose the second position, which is the same result. When comparing the conditions *normal* and *reversed*, the Figures also indicate that the members of the *normal* group did not choose the last five snippets. In the *reversed* condition, the same results were clicked. This suggests that the result's rank (here on positions one to five) is the crucial factor.

Relevance level analysis

In this part of the analysis, we brought in the relevance assessments of the snippets and Webpages we gathered. Every single participant made 200 relevance judgements (ten search tasks, with each ten snippets and ten Websites being rated). At first we compared the assessments for the snippets with those for their related Websites. A paired t-test indicated that there are no significant differences. Accordingly, it can be assumed that the ratings for the snippets largely correspond to the ratings for the associated Websites. Based on these findings, we incorporated only the relevance assessments for the snippets in the next analysis. Using two mixed model analyses, we were able to measure the influence of the position compared to the relevance of a snippet. Besides the position and relevance, we chose the condition as a factor. In the first mixed model analysis, we set the average number of fixations on snippets per search task as the dependent variable (Table 2).

Source	Numerator df	Denominator df	F	Sig.
Position	9	227	7.526	<0.01**
Condition	2	227	13.559	<0.01**
Relevance Value	10	227	3.576	<0.01**
Note: Dependent variable: average number of fixations on snippets per search task. ** Significant at 0.05 level				

Table 2: Type III Tests of fixed effects on viewing behaviour in the mixed model

The results show that the position and relevance of a snippet and the condition of the search engine results page are significant determinants of how often a snippet is viewed ($p < 0.01$). The F values show the degree of influence on the dependent variable. Thus, the condition (*normal*, *swapped*, or *reversed*) has the biggest influence on the viewing frequency of the snippets.

Source	Numerator df	Denominator df	F	Sig.
Position	9	227	8.280	<0.01**
Condition	2	227	1.187	.307
Relevance Value	10	227	33.726	<0.01**
Note: Dependent variable: average number of clicks on snippets per search task. ** Significant at 0.05 level				

Table 3: Type III Tests of fixed effects on clicking behaviour in the mixed model

In the second analysis, the dependent variable was the average number of clicks on snippets per search task (Table 3). Significant determinants of how often a snippet is clicked are position and relevance ($p < 0.01$). The relevance of a snippet is by far the biggest influential factor.

The fourth hypothesis (H4) has to be rejected. Although the viewing behaviour was more influenced by the position than by the relevance of a snippet, the crucial factor for a result to be clicked was the relevance and not its position on the search engine results page.

Discussion

We can confirm the results of Pan *et al.* (2007) concerning viewing behaviour. Participants in the *reversed* condition spent the most time and made the most fixations while checking a search engine results page. This greater scrutiny of search engine results pages in the *reversed* condition indicates an uncertainty caused by the manipulated results lists. The findings also emphasise the importance of a results' position on a results page. Participants in both the *swapped* and *reversed* conditions clicked on the very results that participants in the *normal* condition clicked on to a much lesser degree, or not at all. This behaviour can be explained by the tendency to click on snippets when they are in a high position regardless of their relevance. Nevertheless, there is one limitation concerning clicking behaviour. A Games-Howell post-hoc analysis revealed that there are no significant differences between the three conditions regarding the number of mouse clicks on snippets on the results pages. This could be caused by a too small number of cases to provide statistically significant results. The key difference between our findings and the study by Pan *et al.* (2007) lies in the selection behaviour. The

crucial factor for a result to be chosen was the relevance and not its position. This means that, despite the fact that students were influenced by the position of a result, they made choices on the basis of relevance. This behaviour might be explained by characteristics of our sample. Our participants all were German students of similar age studying mainly library and information studies-related subjects. Thus, we assume that their curriculum contents (e.g., taking courses on Web searching) may have had an impact on their selection behaviour. Furthermore, about ten years lay between both studies, so we see the possibility that today's search engine users are more capable in Web searching. We also need to emphasise some methodological differences. The relevance judgements in our study were made by the same participants who took part in our experiment, while Pan *et al.* (2007) chose five non-participants as judges. This is an advantage of our procedure because it enhances the validity of our results.

An issue that should be investigated further is the impact of Google's Universal Search on user behaviour compared to their behaviour on conventional search engine results pages with ten organic results. It would be interesting to see how different forms of results presentation affect users' selection behaviour in different conditions similar to the ones used in our study. For instance, will Universal Search boxes lead users to click these results, even if other, *classic organic* results would be more relevant? Some research at least suggests that Universal Search results have a huge influence on how users examine the organic results on the same results page (Liu *et al.*, 2015).

Conclusion

We presented an eye-tracking study investigating the effect of results order on users' viewing and clicking behaviour. In a between-subjects design, results were presented in one of three conditions: (1) *normal* condition, (2) *swapped* condition (results one and two are swapped), (3) *reversed* condition (reverse order of the first ten results). Results confirmed the huge influence of results ordering on users' fixations and clicking behaviour, as found by Pan *et al.* (2007). Our findings, however, differ in one important regard: although the viewing behaviour was more influenced by the position than by the relevance of a snippet, the crucial factor for a result to be clicked was the relevance and not its position on the search engine results page.

Our study has some limitations. Firstly, the results are not representative. We exclusively used German participants of similar age, most of them students of library and information studies-related subjects. Furthermore, the number of participants was rather small (n=25). While this is a general limitation of most experimental research, the behaviour of our participants could also have been affected by the prescribed queries (i.e., they were not able to formulate their own queries). Furthermore, an experimenter was present during the procedures in the lab. Both could have added to the unnatural setting in the lab, and therefore have influenced the participants (Höchstötter, 2007).

Interesting directions for future research could be investigating the effect of different forms of results presentation on users' viewing and selection behaviour. For instance, will Universal Search boxes lead users to click these results, even if other, *classic organic* results would be more relevant? An experimental design, as presented in this study, would provide valid results on these kinds of questions.

About the authors

Sebastian Schultheiß is a student in the Master's programme *Information, Media, Library* at the Hamburg University of Applied Sciences, Germany. He received his Bachelor's degree in *Library and Information Management* in 2016. He can be contacted at: s.schultheiss1989@gmail.com.

Sebastian Sünkler is a research assistant and lecturer at the Department of Information at the Hamburg University of Applied Sciences, Germany. Since he was a university student he was interested in the subjects of search engines and Information Retrieval. He can be contacted at: Sebastian.Suenkler@haw-hamburg.de.

Dirk Lewandowski is a professor of information research and information retrieval at the Hamburg University of Applied Sciences, Germany. He is the editor of Aslib Journal of Information Management (formerly: Aslib Proceedings), a ISI-ranked information science journal. He can be contacted at: Dirk.Lewandowski@haw-hamburg.de.

References

Bar-Ilan, J., Keenoy, K., Levene, M., & Yaari, E. (2009). Presentation bias is significant in determining user preference for search results - a user study. *Journal of the American Society for Information Science and Technology, 60*(1), 135-149

Blake, C. (2013). Eye-Tracking: Grundlagen und Anwendungsfelder. In W. Möhring &D. Schlütz (Eds.), *Handbuch standardisierte Erhebungsverfahren in der*

Kommunikationswissenschaft (pp. 367-387). Wiesbaden, Germany: Springer Fachmedien Wiesbaden

Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum, 36*(2), 3-10. Retrieved from https://www.cis.upenn.edu/~nenkova/Courses/cis430/p3-broder.pdf (Archived by WebCite® at http://www.webcitation.org/6t6Oxl3EG)

Buscher, G., Dumais, S. T., & Cutrell, E. (2010). The good, the bad, and the random. In Georg Buscher, Susan T Dumais & Edward Cutrell (Eds.), *Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '10* (pp. 42-49). New York, NY: ACM Press

Ferro, N., Kando, N., Fuhr, N., Lippold, M., Järvelin, K., & Zobel, J. (2016). Increasing reproducibility in IR: findings from the Dagstuhl seminar on "Reproducibility of data-oriented experiments in e-science". *SIGIR Forum, 50*(1), 68-82. Retrieved from http://sigir.org/files/forum/2016J/p068.pdf (Archived by WebCite® at http://www.webcitation.org/6t4zlaoj2)

Goel, S., Broder, A., Gabrilovich, E., & Pang, B. (2010). Anatomy of the long tail: Ordinary people with extraordinary tastes. In B. D. Davison, T. Suel, N. Craswell, & B. Liu (Eds.), *Proceedings of the third ACM international conference on Web search and data mining* (pp. 201-210). New York City, NY: ACM. Retrieved from http://snap.stanford.edu/class/cs224w-readings/goel10longtail.pdf (Archived by WebCite® at http://www.webcitation.org/6t4zq0m57)

Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In Laura A Granka, Thorsten Joachims & Geri Gay (Eds.), *Proceedings of the 27th annual International Conference on Research and Development in Information Retrieval - SIGIR '04* (pp. 478-479). New York, NY: ACM Press. Retrieved from https://www.cs.cornell.edu/people/tj/publications/granka_etal_04a.pdf (Archived by WebCite® at http://www.webcitation.org/6t4sphAop)

Grimmelmann, J. (2010). Some skepticism about search neutrality. *The next Digital Decade: Essays on the Future of the Internet, 31*, 435-459. Retrieved from http://james.grimmelmann.net/essays/SearchNeutrality (Archived by WebCite® at http://www.webcitation.org/6t4zzSQpF)

Hariri, N. (2011). Relevance ranking on Google. *Online Information Review, 35*(4), 598-610

Höchstötter, N. (2007). Suchverhalten im Web - erhebung, analyse und möglichkeiten. *Information, Wissenschaft & Praxis, 58*(3), 135-140. Retrieved from http://www2.bui.haw-hamburg.de/pers/ulrike.spree/ws2007_2008/suchverhalten.pdf (Archived by WebCite® at http://www.webcitation.org/6t4tOi7hT)

Höchstötter, N., & Koch, M. (2009). Standard parameters for searching behaviour in search engines and their empirical evaluation. *Journal of Information Science, 35*(1), 45-65

Hotchkiss, G., Alston, S., & Edwards, G. (2005). Importance of the golden triangle. In *Google eye tracking report: how searchers see and click on Google search results* (pp. 1-8). Kelowna, Canada: Enquiro, EyeTools, Did-It. Retrieved from http://searchengineland.com/figz/wp-content/seloads/2007/09/hotchkiss-eye-tracking-2005.pdf (Archived by WebCite® at http://www.webcitation.org/6t4tDjZrD)

Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management, 42*(1), 248-263. Retrieved from https://eprints.qut.edu.au/4945/1/4945_1.pdf (Archived by WebCite® at http://www.webcitation.org/6t4tI9FgV)

Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., & Gay, G. (2007). Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search. *ACM Transactions on Information Systems, 25*(2), 1-27. Retrieved from https://www.cs.cornell.edu/people/tj/publications/joachims_etal_07a.pdf (Archived by WebCite® at http://www.webcitation.org/6t4tMdc1g)

Keane, M. T., O'Brien, M., & Smyth, B. (2008). Are people biased in their use of search engines? *Communications of the ACM, 51*(2), 49-52. Retrieved from http://irserver.ucd.ie/bitstream/handle/10197/1643/MOB.ACM.v3-1.pdf?sequence=3 (Archived by WebCite® at http://www.webcitation.org/6t4tVItJ1)

Latour, B. (2005). *Reassembling the social: an introduction to actor-network-theory*. Oxford: Oxford University Press. Retrieved from http://droit-public.ulb.ac.be/wp-content/uploads/2013/04/Latour_Reassembling.pdf (Archived by WebCite® at http://www.webcitation.org/72B85xU0v)

Lewandowski, D. (2012). Credibility in Web search engines. In M. Folk & S. Apostel (Ed.), *Online credibility and digital ethos: evaluating computer-mediated communication* (pp. 131-146). Hershey, PA: IGI Global. Retrieved from https://arxiv.org/pdf/1208.1011v1.pdf

Lewandowski, D. (2015). *Suchmaschinen verstehen*. [Understanding search engines.] Berlin: Springer Vieweg

Lewandowski, D. (2017). Is Google responsible for providing fair and unbiased results? In M. Taddeo & L. Floridi (Ed.). *The responsibilities of online service providers* (Vol. 31, pp. 61-77). Berlin: Springer. Retrieved http://searchstudies.org/wp-content/uploads/2017/02/Lewandowski_Fair_and_unbiased_results_Preprint-24407.pdf (Archived by WebCite® at http://www.webcitation.org/6t4tkFKka)

Lewandowski, D., Drechsler, J., & Von Mach, S. (2012). Deriving query intents from Web search engine queries. *Journal of the American Society for Information Science and Technology, 63*(9), 1773-1788. Retrieved from http://eprints.rclis.org/17245/1/JASIST_Query_Intents_Preprint.pdf (Archived by WebCite® at http://www.webcitation.org/6t4ts5sqT)

Liu, Z., Liu, Y., Zhou, K., Zhang, M., & Ma, S. (2015). Influence of vertical result in Web search examination. In R. A. Baeza-Yates, M. Lalmas, A. Moffat, & B. A. Ribeiro-Neto (Eds.), *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '15*, 193-202. New York, NY: ACM. Retrieved from http://www.thuir.cn/group/~yqliu/publications/sigir2015-vertical.pdf (Archived by WebCite® at http://www.webcitation.org/6t4tysh6L)

Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., & Newman, J. K. (2004). The determinants of Web page viewing behavior. In A. T. Duchowski & R. Vertegaal (Eds.), *Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004* (Vol. 1, pp. 147-154). New York, NY: ACM Press. Retrieved from https://pdfs.semanticscholar.org/8578/565b9171eaa829a0635e6c7c15366bffe863.pdf (Archived by WebCite® at http://www.webcitation.org/6t4u69QWS)

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication, 12*(3), 801-823. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00351.x/full (Archived by WebCite® at http://www.webcitation.org/6t50YjIzg)

Petrescu, P. (October 2014). Google organic click-through rates in 2014 [Web log post]. Retrieved from https://moz.com/blog/google-organic-click-through-rates-in-2014 (Archived by WebCite® at http://www.webcitation.org/6t4uHEl2u)

Poole, A., & Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research: current status and future prospects. In *Encyclopedia of human-computer interaction* (pp. 211-219). Retrieved from https://www.researchgate.net/publication/230786738_Eye_tracking_in_human-computer_interaction_and_usability_research_Current_status_and_future_prospects (Archived by WebCite® at http://www.webcitation.org/72B9L3jvG)

Purcell, K., Brenner, J., & Rainie, L. (2012). *Search engine use 2012*. Washington, DC: Pew Research Centre. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf (Archived by WebCite® at http://www.webcitation.org/6t4wVe3Gz)

Röhle, T. (2010). *Der Google-Komplex: über Macht im Zeitalter des Internets*. [The Google complex: power in the age of the Internet]. Bielefeld, Germany: Transcript.

Schaer, P., Mayr, P., Sünkler, S., & Lewandowski, D. (2016). How relevant is the long tail? In N. Fuhr, P. Quaresma, T. Gonçalves, B. Larsen, K. Balog, C. Macdonald, ... N. Ferro (Eds.), *CLEF 2016* (Vol. 9822, pp. 227-233). Cham, Germany: Springer International Publishing. Retrieved from https://arxiv.org/pdf/1606.06081.pdf

Tobii AB. (2016). *Tobii Studio user's manual version 3.4.5*. Stockholm: Tobii AB. Retrieved from https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf (Archived by WebCite® at http://www.webcitation.org/72BF1bqPq)

Tobii Technology. (2010). Tobii eye tracking - an introduction to eye tracking and Tobii eye trackers. Stockholm: Tobii Technology. Retrieved from https://de.scribd.com/document/26050181/Introduction-to-Eye-Tracking-and-Tobii-Eye-Trackers (Archived by WebCite® at http://www.webcitation.org/6t4wv9105)

Tremel, A. (2010). *Suchen, finden - glauben? Die Rolle der Glaubwu?rdigkeit von Suchergebnissen bei der Nutzung von Suchmaschinen*. [Search, find - believe? The role of credibility of search results when using search engines]. Unpublished doctoral dissertation. Ludwig-Maximilians-Universität, Munich, Germany. Retrieved from https://edoc.ub.uni-muenchen.de/12418/1/Tremel_Andreas.pdf (Archived by WebCite® at http://www.webcitation.org/6t4x7tUWR)

Westerwick, A. (2013). Effects of sponsorship, Web site design, and Google ranking on the credibility of online information. *Journal of Computer-Mediated Communication, 18*(2), 194-211. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/jcc4.12006/full (Archived by WebCite® at http://www.webcitation.org/6t50Febtg)

How to cite this paper

Schultheiß, S., Sünkler, S., & Lewandowski, D. (2018). We still trust in Google, but less than 10 years ago: an eye-tracking study *Information Research, 23*(3), paper 799. Retrieved from http://InformationR.net/ir/23-3/paper799.html (Archived by WebCite® at http://www.webcitation.org/72M895A36)

Find other papers on this subject

Check for citations,