Intl. Journal of Human–Computer Interaction, 30: 177–191, 2014

ISSN: 1044-7318 print / 1532-7590 online
DOI: 10.1080/10447318.2013.846790

Taylor & Francis
Taylor & Francis Group

The Role of Search Result Position and Source Trustworthiness in the Selection of Web Search Results When Using a List or a Grid Interface

Yvonne Kammerer and Peter Gerjets
Knowledge Media Research Center, Tuebingen, Germany

Previous research indicates that web users rely to a great extent on the ranking provided by the search engine and predominantly access the first few web pages presented. In case that the information sources presented in the top of the search engine results page (SERP) are of rather low trustworthiness, this might lead to a biased or incomplete view of the topic—especially when dealing with controversial issues. Study 1, thus, systematically investigated whether participants who were asked to search for an unfamiliar and controversial medical issue accessed fewer trustworthy information sources and consequently included less information from trustworthy pages in their argumentation when the search results were ranked from least to most trustworthy on a Google-like SERP than when they were ranked from most to least trustworthy. Results from Study 1 confirmed these assumptions. Furthermore, Study 2 showed that when the same materials were presented in a grid interface, the impact of the position of the search results on their selection was substantially reduced. Irrespective of whether the most trustworthy search results were presented in the top or the bottom row of the grid interface, users predominantly selected the most trustworthy search results from the SERP and included the same amount of information from trustworthy pages in their argumentation.

1. INTRODUCTION

Search engines such as Google, Yahoo!, and Bing have become an integrated part of our personal and professional lives. According to latest reports more than 80% of online adults use search engines at least once a week (e.g., Purcell, Brenner, & Rainie, 2012), and billions of queries are submitted to search engines every day. Search engines help laypersons (i.e., non-experts in a domain) to sort through the huge amounts of information on the web and to find documents relevant to their current information needs. Besides being utilized for searching for simple and uncontroversial facts or products to purchase, they are increasingly used to acquire knowledge about more complex and controversial issues in critical domains such as health, politics, or finance (Estabrook, Witt, & Rainie, 2007; Fox, 2006, 2011; Smith, 2009). For instance, web searches for medical and healthcare information increasingly supplement the interaction with experts (Fox & Jones, 2009; Morahan-Martin, 2004; Stadtler, Bromme, & Kettler, 2009). Considering the potential influence of web information on important personal decisions, the trustworthiness of information sources becomes a pivotal issue. The notion of source trustworthiness is typically embedded in the concept of trust, which is "the firm belief in the competence of an information source to provide accurate and correct information within a specific context" (Houmb, Ray, & Ray, 2006, p. 137).

However, on the web not only official institutions but also companies, journalists, and laypeople provide information about controversial topics. These information providers might have other goals than just providing neutral facts, such as selling or promoting a particular product, lobbying in favor of a particular policy, or exchanging experiences or stating opinions. As a consequence, information sources on the web vary considerably in terms of their trustworthiness, that is, their motivation and ability to provide reliable and accurate information. Moreover, even the information contained in web pages listed among the top search results by a search engine might turn out to be one-sided or commercially biased (Lewandowski, 2011; Mansell & Read, 2009). For instance, this can be due to search engine optimization businesses (cf. Lewandowski, 2013) and the so-called richer-get-richer effect (i.e., highly ranked pages typically benefit from a greater volume of traffic, receive more incoming links, and in turn get higher positions in the search engine ranking; see, e.g., Cho & Roy, 2004).

Accordingly, when making decisions about which search results to select from a search engine results page (SERP), web users should critically evaluate the trustworthiness of the information sources (e.g., Braasch et al., 2009; Rieh, 2002; Taraborelli, 2008). In other words, based on source cues provided in the search results (e.g., the name of the web site, the type of source, or the URL) they should select web pages that are expected to contain trustworthy, that is, reliable and

Address correspondence to Yvonne Kammerer, Knowledge Media Research Center, Schleichstr. 6, Tuebingen 72076, Germany. E-mail: y.kammerer@iwm-kmrc.de

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

unbiased, information (Rieh, 2002). This seems particularly important when search engines are used to acquire knowledge about more complex and controversial issues that do not have a single and definite solution but that are characterized by various competing perspectives and arguments.

Previous empirical findings, however, indicate that not only when searching for simple facts but also when searching for more complex information about controversial topics, laypersons often rely to a great extent on the ranking provided by the search engine and predominantly access the first few web pages presented (e.g., Cutrell & Guan, 2007; Gerjets, Kammerer, & Werner, 2011; Granka, Joachims, & Gay, 2004; Guan & Cutrell, 2007; Nakamura et al., 2007; Pan et al., 2007; Zhang, Cole, & Belkin, 2011). In case that the information sources presented in the top of the list are of rather low trustworthiness, this might lead to a biased or incomplete view of the controversial issue. In Study 1, thus, we systematically investigated, whether participants who were asked to search for an unfamiliar and controversial medical issue—in the context of limited time and a range of alternatives to choose from (cf. Pengnate & Antonenko, 2013)—would select fewer trustworthy search results when these were presented toward the bottom of a Google-like SERP list and as a consequence would include less information from trustworthy pages in their argumentation than when the most trustworthy search results were presented in the top of the list. Furthermore, in Study 2 we examined whether a search engine interface that presents search results in a format different from a single vertical list would reduce the influence of the ranking position of the search results, and thus would lead to a predominant selection and use of the most trustworthy search results in the SERP, irrespective of their position. To examine the impact of the ranking position and of source trustworthiness on users' evaluation and selection behavior on SERPs, in both studies we used a rich multimethod approach including eye-tracking methodologies and log file data as process measures as well as search outcome measures (cf. Keane, O'Brien, & Smyth, 2008; Pan et al., 2007). That is, in the present article we focused on implicit behavioral measures rather than on explicit trustworthiness ratings.

The remainder of this article is structured as follows: Section 2 provides an overview of the theoretical background and related work on how users evaluate and select search results during web search and on how alternative search engine interfaces affect the selection of search results. In section 3, we state our research aims and hypotheses of the two studies. After that, in sections 4 and 5, the method and results of each of the two studies are presented and discussed. Finally, in the concluding section 6, we sum up the outcomes of our research and give some directions for future work.

2. THEORETICAL BACKGROUND AND RELATED WORK

2.1. Evaluating and Selecting Search Results

Whereas in the past two decades much research attention has been paid on users' evaluation and judgment of website trustworthiness (for reviews, e.g., see Pengnate & Antonenko, 2013; Sillence, Briggs, Harris, & Fishwick, 2006), less is known about users' trustworthiness evaluations of web search results (cf. Taraborelli, 2008). The assessment of search results is an early evaluation step that seems crucial for web-based information seeking, because from the large number of web pages usually classified as relevant by a search engine, only a very limited set can be inspected in greater detail (cf. Braasch et al., 2009; Rieh, 2002).

In terms of information-processing and decision theory (e.g., Payne & Bettman, 2004; Payne, Bettman, & Johnson, 1993) the evaluation and selection of search results during web search can be described as a situation that requires choosing between a large number of alternatives. Moreover, for these alternatives usually only sparse information—namely, a title, an excerpt from the respective web page, and its URL—are provided (cf. Wirth, Böcking, Karnowski, & von Pape, 2007). In such decision situations of high uncertainty, individuals typically aim at maximizing the outcome (i.e., gaining valuable information) while minimizing time and cognitive effort—particularly under conditions of limited time and knowledge (Gigerenzer & Goldstein, 1996). Thus, in trying to find an optimal trade-off between cognitive effort and efficient outcome, instead of a systematic, that is, thorough and complete knowledge-based evaluation of all given information individuals often evaluate information in a rather heuristic way (Metzger, 2007; Metzger, Flanagin, & Medders, 2010; Taraborelli, 2008). That is, they typically rely on *heuristic cues* to decide which alternatives to select (e.g., Hilligoss & Rieh, 2008; Metzger et al., 2010; Taraborelli, 2008; Wirth et al., 2007). Such cues, for instance, can be the ranking position of the search result in the SERP (Wirth et al., 2007) or source cues such as the type or reputation of the website indicating the trustworthiness of an information source (Hilligoss & Rieh, 2008; Rieh, 2002). According to the prominence-interpretation theory by Fogg (2003) in particular, *prominent*, that is, easily noticeable cues will be considered in users' evaluations. In addition, a second component that affects a user's evaluation according to the theory is the user's personal *interpretation* of the cue that has been noticed. Whereas the theory originally addressed the assessment of web pages, in our opinion its assumptions can also be applied to the evaluation of search results on SERPs.

Previous research has shown by means of eye tracking and log file analyses that—in the context of limited time and a range of search results to choose from—the ranking position of the search results in the list serves as a strong cue for selection. Users who were asked to search for simple, undisputed facts have been shown to spend most attention to the search results at the top of the first SERP and to predominantly select these links as well as to enter new search terms instead of continuing to examine further SERPs (Cutrell & Guan, 2007; Granka et al., 2004; Guan & Cutrell, 2007). Keane et al. (2008) and Pan et al. (2007) showed that this was even the case when the top search results were the least relevant ones of the SERP. In their

experiments, they had systematically manipulated the order of the 10 first Google search results, presenting them either in the regular Google order or in a systematically reversed order. In the latter case, participants visually inspected more search results before they made a selection decision but generally still paid most attention to the search results on top of the SERP and selected these results most often. As a consequence, in the reversed SERP order participants also showed a lower task performance in terms of finding a page with the correct answer than in the normal SERP order.

Selecting a search result that links to a page with the correct answer, however, still seems to be a rather easy task as compared to the search for information about controversial topics that do not have a single and definite solution (cf. Tu, Shih, & Tsai, 2008). As for this latter type of search tasks, a vast body of conflicting evidence exists that typically is also present in the various web pages returned by a search engine, to solve the task, that is, to get adequately informed about the issue, users need to collect information from multiple web pages (Aula & Russell, 2008). Moreover, in critical domains such as health, politics, or finance, where people might use information found online as the basis for their decisions, a critical evaluation of the trustworthiness of information sources seems to be of particular importance.

Interview and survey studies have shown that web users of different educational backgrounds principally seem to be aware of the importance of trustworthiness evaluations when searching for controversial medical or academic issues on the web (Eysenbach & Köhler, 2002; Flanagin & Metzger, 2010; Rieh & Hilligoss, 2008). However, studies that have examined individuals' actual use of search engines when seeking information on unfamiliar topics indicate that they rely heavily on the ranking of the search engine and—as when searching for simple facts—predominantly select the first few search results (e.g., Bar-Ilan, Keenoy, Levene, & Yaari, 2009; Eysenbach & Köhler, 2002; Gerjets, Kammerer, & Werner, 2011; Hargittai, Fullerton, Menchen-Trevino, & Thomas, 2010). The type or reputation of the website instead seems to be rather seldom considered when deciding which search results to select from the SERPs, as indicated by verbal protocols (e.g., Bar-Ilan et al., 2009; Gerjets et al., 2011; Hargittai et al., 2010).

However, to systematically investigate to what extent the position of the search results versus the source trustworthiness play a role in users' spontaneous[1] selection decisions during web search, the *trustworthiness order* of the search results needs to be experimentally varied, with the search results linked to the most trustworthy sources being presented toward either the top or the bottom of the SERP. This was done in the present article. Specifically, in Study 1 of this article we investigated whether participants who were asked to search for an unfamiliar and controversial medical issue selected fewer trustworthy search results when these were presented toward the bottom of a Google-like SERP list and as a consequence included less information from trustworthy pages in their argumentation than when the most trustworthy search results were presented in the top of the list.

[1]With spontaneous selection decisions, we mean that participants are not prompted to select the best pages or to explain why they select certain pages (cf. Gerjets et al., 2011).

2.2. Alternative Search Engine Interfaces

One reason for the strong influence of the ranking position of search results on their selection might be the list format itself in which conventional search engines typically present their results. In terms of Fogg's (2003) prominence-interpretation theory in such a format the position of search results constitutes a prominent cue. In contrast, a search engine interface that presents search results in a format different from a single vertical list might reduce the influence of the position of the search results on users' selections. Although to this date only few empirical studies have examined the effects of alternative search engine interfaces on the selection of search results, their findings are promising. Resnick, Maldonado, Santos, and Lergier (2001) tested a conventional list interface against a tabular interface in which the columns of a table corresponded to the different elements of the search results (title, excerpt, URL, and metadata). In two fact-finding tasks participants were instructed to select the search result that best met the objective of the search query from a predefined SERP. Results showed that, whereas in the list interface the majority of the participants selected the first search result that they found appropriate, in the tabular interface at least half of the participants chose a more exhaustive approach examining more (or all) search results before they made the selection.

In regard to more complex search tasks about science-related issues, Salmerón, Gil, Bråten, and Strømsø (2010) compared university students' navigation behavior when using a standard Google-like list interface or a graphical-overview interface similar to the search engine Kartoo (closed down in January 2010) in which search results were displayed by means of a graphical overview indicating the semantic relationships between the search results. Navigation results revealed that in the graphical overview interface students' navigation patterns were significantly more heterogeneous than in the list interface. Whereas in the list interface students adhered to a linear top-to-bottom navigation pattern, the graphical overview interface supported a more free selection sequence of the available search results. In addition, the exploration of the web pages by using the graphical overview interface resulted in better integration of information from different web pages than when working with the standard list interface. Finally, a study by Kammerer and Gerjets (2012) showed that a tabular interface that presented search results grouped according to objective, subjective, or commercial information resulted in decreased visual attention to and selection of commercial search results and in increased selection of objective search results than a

standard list interface, when university students searched for a controversial medical issue.

Recently some search engines have emerged that use a grid interface displaying search results in a grid with n-by-n cells (e.g., Viewzi[2]; http://www.horizobu.com). In such a grid format, the ranking position of the search results is assumed to be less prominent, as it remains unclear to the user whether the ranking within a SERP is aligned horizontally (i.e., line by line, according to the regular western reading direction) or vertically (i.e., column by column, according to a list structure), or whether there exists a ranking at all. Therefore, in Study 2 of this article, we examined whether a grid interface would result in a predominant selection and use of the most trustworthy search results in the SERP, irrespective of their position in the SERP (i.e., irrespective of the trustworthiness order), when participants searched for an unfamiliar and controversial medical issue.

3. AIMS AND HYPOTHESES

Study 1 aimed at expanding the findings by Keane et al. (2008) and Pan et al. (2007) to a web search scenario in which (a) laypeople search for a controversial medical issue (i.e., the treatment of Bechterew's disease) instead of simple facts and (b) the "trustworthiness order" of search results is systematically varied. For the latter purpose, search results were presented either in an *optimal trustworthiness order* with the search results being ordered from the most trustworthy search results at the top of a Google-like list to the least trustworthy search result at the bottom of the list, or in a *reversed trustworthiness order*, that is, from the least trustworthy search result at the top to the most trustworthy ones at the bottom. The latter, reversed trustworthiness order represented the critical condition in which simply relying on the position of the search results in a SERP would result in the selection of search results that were of rather low trustworthiness.

We expected the following differences between a SERP with optimal trustworthiness order and a SERP with reversed trustworthiness order: First, in line with results from the effects of a systematically reversed "Google order" in fact-finding tasks (Pan et al., 2007) we hypothesized that in a SERP with reversed trustworthiness order, with the search result at the top being the least trustworthy ones, participants would visually inspect more search results before making their first selection decision than in a SERP with optimal trustworthiness order (Hypothesis 1). If participants exclusively relied on the position of the search results in the SERP, in contrast, no such differences should be found between the two trustworthiness orders. However, because of the expected strong influence of the ranking position (cf. Keane et al., 2008; Pan et al., 2007), we hypothesized that in a SERP with reversed trustworthiness order in the context of limited time and a range of alternatives to choose from participants would select more of the least trustworthy search results (at the top of the list) and less of the most trustworthy results (at the bottom of the list) during their web search than in a SERP with optimal trustworthiness order (Hypothesis 2). Likewise, in a SERP with reversed trustworthiness order participants were expected to spend more time reading the least trustworthy web pages and less time reading the most trustworthy web pages (Hypothesis 3). As a consequence, they were assumed to also have lower search outcomes in terms of arguments listed from the most trustworthy sources than in a SERP with optimal trustworthiness order (Hypothesis 4).

In Study 2 we tested the general assumption that in a search engine interface that does not present search results in a conventional list interface (as in Study 1) but in a grid interface, the impact of the ranking position of the search results on their selection would be substantially reduced, and thus no difference between trustworthiness orders should be found. Accordingly, we hypothesized that in both trustworthiness orders participants would scan a similar (rather high) number of search results before making their first selection decision (Hypothesis 1). Furthermore, we hypothesized that because of the reduced impact of the position of the search results in a grid SERP, in both trustworthiness orders the most trustworthy search results would be predominantly selected during web search, without any differences between trustworthiness orders (Hypothesis 2). Likewise, participants were expected to spend most time reading the most trustworthy web pages, without any differences between trustworthiness orders (Hypothesis 3). As a consequence, we hypothesized that in the reversed trustworthiness order search outcomes in terms of arguments listed from the most trustworthy sources would be as high as in the optimal trustworthiness order (Hypothesis 4).

4. STUDY 1

4.1. Method

Participants. Participants were 40 university students (10 male, 30 female; $M = 24.75$ years, $SD = 4.22$) from social and natural sciences and humanities at a large German university; participation was rewarded with either course credit or payment. Pharmacy and medical students were excluded from participation, as we focused on medical laypersons in this study. Participants had normal or corrected-to-normal vision. All participants reported to use Google as their primary search engine.

Task and materials. The task scenario used in the experiment was to seek information on the World Wide Web about two competing therapies ("radon therapy" and "infliximab therapy") for Bechterew's disease (i.e., a chronic inflammatory rheumatic disease affecting the spine) in order to give informed advice to a fictitious friend who was recently diagnosed with the disease (cf. Kammerer & Gerjets, 2012; Stadtler & Bromme, 2008). After their web research, participants were asked to list arguments in favor or against the therapies that might form the basis of a decision.

[2] See, for example, http://zootool.com/watch/ej4en3/. The Viewzi search engine was closed down in January 2011.

For their web research, participants were provided with two mock Google-like SERP lists, one for each therapy. Each of the SERPs contained nine search results (for an example screenshot, see Figure 1). Participants could access all web pages associated with the search results. The search results or web pages, respectively, were all relevant to the search topic in regard to the content of information provided. All search results included the keywords *bechterew's disease* and *radon* or *infliximab*, respectively, and addressed the therapy or treatment of Bechterew's disease and its effectiveness or side effects. The search results, however, differed in their trustworthiness reflecting the given heterogeneity of information sources on the web. Each SERP included web pages provided by official institutions and specialist media (e.g., Department of Health, medical magazine), industry and companies (e.g., health farms or drug companies), and laypeople (e.g., forum pages). Web materials were presented with Microsoft Internet Explorer 7 but were actually stored locally to guarantee a standardized and controlled experimental setting (cf. Wiley et al., 2009). Search result links that had already been selected were marked in purple color; not-yet-selected links were displayed in blue.

Experimental design. The *trustworthiness order* of the SERPs was varied as between-subjects factor, presenting the search results on a SERP either in an *optimal trustworthiness order* or in a *reversed trustworthiness order*. Trustworthiness order was defined in a pilot study where 24 university students had received two lists of search results (one list for "radon therapy" and one list for "infliximab therapy," with nine search results each) that were presented in random order to the participants. Participants' task was to order the search results in each list according to the expected trustworthiness of the corresponding web pages from 1 (*most trustworthy*) to 9 (*least trustworthy*). Based on the mean ranks two different trustworthiness orders were constructed for the SERPs: an optimal trustworthiness order, with the search results rated as most trustworthy being presented at the top of the list and those rated least trustworthy being presented at the bottom, and a reversed trustworthiness order with the search results rated as least trustworthy being presented at the top of the list and those rated as most trustworthy at the bottom. Search results that were ranked as most trustworthy linked to sites from official institutions or specialist media, whereas the search results of lower trustworthiness were predominantly those that linked to commercial websites or to contributions in forums (cf. Pengnate & Antonenko, 2013). Participants were randomly assigned to the two experimental conditions, with 20 participants serving in each condition.

In addition, *search result trustworthiness* was considered as a within-subjects factor, differentiating the search results into the three most-trustworthy search results, the three medium-trustworthy search results, and the three least-trustworthy search results in a SERP. The three search result categories differed significantly in their average rankings, $F(2, 22) = 70.59$, $p < .001$, partial $\eta^2 = .87$. Most trustworthy search

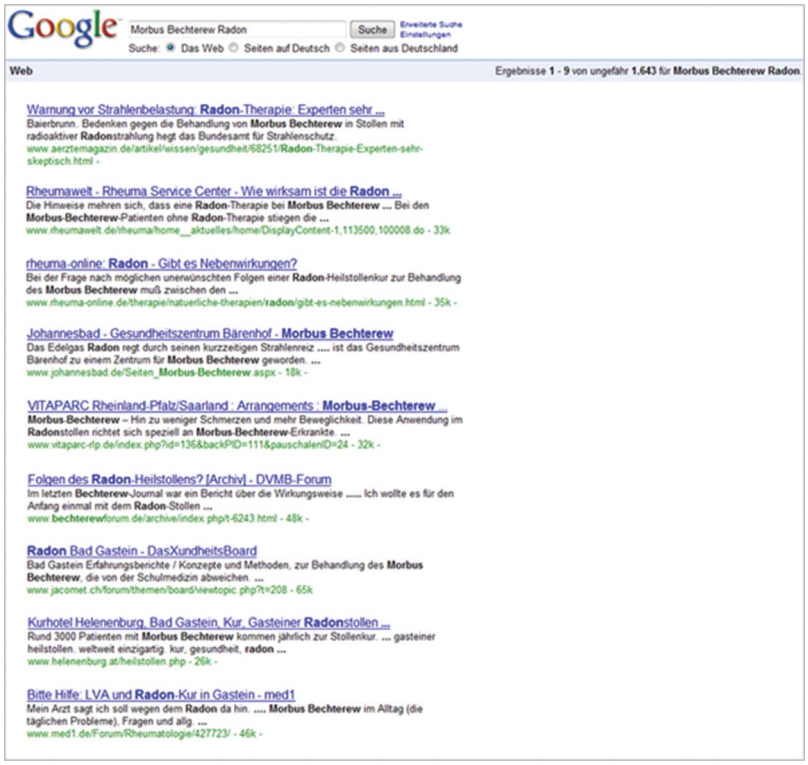

FIG. 1. Example screenshot of a list interface.

results with an average ranking of 3.19 ($SD = 0.82$) were significantly higher ranked (both $ps < .001$) than medium trustworthy search results with an average rank of 5.24 ($SD = 0.55$) and least trustworthy search results with an average rank of 6.59 ($SD = 0.66$). Medium trustworthy and least trustworthy search results also differed significantly from each other ($p < .001$).

Control variables. Demographics (gender, age), computer and web search experience (six items, Cronbach's $\alpha = .85$), and prior knowledge on Bechterew's disease, other rheumatic diseases, and respective therapies (10 items, Cronbach's $\alpha = .64$) were assessed as control variables (see Table 1 for means and standard deviations). Except for gender and age, items had to be rated on 5-point Likert-type response scales ranging from 1 (*totally disagree* or *very low*) to 5 (*totally agree* or *very high*). Analyses of the respective data revealed no differences between the two experimental conditions, that is, for gender, $\chi^2(1, N = 40) = 0.53$, $p = .465$; for age, $t(38) = 0.15$, $p = .884$; for computer and web search experience, $t(38) = –0.56$, $p = .580$; and for prior domain knowledge, $t(38) = –0.94$, $p = .352$.

Procedure. Participants were tested in individual sessions of approximately 1 hr. First, they were given a brief written introduction to the study including the consent form and were then administered a computer-based questionnaire to assess the control variables. Then, they underwent a practice task for approximately 2 min to get acquainted with the experimental setup. This practice task, which was about two competing weight loss methods, was structured in the same way as the subsequent main task. After the practice task, participants received the instruction for the main task (i.e., a request from a fictitious friend asking for advice regarding the two therapies) as well as a brief description of the Bechterew's disease (e.g., symptoms, causes, course of the disease) to ensure that participants did not spend parts of their web search to inform themselves about these issues rather than the two therapies. Participants were told that for each SERP they had 4 min (cf. Pengnate & Antonenko, 2013) to inform themselves about the respective therapy. Then, they were calibrated on the eye-tracking system and started their web search regarding the first of the two therapies. After 4 min, the information search regarding the first therapy was interrupted and the search for the second therapy started. The order in which participants were provided with the two SERPs was counterbalanced across participants.

Participants could access all web pages corresponding to the 18 search results presented. However, to ensure a controlled experimental setting they were not allowed to generate new SERPs by changing the search terms. Thus, as opposed to searching the open web where users can repeatedly refine their search terms and have access to an enormous amount of web pages, the experimental setting of the present study provided participants with a restricted information space. Therefore, the total search time was limited to 8 min to ensure that participants preselected and focused only on some of the available information sources—like they would in a real search situation on the web—rather than reading all available information (cf. Klöckner, Wirschum, & Jameson, 2004; Pengnate & Antonenko, 2013).

After finishing the search task, participants were given 5 min to list arguments for or against each of the two therapies that might form the basis of a decision to undergo one or the other therapy.

Dependent variables. Participants' eye movements during task processing were recorded by a 50 Hz Tobii 1750 remote eye-tracking system, supported by the software ClearView 2.7.1. For each of the nine search results on a SERP a polygonal "area of interest" covering the title, excerpt, and URL was defined to determine whether a participant was looking at a search result. Following Pan et al. (2007) we considered a search result as "viewed" by the participant if there was at least one fixation within a defined area of interest covering the search result. The minimum fixation duration was set to 80 ms with a fixation radius of 30 pixels. To test Hypothesis 1 that in a SERP with reversed trustworthiness order participants would visually inspect more search results before making their first selection decision than in a SERP with optimal trustworthiness order, the *number of search results fixated before accessing the first web page* was determined (cf. Pan et al., 2007).

The ClearView 2.7.1 analysis software also captured which search results participants selected during search and how much time they spent on each web page. To test Hypothesis 2 that in a SERP with reversed trustworthiness order participants would select more of the three least trustworthy search results and fewer of the three most trustworthy search results than in a SERP with optimal trustworthiness order, the *number of search results selected* from a SERP was analyzed as a function of search result trustworthiness. Likewise, to test Hypothesis 3 the *time spent on the web pages* as a function of their trustworthiness was analyzed, that is, how much time participants spent on the three most trustworthy, the three medium trustworthy, and the three least trustworthy web pages of a SERP. For all three dependent variables data for both SERPs (i.e., the "radon SERP" and the "infliximab SERP") were collapsed.

TABLE 1
Means (and Standard Deviations) for Control Variables as a Function of Trustworthiness Order in a List Interface

Variable	Optimal TO	Reversed TO
Age	24.85 (5.07)	24.65 (3.38)
Gender	4 m, 16 f	6 m, 14 f
Computer and web search experience, 1 (*low*) to 5 (*high*)	3.27 (0.76)	3.38 (0.55)
Prior domain knowledge, 1 (*low*) to 5 (*high*)	1.29 (0.36)	1.41 (0.44)

Note. TO = trustworthiness order; m = male; f = female.

Finally, as the search outcome, the arguments participants listed in favor and against the two therapies were analyzed. Specifically, the *number of arguments from the three most trustworthy pages* listed by the participants were coded. In total we defined 22 statements from these pages that could be used as arguments, such as "radon therapy reduces joint inflammation," "infliximab improves patients' spinal mobility," "long-term effects of radon therapy are unknown," or "infliximab therapy increases the risk of infections." Two raters familiar with the search task and the web materials as well as with the list of arguments scored the arguments of 30% of the participants by classifying them as one of the arguments or by identifying an argument as a false or nonargument. Interrater reliability computed on this subsample of protocols yielded a Cohen's kappa of $\kappa = .84$. Disagreements were resolved through discussion between the raters. One rater coded the remaining argument lists.

4.2. Results

An alpha level of .05 (two-sided) was used for the statistical tests reported. To test the hypotheses, analyses of variance (ANOVAs) with trustworthiness order (optimal vs. reversed) as a between-subject factor, and, when applicable, search result trustworthiness (most trustworthy, medium trustworthy, and least trustworthy search results) as a within-subject factor were conducted. In case of significant effects, ANOVAs were followed up by Bonferroni post hoc tests. For means and standard deviations of the dependent variables see Table 2.

With regard to the *number of search results participants fixated before they accessed the first web page,* the ANOVA showed that in line with Hypothesis 1, participants who were presented SERPs with a reversed trustworthiness order fixated significantly more search results before they accessed the first web page than participants who were presented SERPs with an optimal trustworthiness order, $F(1, 38) = 12.74$, $p = .001$, partial $\eta^2 = .25$. Whereas in SERPs with optimal trustworthiness order only 5% of the participants inspected eight or all nine results before the first selection, in SERPs with reversed trustworthiness order it were 35%. This difference was significant, $\chi^2(1, N = 40) = 5.63$, $p = .018$.

During their web search participants, on average, selected 49.6% of the available search results. With regard to the *number of search results selected* as a function of their trustworthiness, a significant effect of search result trustworthiness was found, $F(2, 76) = 18.98$, $p < .001$, partial $\eta^2 = .33$. This effect was qualified by a significant interaction with trustworthiness order, $F(2, 76) = 12.44$, $p < .001$, partial $\eta^2 = .25$. As expected in Hypothesis 2, in SERPs with reversed trustworthiness order significantly fewer of the most trustworthy search result were selected ($p < .001$) and significantly more of the least trustworthy search results ($p = .020$) than in SERPs with optimal trustworthiness order (see Figure 2). The medium-trustworthy search results, which were in the middle of the list in both trustworthiness orders, were selected to a similar extent in both conditions ($p = .843$). The results clearly indicate that the ranking position of a result had substantial influence on whether the users selected it.

Yet it should be noted that in SERPs with reversed trustworthiness order the three most trustworthy, medium trustworthy, and least trustworthy search results were all selected to the same extent (all ps $> .882$). This indicated that the trustworthiness of the search results also played a role in their selection. Otherwise, the three least trustworthy search results at the top of the list should have been predominantly selected. In SERPs with optimal trustworthiness order, in contrast, the most trustworthy search results, which were at the top of the list, were selected significantly more often than both the medium trustworthy and least trustworthy search results (both ps $< .001$), which did not differ from each other ($p = .405$).

With regard to the *time spent on web pages* as a function of their trustworthiness, the ANOVA also showed a significant effect of search result trustworthiness, $F(1.67, 63.59) = 40.86$, $p < .001$, partial $\eta^2 = .52$ (Greenhouse-Geisser corrected), with this effect again being qualified by a significant interaction

TABLE 2
Means (and Standard Deviations) of the Dependent Variables as a Function of Trustworthiness Order in a List Interface

Dependent Variables	Optimal TO	Reversed TO
Number of search results fixated before first click	2.85 (2.00)	5.33 (2.37)
Number of most trustworthy search results selected per SERP	2.53 (0.47)	1.53 (0.62)
Number of medium trustworthy search results selected per SERP	1.25 (0.72)	1.30 (0.86)
Number of least trustworthy search results selected per SERP	0.93 (0.57)	1.40 (0.66)
Time (in s) spent on most trustworthy web pages per SERP	129.89 (37.73)	88.52 (40.19)
Time (in s) spent on medium trustworthy web pages per SERP	30.90 (21.04)	41.24 (29.02)
Time (in s) spent on least trustworthy web pages per SERP	29.88 (24.25)	59.76 (39.44)
Number of arguments from most trustworthy sources	5.85 (2.35)	4.00 (1.97)

Note. TO = trustworthiness order; SERP = search engine results page.

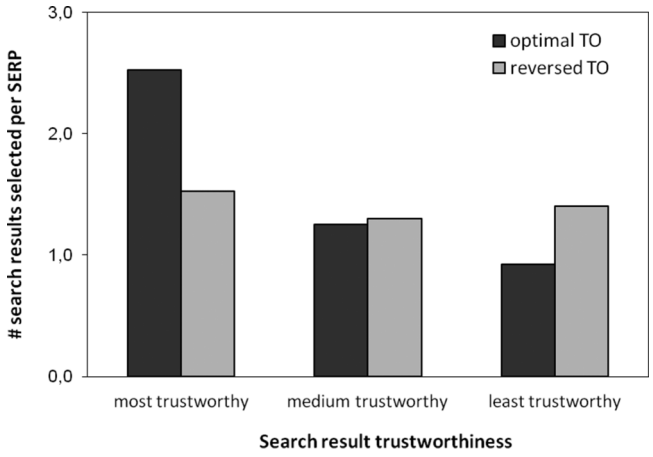

FIG. 2. Number of search results selected per search engine results page (SERP) as a function of search result trustworthiness and trustworthiness order (TO) in a list interface.

with trustworthiness order, $F(1.67, 63.59) = 8.68$, $p = .001$, partial $\eta^2 = .19$ (Greenhouse-Geisser corrected). As expected in Hypothesis 3, participants who were presented SERPs with reversed trustworthiness order spent significantly less time on the three most trustworthy web pages ($p = .002$) and significantly more time on the three least trustworthy pages ($p = .006$) than participants who were presented SERPs with optimal trustworthiness order (see Figure 3). On the medium trustworthy pages participants spent a similar amount of time in both conditions ($p = .205$).

Furthermore, in SERPs with optimal trustworthiness order significantly more time was spent on the most trustworthy pages than both on the medium trustworthy and the least trustworthy pages (both ps $< .001$). In SERPs with reversed trustworthiness order, in contrast, the time spent on the most trustworthy pages and the least trustworthy pages did not differ significantly ($p = .171$).

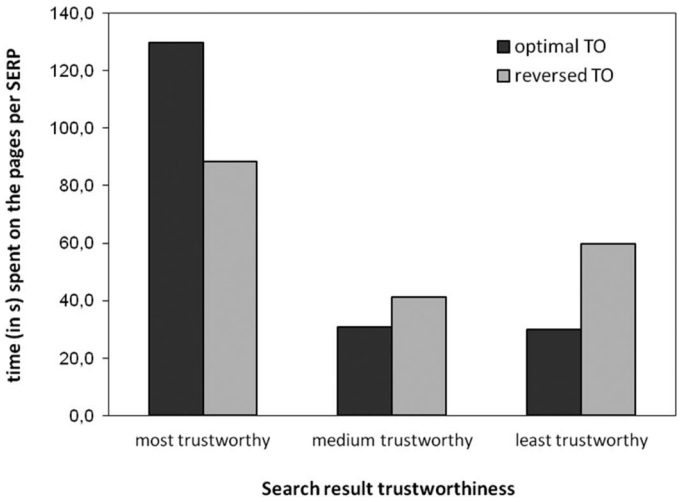

FIG. 3. Time (in seconds) spent on the web pages of a search engine results page (SERP) as a function of search result trustworthiness and trustworthiness order (TO) in a list interface.

With regard to the *number of arguments listed from the three most-trustworthy sources*, the ANOVA showed a significant main effect of trustworthiness order, $F(1, 38) = 7.28$, $p = .010$, partial $\eta^2 = .16$. Participants who had received SERPs with optimal trustworthiness order listed significantly more of these arguments than participants who had received SERPs with reversed trustworthiness order, as it was expected in Hypothesis 4.

4.3. Discussion

The results of Study 1 expand previous findings by Keane et al. (2008) and Pan et al. (2007), showing that also when laypeople search for information on an unfamiliar and controversial medical issue in the context of limited time and a range of search results to choose from, the position of search results in a SERP list has a strong influence on users' selection decisions. In line with our hypotheses, when the top search results were the least trustworthy ones participants selected more of these least trustworthy search results and fewer of the most trustworthy results (Hypothesis 2), spent more time on the least trustworthy web pages and less time on the most trustworthy web pages (Hypothesis 3), and as a consequence after the web search listed fewer arguments from the most trustworthy sources (Hypothesis 4) than when the search results in the top of the SERP list were the most trustworthy ones. To conclude, the results indicate that participants evaluated the search results in a heuristic way (cf. Metzger, 2007; Metzger et al., 2010; Taraborelli, 2008), only inspecting a subset of search results before their first selection and in general relying to a certain extent on the search result position as a selection cue (cf. Wirth et al., 2007).

However, it should be noted that participants did not blindly rely on the position of the search results in the list, but the trustworthiness of the search results also seemed to play at least some role in participants' selection decisions: As expected in Hypothesis 1 when the top search results were the least trustworthy ones in the list, participants visually inspected more search results before they accessed the first web page than when the top search results were the most trustworthy ones. This indicates some awareness of the conflict between search result position and source trustworthiness. In addition, participants did not predominantly select the least trustworthy results at the top of the list but selected all search results to a similar extent.

In Study 2 we tested whether the influence of source trustworthiness on users' selection of search results could be further increased, or in other words, whether the influence of the position of the search results could be reduced, when search results were presented in a grid interface similar to the search engines Viewzi or Horizobu instead of a conventional list interface as in Study 1. As in such a grid interface it is unclear to the

user whether the ranking within a SERP is aligned horizontally from left to right (i.e., line by line) or vertically from top to bottom (i.e., column by column), or whether there exists a ranking at all, the position of the search results in a SERP was assumed to be less prominent and thus should have a decreased influence according to Fogg's (2003) prominence-interpretation theory.

5. STUDY 2

5.1. Method

Participants. Participants were 40 university students (7 male, 33 female; $M = 23.33$ years, $SD = 2.86$) from social and natural sciences and humanities at the same university as participants in Study 1; participation was rewarded with either course credit or payment. As in Study 1 pharmacy and medical students were excluded from participation. Furthermore, because we used the same task scenario as in Study 1, participants were not allowed to participate in both studies. Therefore, the data for Study 2 were gathered at the same time as the data of Study 1, with participants being recruited with the same announcement and from the same population and then being randomly assigned to one of the two studies. Participants had normal or corrected-to-normal vision. All participants reported to use Google as their primary search engine.

Task and materials. Task and materials were the same as in Study 1. However, instead of a standard list interface participants received the SERPs in the form of a grid interface with the same search results as in Study 1 being presented in a three-by-three grid (see Figure 4). Search results were arranged line by line with regard to their trustworthiness, that is, from left to right in each of the three rows, following the regular western reading direction. Accordingly, in the optimal trustworthiness order the search result rated as most trustworthy was presented in the upper left corner of the grid, and the search result rated as least trustworthy in the bottom right corner of the grid. In the reversed trustworthiness order, the order was reversed, with the search result presented in the upper left corner being the least trustworthy one. Participants were unknown about the ordering of the search results.

Experimental design. As in Study 1 the *trustworthiness order* (optimal or reversed) of the SERPs was varied as between-subjects factor and *search result trustworthiness* (most trustworthy, medium trustworthy, and least trustworthy search results) was considered as a within-subjects factor. Participants were randomly assigned to the two experimental conditions, with 20 participants serving in each condition.

Control variables. Demographics (gender, age), web search experience and skills (six items; Cronbach's $\alpha = .79$), and prior knowledge on Bechterew's disease, other rheumatic diseases, and respective therapies (10 items; Cronbach's $\alpha = .62$) were assessed as control variables (see Table 3 for means and standard deviations). Analyses of the respective data revealed no differences between the two experimental conditions, that is, for gender, $\chi^2(1, N = 40) = 1.56$, $p = .212$; for age, $t(38) = -0.83$, $p = .414$; for web search experience and skills, $t(38) = 0.97$, $p = .340$; and for prior domain knowledge, $t(38) = -0.06$, $p = .952$.

Procedure and dependent variables. Testing procedure and dependent variables were equivalent to those used in Study 1. Again, a 50 Hz Tobii 1750 remote eye-tracking system supported by the software ClearView 2.7.1 was used to determine the *number of search results fixated before accessing the first web page* as well as to capture the *number of search results selected* from a SERP and the *time spent on the web pages* as a function of search result trustworthiness. As the search outcome, again the *number of arguments listed from the three most trustworthy pages* were coded. The same two raters as in Study 1 scored 20% of the argument lists. Interrater reliability computed on this subsample of protocols yielded

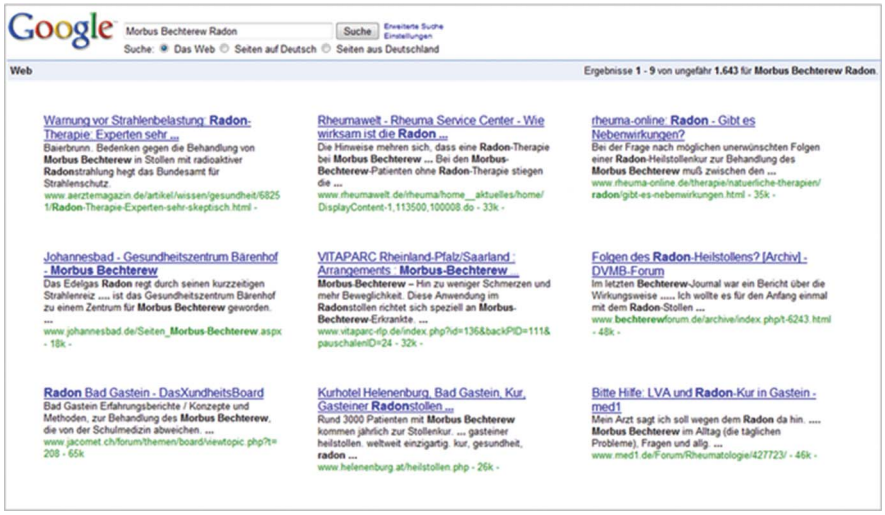

FIG. 4. Example screenshot of a grid interface.

TABLE 3
Means (and Standard Deviations) for Control Variables as a Function of Trustworthiness Order in a Grid Interface

Variable	Optimal TO	Reversed TO
Age	22.95 (2.93)	23.70 (2.81)
Gender	2 m, 18 f	5 m, 15 f
Computer and web search experience, 1 (*low*) to 5 (*high*)	3.43 (0.62)	3.24 (0.64)
Prior domain knowledge, 1 (*low*) to 5 (*high*)	1.35 (0.59)	1.36 (0.44)

Note. TO = trustworthiness order; m = male; f = female.

a Cohen's kappa of $\kappa = .78$. Disagreements were resolved through discussion between the raters. One rater coded the remaining argument lists.

According to Hypothesis 1 to 4 with respect to all four dependent variables, in the grid interface no differences were expected between trustworthiness orders.

In addition, to test in which order participants inspected the search results in the grid interface, participants' individual viewing sequences were compared to a line-by-line sequence from left to right (Search Result 1, Search Result 2, Search Result 3, Search Result 4, etc.) and a top-to-bottom column-by-column sequence from top to bottom (Search Result 1, Search Result 4, Search Result 7, Search Result 2, etc.) by means of the Levenshtein distance measure (cf. Josephson & Holmes, 2002; Salmerón et al., 2011). Levenshtein distance values were transformed into similarity percentages (i.e., 0–100% similarity to the line-by-line or column-by-column sequence). The similarity percentages were averaged across the two SERPs.

5.2. Results

An alpha level of .05 (two-sided) was used for the statistical tests reported. ANOVAs and Bonferroni post hoc tests equivalent to those in Study 1 were conducted to test our hypotheses. For means and standard deviations of the dependent variables, see Table 4.

With regard to the *number of search results participants fixated before they accessed the first web page,* in line with Hypothesis 1, no significant effects of trustworthiness order, $F(1, 38) = 1.71$, $p = .199$, were found, with 4.86 search results ($SD = 2.70$), on average, being fixated before the first click. In SERPs with optimal trustworthiness order, 40% of the participants inspected eight or all nine results before the first selection, and in SERPs with reversed trustworthiness order, it was 30% of the participants, $\chi^2(1, N = 40) = 0.44$, $p = .507$.

The analysis of participants' *viewing sequences*, that is, the order in which they inspected the search results on a SERP, on average revealed a quite low similarity to both a line-by-line viewing sequence and a column-by-column viewing sequence, without any significant differences between trustworthiness orders (both Fs < 1.14). A hierarchical cluster analysis using the Ward method with the two grouping variables "similarity to line-by-line sequence" and "similarity to column-by-column sequence" was conducted. The cluster analysis identified three subgroups of grid interface users: The viewing sequences of the first group that composed the majority ($n = 26$) of the participants (12 in the optimal and 14 in the reversed trustworthiness order) had a rather low similarity to the line-by-line sequence ($M = 47.21\%$) and an even lower similarity to the column-by-column sequence ($M = 28.55\%$). The viewing sequences of the second group, comprising nine participants (five in the optimal and four in the reversed trustworthiness order), had a very low similarity to the line-by-line sequence ($M = 33.26\%$) and a somewhat higher similarity to the column-by-column sequence ($M = 55.56\%$). Finally, the viewing sequences of the third group, comprising only five participants (three in the optimal and two in the reversed trustworthiness order) had a high similarity to the line-by-line sequence ($M = 82.38\%$) and a low similarity to the column-by-column sequence ($M = 28.19\%$).

TABLE 4
Means (and Standard Deviations) of the Dependent Variables as a Function of Trustworthiness Order in a Grid Interface

Dependent Variables	Optimal TO	Reversed TO
Number of search results fixated before 1st click	4.48 (2.65)	5.45 (2.03)
Number of most trustworthy search results selected per SERP	2.35 (0.61)	2.28 (0.40)
Number of medium trustworthy search results selected per SERP	1.25 (0.72)	1.18 (0.82)
Number of least trustworthy search results selected per SERP	1.03 (0.75)	1.28 (0.83)
Time (in s) spent on most trustworthy web pages per SERP	123.77 (41.96)	112.88 (28.38)
Time (in s) spent on medium trustworthy web pages per SERP	38.30 (29.53)	35.78 (27.68)
Time (in s) spent on least trustworthy web pages per SERP	33.62 (24.40)	39.49 (23.53)
Similarity to line-by-line sequence (in %)	51.37 (20.32)	45.56 (13.44)
Similarity to column-by-column sequence (in %)	36.10 (16.51)	33.06 (11.88)
Number of arguments from most trustworthy sources	5.55 (2.09)	5.40 (2.11)

Note. TO = trustworthiness order; SERP = search engine results page.

During their web search participants, on average, selected 51.9% of the available search results. With regard to the *number of search results selected* as a function of their trustworthiness, in line with Hypothesis 2 a significant main effect of search result trustworthiness was found, $F(2, 76) = 45.46$, $p < .001$, partial $\eta^2 = .55$. Irrespective of trustworthiness order, the three most trustworthy search results ($M = 2.31$, $SD = 0.50$) were selected significantly more often than both the medium trustworthy ($M = 1.21$, $SD = 0.76$) and least trustworthy search results ($M = 1.15$, $SD = 0.79$; both ps $< .001$), which were selected to a similar extent ($p > .99$); also see Figure 5. Besides, there was neither a main effect of trustworthiness order nor a significant interaction between trustworthiness order and search result trustworthiness (both Fs < 1).

As expected in Hypothesis 3 with regard to the *time spent on web pages* as a function of their trustworthiness, a similar pattern was found, that is, a significant main effect of search result trustworthiness, $F(1.72, 65.46) = 67.90$, $p < .001$, partial $\eta^2 = .64$, with significantly more time spent on the most trustworthy pages ($M = 118.33$, $SD = 35.79$) than both on the medium trustworthy ($M = 37.04$, $SD = 28.28$) and least trustworthy pages ($M = 36.55$, $SD = 23.85$; both ps $< .001$), irrespective of trustworthiness order (see Figure 6). Besides, there was no effect of trustworthiness order, $F(1, 38) = 2.86$, $p = .099$, or a significant interaction between the two factors ($F < 1$).

With regard to the *number of arguments listed from the three most-trustworthy sources*, as expected in Hypothesis 4 no significant differences were found between trustworthiness orders ($F < 1$), that is, participants who had received SERPs with reversed trustworthiness order performed as well as participants who had received SERPs with optimal trustworthiness order.

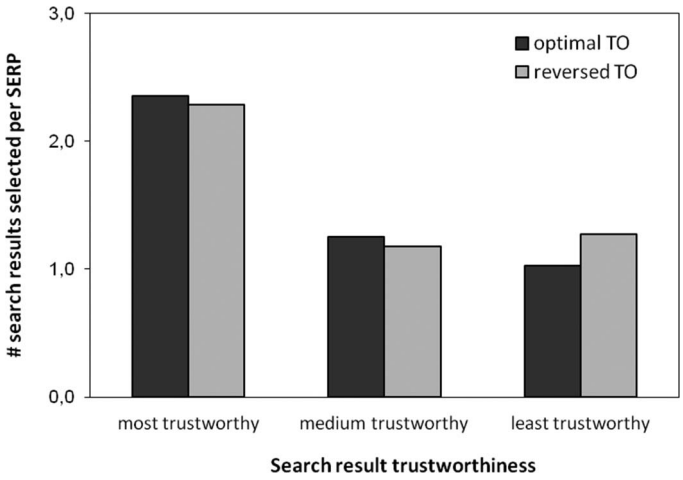

FIG. 5. Number of search results selected per search engine results page (SERP) as a function of search result trustworthiness and trustworthiness order (TO) in a grid interface.

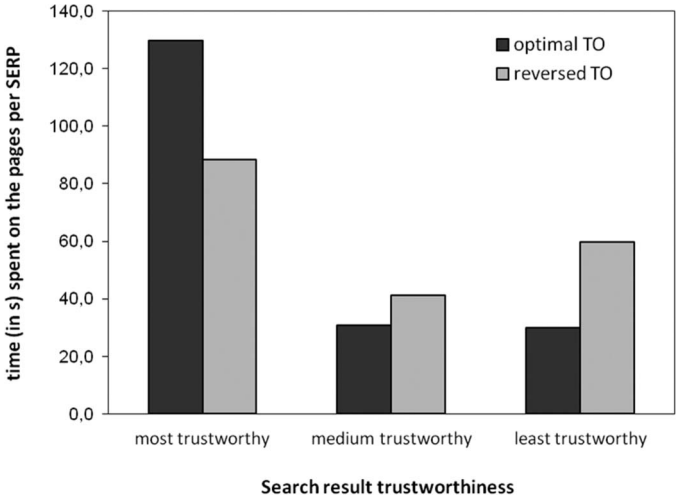

FIG. 6. Time (in seconds) spent on the web pages of a search engine results page (SERP) as a function of search result trustworthiness and trustworthiness order (TO) in a grid interface.

5.3. Discussion

The results of Study 2 confirmed that in a grid search interface, in which search results were presented in a three-by-three grid, the impact of the position of the search results on their selection was substantially reduced. Accordingly, as expected, no differences between trustworthiness orders were found: Irrespective of the trustworthiness order, grid interface users visually inspected about half of the search results before making their first selection decision (Hypothesis 1), predominantly selected the three most trustworthy search results from the SERP (Hypothesis 2), spent more time on the three most trustworthy web pages than on the rest of the pages (Hypothesis 3), and listed the same amount of arguments from the most trustworthy sources (Hypothesis 4).

In sum, the results of Hypothesis 1 indicate that—as in a list interface—in the context of limited time and knowledge and a range of search results to choose from, users applied a heuristic strategy to evaluate the search results. However, instead of relying on the position cue, which is less prominent or less clear in a grid interface, they seemed to rely more on source cues such as the name or type of site, as indicated by the predominant selection of most trustworthy search results. Furthermore, analyses of participants' viewing sequences revealed that the majority of participants inspected the search results in the grid interface neither line by line nor column by column, but in a rather nonsystematic way. This indicates that the grid interface similar to the graphical-overview interface examined by Salmerón et al. (2011) supports a rather free exploration of the search results. To conclude, Study 2 provides promising results that the strong influence of the ranking position of the search results on their selection can be reduced substantially by presenting search results in a grid interface and, thus, stimulates laypersons who search for an unfamiliar and controversial issue, to

select and use trustworthy information sources irrespective of their position in the SERP.

6. GENERAL DISCUSSION

In the present research we examined how laypeople (i.e., nonexperts in a domain) who were given the task to seek information on the web for an unfamiliar and controversial medical issue selected search results from a SERP and how this influenced their subsequent search outcomes. Results from the two studies revealed that the majority of participants did not systematically evaluate all search results available but applied a rather heuristic selection strategy (cf. Metzger, 2007; Metzger et al., 2010; Taraborelli, 2008) and inspected only some results before they selected the first one from a SERP. Yet, when the top results were of rather low trustworthiness (i.e., the reversed trustworthiness order in Study 1) or when the ordering of the results was somewhat ambiguous (as in the grid interface used in Study 2), participants visually inspected more results before their first selection as compared to a situation where the top results were the most trustworthy ones and the ordering was unambiguous (i.e., the optimal trustworthiness order in Study 1). This indicates that in the former cases web users were able to adapt their behavior choosing a somewhat more exhaustive evaluation strategy (cf. Resnick et al., 2001).

However, as shown in Study 1 in a standard list interface the ranking position of the search results in the SERP served as a strong heuristic cue for selection decisions, which also affected users' subsequent search outcomes. Based on Fogg's (2003) prominence-interpretation theory, we have argued that this might be due to the prominence of the position of items in list interfaces. Evidence for this assumption is given by the results of Study 2: In a grid interface in which it remains unclear to the user whether the ranking within a SERP is aligned horizontally or vertically, or whether there exists a ranking at all, the position no longer had an impact on users' search result selections or their subsequent search outcomes. Instead, the most trustworthy search results—as defined by a different group of university students, who were comparable in age and educational background to the participants of the experimental study—were predominantly selected irrespective of the position in the SERP. As a consequence in both conditions (i.e., optimal and reversed trustworthiness order) an equal number of arguments from the most trustworthy sources were listed after search. To conclude, from a practical perspective, redesigning the interface of search engines by displaying search results in a grid format instead of a list format might be a promising way to foster the selection of trustworthy information sources irrespective of the position of the search result in the SERP. Moreover, from a theoretical perspective, our research shows that the assumptions of Fogg's prominence-interpretation theory, in particular with respect to the "prominence" component, can also be applied to the evaluation of search results on SERPs.

An open question, however, is *why* the position of the search results, in case that it is prominent, has such a strong influence. That is, speaking in terms of the prominence-interpretation theory, the question is how the ranking position is *interpreted* by users. Hargittai et al. (2010) found that many college students believe that the search results presented first by a search engine are the most trustworthy ones. Similarly, in a large-scale survey by Purcell et al. (2012), two thirds of American search engine users reported to trust in search engine rankings believing that they are fair and unbiased. Among 18- to 29-year-old search engine users, the percentage was 72%. Two other common beliefs are that the ranking position reflects the frequency with which a page has been visited (i.e., its popularity) or the relevance of a web page to the entered search terms (cf. Nakamura et al., 2007).

Thus, to shed light on the issue of how search engine users interpret the ranking position as well as to get further insights into why they select certain search results, future research is needed that combines behavioral analyses, as used in the two present studies, with concurrent or retrospective interviews (cf. Hargittai et al., 2010) or with rating scales where users are asked to judge each search result according to its perceived relevance, trustworthiness, or popularity (cf. Flanagin & Metzger, 2007, for website ratings). These measures would also provide additional evidence about whether the grid interface users indeed selected the most trustworthy search results because of their high source trustworthiness (indicated by the type or name of the website in the title or the URL) or whether other factors such as perceived relevance as well affected their search result selections. According to Pattanaphanchai, O'Hara, and Hall (2013), for instance, besides the type or reputation of the source, accuracy (i.e., whether the information is free of errors), currency (i.e., how up-to-date the information is), and relevance (i.e., how well the content meets the user's information needs) are important criteria to assess the trustworthiness of web information. Based on Fogg's (2003) prominence-interpretation theory, we expect that users' search results evaluations in grid interfaces could be even further improved by making such cues more prominent, for instance, by highlighting the source type or by providing the last-modified date in the search results (cf. Nakamura et al., 2007). Another promising approach might be to add social information to search results, such as the number of social bookmarks of a page, aggregated user ratings (e.g. presented as rating stars), or social annotations by people of the user's online network (cf. Nakamura et al., 2007; Taraborelli, 2008). In list interfaces, because of the overruling effect of the search result position, on the contrary, we wouldn't expect any substantial influences of such additional cues. First empirical evidence for this assumption is provided by Muralidharan, Gyongyi, and Chi (2012), who showed that social annotations presented in search results of Google SERPs remained mostly unnoticed by users. Similarly, Terbeck (2011) found that user ratings presented as rating stars in shopping search results only had an effect only on users' selection decisions when the search

results were presented in the top positions of the SERP list. In general, users still paid most attention to the top results. To conclude, it seems difficult to break users habits in list interfaces.

Related to that it should be stressed that participants in the present article were unknown about the ordering of the search results in the grid interface, which was also reflected in the nonsystematic viewing sequences of the majority of the participants. It is conceivable that in case users know that search results are ordered line by line according to their trustworthiness (or according to the search engine's ranking algorithm, respectively), the positive effects of the grid interface would be reduced. Therefore, the ordering of the search results within the grid should not be communicated to the users, as it is (or was) the case in the search engines Horizobu and Viewzi. However, it is also an open question how an increased amount of rows or columns within one SERP would affect the study results. It is conceivable that in a grid search interface in which a large number of search results are presented within one page, some order effects would be present, such that users might focus on the search results in the upper part of the grid. This might be one reason why the search engine Horizobu presents only six results within one page.

Further research is needed to extend the findings of the present two studies to a broader range of users as well as to other contexts, such as more natural search situations with real information needs, without time constraints, and with the open web at users' disposal. When interpreting our results it should be considered that for the sake of experimental control, we introduced some constraints in our studies. That is, participants conducted their web search in a lab setting for an artificially designed search task with a finite set of only 18 search results and a predefined search time of 8 min (for details, see the Study 1 Procedure section). Furthermore, participants were all university students, constituting a rather homogeneous sample with web search experience and skills presumably higher than average (cf. Van Deursen & Van Dijk, 2010; Van Deursen, Van Dijk, & Peters, 2012). At the same time, however, participants had low prior knowledge about the topic of Bechterew's disease. According to dual-process theories from social psychology such as the elaboration-likelihood model (e.g., Petty & Cacioppo, 1986) or the Heuristic-Systematic Model (e.g., Chen & Chaiken, 1999) it can be assumed that individuals who seek information on a topic of high personal relevance and on which they possess high prior knowledge will engage in a more thorough and systematic evaluation of the search results (cf. Metzger, 2007; Wirth et al., 2007). As a consequence, for these individuals neither in the grid interface nor in the list interface any differences might be found between the two trustworthiness orders. This should be examined in future studies.

Finally, it should be noted that another way to support web users in the selection of trustworthy information is to further improve a search engine's ranking algorithm. In this vein, in February 2011 Google announced a significant improvement in their ranking algorithm that should

> reduce rankings for low-quality sites—sites which are low-value add for users, copy content from other web sites or sites that are just not very useful. At the same time, it will provide better rankings for high-quality sites—sites with original content and information such as research, in-depth reports, thoughtful analysis and so on. (Singhal & Cutts, 2011, para. 2)

Still, we believe that the added value of a grid interface is that it stimulates search engine users to evaluate the search results on their own instead of relying on the ranking algorithm, which can help individuals to become more competent web users in the long run.

REFERENCES

Aula, A., & Russell, D.M. (2008, June). *Complex and exploratory Web search*. Paper presented at the NSF IS3 Information Seeking Support Systems Workshop, Chapel Hill, NC. Retrieved from http://www.ils.unc.edu/ISSS/papers/papers/aula.pdf

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*, 135–149.

Braasch, J., Lawless, K., Goldman, S., Manning, F., Gomez, K. & MacLeod, S. (2009). Evaluating search results: An empirical analysis of middle school students' use of source attributes to select useful sources. *Journal of Educational Computing Research*, *41*, 63–82.

Chen, S., & Chaiken, S. (1999). The Heuristic-Systematic Model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). New York, NY: Guilford.

Cho, J., & Roy, S. (2004). Impact of search engines on page popularity. In S. Feldman, M. Uretsky, M. Najork, & C. Wills (Eds.), *Proceedings of the 13th international conference on World Wide Web* (WWW '04; pp. 20–29). New York, NY: ACM Press.

Cutrell, E., & Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in Web search. In M. B. Rosson & D. Gilmore (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07; pp. 407–416). New York, NY: ACM Press.

Estabrook, L., Witt, E., & Rainie, L. (2007). Information searches that solve problems. *PEW Internet & American Life Project*. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2007/Pew_UI_LibrariesReport.pdf.pdf

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the World Wide Web? Qualitative studies using focus groups, usability tests, and in-depth interviews. *British Medical Journal*, *324*, 573–577.

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, *9*, 319–342.

Flanagin, A. J., & Metzger, M. J. (2010). *Kids and credibility: An empirical examination of youth, digital media use, and information credibility*. Cambridge, MA: MIT Press.

Fogg, B. J. (2003). Prominence-interpretation theory: Explaining how people assess credibility online. In G. Cockton & P. Korhonen (Eds.), *Proceedings of CHI '03 extended abstracts on Human factors in computing systems* (pp. 722–723). New York, NY: ACM Press.

Fox, S. (2006). *Online health search 2006*. Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf

Fox, S. (2011). *Health topics*. Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2011/PIP_HealthTopics.pdf

Fox, S., & Jones, S. (2009). *The social life of health information*. Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, *21*, 220–231.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Granka, L., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in www search. In M. Sanderson (Ed.), *Proceedings of the 27th Annual ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '04; pp. 478–479). New York, NY: ACM Press.

Guan, Z., & Cutrell, E. (2007). *An eye tracking study of the effect of target rank on Web search*. In M. B. Rosson & D. Gilmore (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07; pp. 417–420). New York, NY: ACM Press.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: young adults' evaluation of Web content. *International Journal of Communications*, pp. 468–494.

Hilligoss, B., & Rieh, S. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing and Management*, *44*, 1467–1484.

Horrigan, J. (2006). *The Internet as a resource for news and information about science*. Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/Reports/2006/The-Internet-as-a-Resource-for-News-and-Information-about-Science.aspx

Houmb, S. H., Ray, I., & Ray, I. (2006), Estimating the relative trustworthiness of information sources in security solution evaluation. In K. Stølen, W. H. Winsborough, F. Martinelli, & F. Massacci (Eds.) *Trust management, 4th International Conference, iTrust* (LNCS Vol. 3986, pp. 135–149). Berlin, Germany: Springer.

Josephson, S., & Holmes, M. E. (2002). Visual attention to repeated internet images: testing the scanpath theory on the world wide web. In A. T. Duchowski (Ed), *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications* (ETRA '02; pp. 43–49). New York, NY: ACM Press.

Kammerer, Y., & Gerjets, P. (2012). Effects of search interface and Internet-specific epistemic beliefs on source evaluations during Web search for medical information: An eye-tracking study. *Behaviour & Information Technology*, *31*, 83–97.

Keane, M. T., O'Brien, M., & Smyth, B. (2008). Are people biased in their use of search engines? *Communications of the ACM*, *51*, 49–52.

Klöckner, K., Wirschum, N., & Jameson, A. (2004). Depth- and breadth-first processing of˝search result lists. In *CHI '04 extended abstracts on Human factors in computing systems* (p. 1539). New York, NY: ACM Press.

Lewandowski, D. (2011). The influence of commercial intent of search results on their perceived relevance. In S. Sadagopan & K. Ramamritham (Eds.), *Proceedings of the 2011 iConference* (iConference '11; pp. 452–458). New York, NY: ACM Press.

Lewandowski, D. (2013). Credibility in Web search engines. In S. Apostel & M. Folk (Eds.), *Online credibility and digital ethos: Evaluating computer-mediated communication* (pp. 131–146). Hershey, PA: IGI Global

Mansell, P., & Read, J. (2009). Posttraumatic stress disorder, drug companies, and the Internet. *Journal of Trauma & Dissociation*, *10*, 9–23.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*, 2078–2091.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, *60*, 413–439.

Morahan-Martin, J.M. (2004). How Internet users find, evaluate, and use online health information: A cross-cultural review. *CyberPsychology & Behavior*, *7*, 497–510.

Muralidharan, A., Gyongyi, Z., & Chi, E. H. (2012). Social annotations in Web search. In J. A. Konstan, E. H. Chi, & K. Höök (Eds.), *Proceedings of the 2012 ACM annual conference on Human factors in computing systems* (CHI '12; pp. 1085–1094). New York, NY: ACM Press.

Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., . . . Tanaka, K. (2007). Trustworthiness analysis of Web search results. In L. Kovács, N. Fuhr, & C. Meghini (Eds.), *Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries* (ECDL 2007; LNCS Vol. 4675, pp. 38–49). Berlin, Germany: Springer.

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*, 801–823.

Pattanaphanchai, J., O'Hara, K., & Hall, W. (2013). Trustworthiness criteria for supporting users to assess the credibility of web information. In D. Schwabe, V. Almeida, H. Glaser, R. Baeza-Yates, & S. Moon (Eds.), *Proceedings of the 22nd international conference on World Wide Web companion* (WWW '13 Companion; pp. 1123–1130). New York, NY: ACM Press.

Payne, J. W., & Bettman, J. R. (2004). Walking with the scarecrow: The information-processing approach to decision research. In D. J. Koehler & N. Harvey (Eds.) *Blackwell handbook of judgment and decision making* (pp. 110–132). Oxford, UK: Blackwell.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge University Press.

Pengnate, S., & Antonenko, P. (2013). A multi-method evaluation of online trust and its interaction with metacognitive awareness: an emotional design perspective. *International Journal of Human–Computer Interaction*, *29*, 582–593.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, NY: Springer.

Purcell, K., Brenner, J., & Rainie, L. (2012). *Search engine use 2012*. Washington, DC: Pew Internet & American Life Project. Retrieved from http://pewinternet.org/~/media//Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf

Resnick, M. L., Maldonado, C. A., Santos, J. M., & Lergier, R. (2001). Modeling on-line search behavior using alternative output structures. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Conference*, 1166–1171.

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, *53*, 145–161.

Rieh, S. Y., & Hilligoss, B. (2008). College students' credibility judgments in the information-seeking process. In M. J. Metzger, & A. J. Flanagin (Eds.), *Digital media, youth, and credibility. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning* (pp. 49–72). Cambridge, MA: MIT Press.

Salmerón, L., Gil, L., Bråten, I., & Strømsø, H. I. (2010). Comprehension effects of signalling relationships between documents in search engines. *Computers in Human Behavior*, *26*, 419–426.

Sillence, E., Briggs, P. Harris, P, & Fishwick, L. (2006). A framework for understanding trust factors in Web-based health advice. *International Journal of Human–Computer Studies*, *64*, 697–713.

Singhal, A., & Cutts, M. (2011, February). *Finding more high-quality sites in search*. Retrieved from http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html

Smith, A. (2009). The Internet's role in campaign 2008. Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/Reports/2009/6--The-Internets-Role-in-Campaign-2008.aspx

Stadtler, M., & Bromme, R. (2008). Effects of the metacognitive tool met.a.ware on the Web search of laypersons. *Computers in Human Behavior*, *24*, 716–737.

Stadtler, M., Bromme, R. & Kettler, S. (2009). Dr. Google—geschätzter Kollege? Die Rolle des Internet in der Arzt-Patient-Kommunikation [Dr. Google—A valued colleague? The role of the Internet in physician–patient communication]. *Zeitschrift für Allgemeinmedizin*, 85, 254–259.

Taraborelli, D. (2008). How the web is changing the way we trust. In A. Briggle, K. Waelbers, & P. A. E. Brey (Eds.), *Current issues in computing and philosophy* (pp. 194–204). Amsterdam, the Netherlands: IOS Press.

Terbeck, H. (2011). *Der Einfluss sozialer Empfehlungen auf das Selektionsverhalten in Suchmaschinen – eine Eyetracking-Studie der Google-Ergebnisseite* [The influence of social recommendations on the selection behavior in search engines—An eyetracking study of the Google search results page] (Unpublished bachelor's thesis). Hamburg University of Applied Sciences, Hamburg, Germany. Retrieved from http://hendrik-terbeck.de/bachelorarbeit/soziale_empfehlungen_eyetracking_google_terbeck_2011.pdf

Tu, Y. W., Shih, M., & Tsai, C. C. (2008). Eighth graders' web searching strategies and outcomes: The role of task types, web experiences and epistemological beliefs. *Computers and Education*, *51*, 1142–1153.

Van Deursen, A. J. A. M., & Van Dijk, J. A. G. M. (2010). Measuring Internet skills. *International Journal of Human–Computer Interaction*, *26*, 891–916.

Van Deursen, A. J. A. M., Van Dijk, J. A. G. M., & Peters, O. (2012). Proposing a survey instrument for measuring operational, formal, information and strategic Internet skills. *International Journal of Human–Computer Interaction*, *28*, 827–837.

Wiley, J., Goldman, S., Graesser, A., Sanchez. C., Ash, I., & Hemmerich, J. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal*, *46*, 1060–1106.

Wirth, W., Böcking, T., Karnowski, V., & von Pape, T. (2007). Heuristic and systematic use of search engines. *Journal of Computer-Mediated Communication*, *12*, 778–800.

Zhang, X., Cole, M., & Belkin, N. (2011). Predicting users' domain knowledge from search behaviors. In W.-Y. Ma, J.-Y. Nie, R. Baeza-Yates, T.-S. Chua, & W.B. Croft (Eds.), *Proceedings of the 34th international ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '11; pp. 1225–1226). New York, NY: ACM Press.

ABOUT THE AUTHORS

Yvonne Kammerer is a post-doc at the Knowledge Media Research Center in Tübingen, Germany. Her research interests include information processing and source evaluation during Web search, hypertext-based learning, and multiple-documents comprehension. In her work she combines performance measures (e.g., search or learning outcomes) and process measures (e.g., eye tracking, logfiles).

Peter Gerjets is head of the Hypermedia Lab at the Knowledge Media Research Center in Tuebingen, Germany. His research focuses on individual learning in hypermedia environments, information processing during web search, learning with dynamic visualizations, designing multitouch-interfaces, and supporting learners' competences to use hypermedia environments and the web.