

Searching the Web for conflicting topics: page and user factors

Ladislao Salmerón ¹

Yvonne Kammerer ²

Pilar García-Carrión ³

¹ ERI Lectura & Department of Developmental and Educational Psychology, University of
Valencia, Spain

² Knowledge Media Research Center, Tübingen, Germany

³ Department of Developmental and Educational Psychology, University of Seville, Spain

This is a draft. Check the original reference at *Computers in Human Behavior* (April 2013)

Correspondence may be sent to:

Ladislao Salmerón

ERI Lectura & Department of Developmental and Educational Psychology

University of Valencia

Avd. Blasco Ibáñez, 21. 46010 - Valencia (Spain)

Phone # (+34) 96 398 3847

e-mail: ladislao.salmeron@valencia.edu

Author notes

This research was funded by the Spanish Minister of Science and Innovation (Secretaría General de Universidades, EDU2011-25885). We thank three anonymous reviewers for their helpful input in a previous version of the paper.

Web users tend to search only the pages displayed at the top of the search engine results page (the 'top link' heuristic). Although it might be reasonable to use this heuristic to navigate simple and unambiguous facts, it might be risky when searching for conflicting socio-scientific topics, such as potential measures to reduce greenhouse gas emissions. In the present study, we explored the extent to which students consider other Web page characteristics, such as topic relevance and trustworthiness, when searching and bookmarking pages concerning a conflicting topic. We also examined the extent to which prior background knowledge moderates students' behavior. The results revealed that while the study participants actually used a 'top link' heuristic to navigate the results, they engaged in more systematic processes to bookmark pages for further study. Furthermore, the students' background knowledge was related to the assessment of Web page trustworthiness. We discuss these results from the perspective of a dual-processing model.

1. Introduction

Current web literacy models stress the importance of students' skills to search, evaluate, and use information from the World Wide Web (Brand-Gruwel, Wopereis, & Walraven, 2009; Eisenberg, & Johnson, 2002; OECD, 2011; Rouet & Britt, 2011). Those skills are particularly necessary when searching for complex information concerning conflicting scientific, or socio-scientific, topics that do not convey a single clear-cut solution, such as potential measures to reduce greenhouse gas emissions, possible reasons for the extinction of the dinosaurs, or the effectiveness of methods to lower cholesterol. Relevant Web pages might present only one side of a contentious issue, and may include false or unsubstantiated claims. Furthermore, as students are confronted with an increasing overload of information on the Web, they may turn to the assistance of search engines to simplify their search.

In theory, students have two alternatives for processing information when faced with the search engine results page (SERP). First, they might decide to visit only a few pages listed at the top of this page (Granka, Joachims, & Gay, 2004; Wirth, Böcking, Karnowski, & Pape, 2007). This behavior represents a trade-off between the quality of the information obtained and the cognitive cost associated with the search. The quality of the pages offered on top of the first SERP is due to the fact that search engines rely, to a substantial extent, on collective intelligence to rank Web pages (Lewandowski, 2013). This means that when a web page is linked or visited by many other users, its position in the SERP will improve. However, information contained on Web pages listed among the top search results can also be one-sided or commercially biased, as shown by Lewandowski (2011) and Mansell and Read (2009). This can be due to search engine optimization businesses (cf. Lewandowski, 2013), and the so-called rich-get-richer effect (i.e., highly-ranked pages are predominantly selected by users and in turn reach higher and higher positions in the search engine ranking; for example, see Cho & Roy, 2004).

Alternatively, students might engage in a systematic evaluation of the Web pages by considering not just the position of a page in the SERP, but also its topic relevance and trustworthiness. This option might provide an optimal search outcome and prevent students from accessing erroneous information from biased Web pages. However, a thorough search for information may come at a cost of mental effort (Rieh, Kim, & Markey, 2012), and might require particular cognitive dispositions (e.g., background knowledge about the topic) that are not possessed by all students.

The two alternative ways of processing Web pages from SERPs that have been sketched above may not only influence the way students access information, but also the way they use this information. Indeed, while students generally do not use the Web to just retrieve and evaluate Web pages, they ultimately must make use of the information provided in those pages, for example, to write an essay on the topic (Brand-Gruwel et al., 2009; Eisenberg, & Johnson, 2002; OECD, 2011; Rouet & Britt, 2011). Although some students may just uncritically use the information provided at the top of the SERP (cf. Wiley et al., 2009), others might systematically evaluate the topic relevance and trustworthiness of a Web page before using particular information. In addition, background knowledge might be required in order to systematically evaluate Web pages before using them.

In the present paper, we investigated whether university students searching the Web for a conflicting socio-scientific topic selected the first results provided by a search engine, or, rather, engaged in a systematic search of Web pages, both during the search for, and use of, information. Since tuning into one processing mode or another might be dependent on students' dispositions, we studied the potentially moderating effect of background knowledge on both search for, and use of, information.

2. Theoretical background

2.1 Dual-processing models of web literacy

Two major aspects of web literacy have been studied in previous research: how people search Web pages (e.g., Pan et al., 2007), and how people assess the credibility of Web pages (e.g., Metzger, 2007). However, to date, both lines of research have not converged to create a unified account. This is clearly exemplified by the use of different methodologies used to study web literacy. In most Web searching research, participants have been provided with a set of fact-finding tasks, for which they must find the correct answer in a set of Web pages that are presented via a search engine interface, and that do not tend to vary in their credibility level (e.g., Cutrell & Guan, 2007; Granka, Joachims, & Gay, 2004; Lorigo et al., 2006; Pan et al., 2007; see Gerjets, Kammerer, & Werner, 2011; Kammerer & Gerjets, 2012, and Wiley et al., 2009 for recent exceptions). In this area, researchers primarily focus on objective behaviors, such as which Web pages are visited and for how long, as well as participants' task performance (i.e., the percentage of correctly solved tasks and the time taken to solve them). Conversely, in studies of Web credibility, participants are usually presented with a preselected set of Web pages, which vary with regard to their credibility, and are asked to judge each page, according to its believability, accuracy, trustworthiness, or bias, etc., using rating scales (e.g., Flanagin & Metzger, 2007, 2010; Fogg et al., 2003). Alternatively, they use interview studies may be used to examine how participants judge the credibility of certain Web page types (e.g., blogs, wikis, commercial pages, institutional pages, etc.) in general (e.g., Metzger, Flanagin, & Medders, 2010). Therefore, in this area, the main dependent variables are individuals' subjective reports, which allow the determination of how individuals consider certain credibility cues on Web pages in order to make a global credibility assessment.

One way of connecting these two lines of research is to consider a common theoretical framework that can account for the major aspects studied in both. We suggest that potential candidates for such a connection are dual-processing models for information processing. Relevant dual-process models have been proposed in several research fields, such as the psychology of reasoning (e.g. Evans, 2008), judgment and decision-making (e.g. Kahneman & Frederick, 2005), and social cognition (e.g. Petty & Cacioppo, 1981, and for a review see

Evans, 2008). In this vein, we follow recent studies that have adapted this framework to Web searching (Wirth, Böcking, Karnowski, & von Pape, 2007) and Web credibility (Metzger, 2007). The dual-processing framework distinguishes two processing modes that are available while engaging in information-processing tasks. People may use systematic processes in order to perform a complete evaluation of collected information, based on various different characteristics (e.g., the topic relevance, trustworthiness, or completeness of Web pages or the information therein, respectively). These processes are slow and deliberate, and consequently demand the use of a large amount of cognitive resources. Conversely, people may instead rely on heuristic processes that only focus on a limited set of characteristics of the existing information, which, nevertheless, tend to be sufficiently informative (e.g. the high position of a Web page in the SERP). These processes are fast and automatic, and therefore demand fewer cognitive resources to operate.

Dual-processing models assume that individuals' use of systematic or heuristic processes is moderated by task characteristics and cognitive requirements (Evans, 2008). Effortful and systematic processing will occur whenever the task motivates students (e.g., the Web search task will be graded for a course) and/or when students possess the ability to complete the task (e.g., having sufficient prior background knowledge to accurately process the Web page information). Otherwise, faster and heuristic processing might take place.

2.2 Heuristic processing in web search and credibility assessment

Web searching (and the subsequent use of information) around conflicting topics can be described as a situation that requires choosing between a large number of alternatives (i.e., Web pages) that vary in many dimensions (i.e., position on the SERP, topic relevance, page trustworthiness, completeness, etc.; cf. Braasch et al., 2009; Savolainen & Kari, 2006). These dimensions can be assessed through superficial or deep relevance cues (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011). For example, students might consider that a Web page is

topically relevant after realizing that its title shares some words with the search query (i.e., a superficial cue), or after a thorough reading of the Web page content (i.e., a deep cue).

As discussed previously, a systematic analysis of all information found while searching on the Web is costly in terms of time and cognitive resources. Therefore, in many situations, students may instead use fast heuristics, considering only a few of the important dimensions of the situation (e.g., the top links on a SERP). Gigerenzer and Gaissmaier (2011) suggested that heuristics may be described with regard to three main components, or building blocks: 1. Search rules indicate the order in which the information search will take place; 2. Stopping rules specify when the user stops the search; and 3. Decision rules specify how the final decision concerning choosing an alternative is reached.

In the context of a Web search, a common behavior is the ‘top link’ heuristic (or ‘Google trust’ heuristic, as referred to by Pan et al., 2007), that is, choosing one of the first few search results presented by the search engine, without evaluating all search results on the SERP (or on subsequent SERPs). Although several empirical studies have shown its use (Cutrell & Guan, 2007; Granka et al., 2004; Lorigo et al., 2006; Pan et al., 2007), to the best of our knowledge no systematic account of the heuristic in terms of the building blocks described by Gigerenzer and Gaissmaier (2011) has been proposed to date. Therefore, we tentatively conceive the ‘top link’ heuristic as the following three building blocks: 1. Search rule: search through Web pages in the order listed by the SERP, from top to bottom; 2. Stopping rule: stop when finding the first Web page from the top of the SERP that seems appropriate (e.g. because it comprises keywords that match the search goal); 3. Decision rule: infer that the Web page identified in step 2 is sufficiently good to obtain the information from.

In a seminal study evaluating the use of the ‘top link’ heuristic, Pan et al. (2007) provided undergraduate students with two different SERP scenarios (normal SERP and reversed SERP), to allow them to search for undisputed information (e.g. ‘Find the homepage of Michael Jordan, the statistician’, or ‘Where is the tallest mountain in New York located?’). In the normal SERP

scenario, the Web search results at the top of the list were more topically relevant than the others, whereas in the reversed SERP scenario the order of the results presentation was reversed, with the pages at the top being those that were the least topically relevant. Eye tracking and log file data revealed that while inspecting the SERP, participants looked at the top three results thoroughly and primarily ignored the results at the bottom. This pattern arose even in the reversed SERP scenario, in which the top pages were less topically relevant to the information goals of the participants. Similarly, students primarily clicked on the top pages, on both the regular and reversed SERPs. As a consequence, they showed a lower task performance in terms of finding a page with the correct answer in the reversed SERP scenario than in the normal SERP scenario. Therefore, the results clearly showed that participants applied a 'top link' heuristic to both navigate the Web pages and to answer the search questions, without making a systematic assessment of the topic relevance of the pages. In merely analyzing superficial cues of topic relevance, such as the presence of keywords in the Web page titles, the students appeared to have decided that those pages were sufficient for their task (cf. Rouet et al., 2011).

We assume that when seeking information about conflicting topics, students would habitually also apply a 'top link' heuristic to navigate Web pages. For example, this was indicated in a study by Gerjets, Kammerer, and Werner (2011), whereby university students that were given the task of searching for information about the effectiveness of two competing weight loss methods selected more search results toward the top than the bottom of the SERPs. However, would students still apply a 'top link' heuristic when deciding which of the pages to use for their coursework? We suggest that, in this case, students would be motivated to use topically relevant pages, so instead of following a 'top link' heuristic to choose the pages for further study, they would instead assess the topic relevance of the pages to make these decisions. The first aim of the present study was to test this assumption.

Previous research indicates that students' background knowledge about the topic plays an important role in their credibility assessment of Web pages, as will be outlined in the following section.

2.3 Background knowledge and trustworthiness evaluation of Web pages

An assessment of the credibility, or trustworthiness, of a Web page could prevent students from using biased information about conflicting topics. While still in its infancy, existing research suggests that background knowledge about a topic positively influences student evaluation of the trustworthiness of Web pages, as shown by self-reports (i.e., ratings or verbal protocols). For example, Eastin (2001) found that during a Web search for medical information, participants were more accurate in their credibility assessments of well-known content than of lesser-known information. Moreover, Fogg et al. (2003) examined which criteria and features domain experts and novices reportedly used to evaluate the credibility of health and finance Web sites. The results showed that, whereas domain novices primarily based their credibility evaluations on the Web page design (e.g., colors, layout, pictures), domain experts most often relied on author or publisher information, followed by credibility evaluations relating to references provided on the sites, or based on perceived motives or biases. In addition, Bråten, Strømsø, and Salmerón (2011) found that when reading multiple documents about climate change, undergraduates with little knowledge of the subject matter trusted different documents to the same extent, irrespective of the type of source, whereas students with greater knowledge judged an article issued by a company with vested interests in the addressed issue as being less trustworthy than other documents.

Why do students unfamiliar with the search topic base their evaluations of the trustworthiness of Web pages on superficial cues, or do not assess trustworthiness at all, even in situations where they risk using biased information? As discussed previously, dual-processing models suggest that choice of systematic, over heuristic, processing in making multi-dimensional decisions critically depends on the costs and benefits of the strategies, relative to the resources available to the decision-maker (e.g. Payne, 1982). The assessment of page trustworthiness might compete for users' limited resources. In accordance with this, Bråten et al. (2011) proposed a limited-resources hypothesis to explain users' limited evaluations of trustworthiness in complex learning scenarios. Existing evidence suggests that students with little background knowledge invest

greater effort in comprehending the content of Web documents than do knowledgeable users: They show more difficulties in identifying and relating important ideas in hypermedia documents (for a review, see Amadiou, Tricot, & Mariné, 2011). As a consequence, they possess fewer available cognitive resources to engage in additional cognitive processes than knowledgeable students, as demonstrated by subjective reports of cognitive load while studying Web documents (Amadiou, Van Gog, Paas, Tricot, & Mariné, 2009). In contrast, students with greater background knowledge on the topic are not overloaded by the task of comprehending the document message, and so may remain capable of devoting additional cognitive resources to performing profound evaluations of the trustworthiness of Web pages. Therefore, we assume that students with greater background knowledge would also be more likely to consider the trustworthiness of Web pages in making their decisions around which one to use for their coursework than would students with little knowledge. Please note that, to the best of our knowledge, no study has yet investigated the interaction between background knowledge and evaluations of trustworthiness when it comes to students' actual use of information from Web pages, such as bookmarking a page for later study.

Furthermore, if the limited-resources hypothesis holds true, students' assessment of page trustworthiness might also depend on how cognitively demanding the search task is. In a situation in which users can effectively apply a 'top link' heuristic to obtain relevant pages (because the top pages are also highly relevant to the search goal), they might have more free resources available to critically consider Web page trustworthiness. In contrast, in a situation in which users do not obtain relevant results by applying the 'top link' heuristic (e.g., in that the top results are not the most relevant pages), they would turn to a more analytic assessment of page relevance (Wirth et al., 2007), which might overload their cognitive resources and, thus, might limit their capacity to assess Web page trustworthiness.

2.4 Research hypotheses

The aim of our study was to examine the extent to which students searching the Web for pages relating to a conflicting topic considered the link position, the topic relevance, the trustworthiness of the pages, or the information therein in their navigation behavior, as well as their later use of information. We expected that learners would apply a ‘top link’ heuristic to navigate and guide their search through the SERP, but would turn to a more systematic assessment of Web page features when deciding which information to use, in accordance with the results of previous studies, in which Web users searched for simple, undisputed facts (e.g., Pan et al., 2007). Specifically, we expected that in a SERP scenario similar to the one used by Pan et al. (2007), participants would visit more of the topically relevant pages and would spend more time on these pages when they were presented at the top, rather than at the bottom, of the SERP list (Hypothesis 1a). Similarly, participants were expected to visit more of the comparatively irrelevant pages, and to spend more time on these pages, when they were presented at the top, rather than at the bottom of the SERP list (Hypothesis 1b). However, when deciding which information to use for further study (by bookmarking it), we expected that learners would not apply the ‘top link’ heuristic. Specifically, Hypothesis 2 predicted that they would bookmark pages that were topically relevant for their search goal for later study, even if the pages were presented toward the bottom of the SERP list, and those on top of the SERP were only superficially relevant (i.e., the title included the terms of the search query).

Furthermore, the limited-resources hypothesis (Bråten et al., 2011) predicts that learners will engage in evaluations of trustworthiness if they have sufficient cognitive resources available after assessing topic relevance. This assessment may influence both evaluation and use of Web page information. To test these predictions, we analyzed students’ bookmarking, as well as the reported criteria they had used to select those particular pages. Specifically, when the task was less cognitively demanding, such as when topically relevant search results were presented at the top of the SERP, rather than at the bottom, we expected that participants would bookmark Web pages that were not only topically relevant, but which were also trustworthy (Hypothesis 3a), and to a greater extent would base their selections on the trustworthiness of the information

provided by the Web page (Hypothesis 3b). Similarly, we predicted that participants with a high level of prior background knowledge on the topic would bookmark topically relevant and trustworthy pages (Hypothesis 4a), and would use evaluations of trustworthiness as selection criteria (Hypothesis 4b), more often than those with a lower level of background knowledge.

3. Method

3.1 Participants

The participants were 67 undergraduate psychology students from a large Spanish University (mean age 22.27 years, $SD = 2.38$, 86.6% female), and they received course credit for participation. On average, the participants had been using computers for 10.26 years ($SD = 3.10$) and the Internet for 7.58 years ($SD = 2.36$).

3.2 Materials

3.2.1 Task and Web materials

The participants were given the task of seeking information on the Web regarding the "Reduction of greenhouse gas emissions", which is an important and controversial socio-scientific issue. To complete the experimental task, we provided 10 search results linked to real Web pages retrieved from Google[®], with the queries "Reduction of greenhouse gas emissions". All search results included the query terms in their titles and provided different views of each topic (see Table 1). However, not all pages were topically relevant for the task (cf. Rouet et al., 2011). For example, one page was an online magazine informing readers of an agency that provided Spanish companies with certificates detailing their reduction of greenhouse gas emissions. However, this page did not present any solution to the problem of how to reduce such emissions. We edited the Web pages to disable the hyperlinks therein, and to make them similar in length (607-725 words).

- Insert Table 1 about here -

Eight graduate students with experience in text analysis rated the Web pages according to the following three criteria: (a) the page's relevance to the topic (i.e., topic relevance), (b) the degree to which the page covered the topic (i.e., completeness), and (c) the extent to which the page was trustworthy (i.e., trustworthiness), on scales from 1 (low) to 5 (high). Seven pages were considered rather relevant to the topic ($M = 3.8$), and three were considered to be rather irrelevant ($M = 2.7$). Of the relevant pages, three were considered to be more complete ($M = 4.1$), and four were considered less complete ($M = 2.7$). Finally, from the three relevant and complete pages, one page was rated as more trustworthy ($M = 4.5$, i.e., a page of the United Nations agency on climate change describing the measures proposed by the agency to reduce greenhouse gas emissions) than the remaining two pages ($M = 3.6$, personal blog and a page from an Non-governmental organization (NGO)).

3.2.2 Background knowledge questionnaire

To assess the participants' background knowledge on the topic of climate change, we used a reliable multiple-choice test that had been used in several previous studies (e.g., Bråten, Strømsø, & Salmerón, 2011; Gil, Bråten, Vidal-Abarca, & Strømsø, 2010; Salmerón, Gil, Bråten, & Strømsø, 2010). This test consisted of 17 items with five response options for each item. Diverse aspects of climate change were assessed, with items referring to both scientific (e.g., the greenhouse effect), and political (e.g., the Kyoto Protocol) issues, which were also discussed in the 10 Web pages. The participants' background knowledge score was the sum of correct responses out of the 17 items. The test-retest reliability of the questionnaire was computed in an independent sample of undergraduates from the University of Valencia ($n = 80$), with 2 weeks between the test and the retest. This yielded a reliability estimate (Pearson's r) of .73 (Gil et al., 2010).

3.3 Experimental design

The order of the pages on the SERP (i.e., the *SERP type*) was varied as a between-subjects factor, and they were presented either in a *normal* or in a *reversed* order. In the *normal* SERP condition, the Web pages were organized according to their topic relevance and completeness, as rated by the graduate students. Accordingly, the more relevant and more complete pages were presented at the top of the list. In addition, to test students' evaluations of trustworthiness, the first two pages of the list were the less trustworthy pages, and the third page the more trustworthy page (i.e., the United Nations Web page). That is, the two pages at the top were the two relevant-complete, but less trustworthy pages, page three was the relevant-complete trustworthy page, pages 4 to 7 were the relevant, but less complete, pages, and the final two pages were those that were rather irrelevant. In the *reversed* SERP condition, the Web pages were displayed in reverse order. The participants were randomly assigned to the two experimental conditions, with 34 participants working with the normal SERP and 33 participants working with the reversed SERP.

In addition, the participants' background knowledge of climate change (see 3.2.2 for details) was used as a second continuous factor.

3.4 Procedure

The participants performed the task in groups of four or five at the University laboratories. Before beginning the search task, they filled out the background knowledge test. We then asked the participants to conduct a Web search, giving the following instructions: "Imagine you are taking an introductory course on Climate Science. The professor has requested you to work in small groups to write a report on the topic of climate change. Your group has decided to divide the work into several subtopics, and you will be in charge of the subtopic "Reduction of greenhouse gas emissions". For the next group meeting, your task is to search the Internet to look for relevant pages and to bookmark the two pages that best cover your subtopic, so that your colleagues would have a complete understanding of the topic by reading only those two pages." For their Web search, the participants were presented with a mock SERP of a well-

known search engine (Figure 1), from which they could access 10 preselected Web pages (see 3.2.1 for details) without time constraints. Navigation behaviors were recorded through navigation log files. Having explored the Web pages, the participants were asked to bookmark the two pages that best covered the topic. For this task, they were again presented with the SERP and had to tick checkboxes next to the two pages that they wanted to bookmark. Finally, for each of the two pages that they had bookmarked, they were required to indicate whether they had considered the following three criteria to bookmark the pages: 1) the link position in the list, 2) whether the page provided topically relevant information, and 3) whether the page provided trustworthy information.

- Insert Figure 1 about here -

3.5 Dependent variables

3.5.1 Navigation behavior

Page visits and reading times were analyzed as dependent variables, with respect to navigation behavior during the Web search. Specifically, to test our hypothesis regarding the use of a 'top link' heuristic to search Web pages (i.e., Hypotheses 1a and 1b), we computed the percentage of participants that first clicked on the link at the top of the list. Furthermore, we computed the percentage of (a) relevant and complete pages (relevant-complete: i.e., pages 1-3 in the normal SERP, pages 8-10 in the reversed SERP), (b) relevant, but incomplete, pages (relevant-incomplete: i.e., pages 4-7 in the normal and in the reversed SERPs), and (c) rather irrelevant pages (irrelevant: i.e., pages 8-10 in the normal SERP, pages 1-3 in the reversed SERP) that were visited by the participants and how much time they spent on the respective pages. In addition, we separately computed whether the relevant-trustworthy page (i.e., the page of the United Nations agency) was visited and for how long. According to the 'top link' heuristic (Hypotheses 1a), we expected that fewer participants would visit the reversed SERP, and that the page would be visited for a shorter time on average, compared to the normal SERP.

3.5.2 Information use (bookmarking)

With respect to the page bookmarks, to test our hypothesis regarding participants' topic relevance evaluations (i.e., Hypothesis 2), equivalent to the analyses for the navigation behavior (see section 3.5.1), we computed how many of the two pages that were bookmarked (in percentages) were (a) relevant and complete pages, (b) relevant, but incomplete, pages, and (c) rather irrelevant pages, as dependent variables. In addition, to test our hypotheses regarding the participants' evaluations of trustworthiness (Hypothesis 3a and 3b, and Hypothesis 4a and 4b), we analyzed whether or not the relevant-trustworthy page was bookmarked, as well as the percentage of participants that had considered 1) the link position, 2) the topic relevance, or 3) the trustworthiness of the information on the Web page as criteria for bookmarking the different types of pages.

4. Results

SERP type (normal and reversed) and the participants' background knowledge (z-standardized) were used as factors in all analyses. With regard to heterogeneous regression slopes, the interaction term between SERP type and background knowledge was also included in the analyses.

4.1 Navigation behavior

Table 2 shows means and standard errors (corrected for the influence of background knowledge) for the navigation behavior variables as a function of SERP type.

- Insert Table 2 about here -

To answer Hypothesis 1 that the participants would apply a 'top link' heuristic to navigate to Web pages in both SERP types (normal and reversed), irrespective of the topic relevance of the pages, we computed the percentage of participants that first clicked on the link at the top of the list. Across SERP types, the majority visited the top link first ($M = 67.16\%$). A logistic regression analysis showed no effects of SERP condition or background knowledge, $b = 0.30$,

$SE = 0.27, p = .25$ and $b = 0.11, SE = 0.26, p = .67$, respectively. In the normal SERP group, 73.53% of the participants visited the top link first, and in the reversed SERP group this number was 60.81%. To further explore Hypothesis 1a, we computed how many of (a) the relevant and complete pages, (b) the relevant, but incomplete, pages, and (c) the rather irrelevant pages were visited for each SERP condition. The results of a repeated-measures ANCOVA, with Web page type as a within-subjects factor, SERP type as a between-subjects factor, and background knowledge (z-standardized) as a continuous factor, showed a significant main effect of Web page type, $F(1.83, 116.89) = 12.75, p < .001$, partial $\eta^2 = .17$ (Greenhouse-Geisser corrected). This effect was qualified by a significant interaction with SERP type, $F(1.83, 116.89) = 4.49, p = .02$, partial $\eta^2 = .07$ (Greenhouse-Geisser corrected). In line with Hypothesis 1a, Fisher LSD *post hoc* tests showed that, as expected, significantly more of the relevant-complete pages were visited on the normal SERP, where they were at the top of the list, than on the reversed SERP, where they were at the bottom of the list ($p = .04$). However, contrary to what was expected in Hypothesis 1b, the participants in the reversed SERP group did not visit more of the irrelevant pages, which in this condition were located at the top, than in the normal SERP group, in which these pages were at the bottom ($p = .28$). Relevant-incomplete pages, which were in the middle of the list in both SERPs, were also visited to a similar degree across the two SERP types ($p = .37$). In addition, there were no significant main effects of SERP type or prior knowledge, or any other significant interactions (all $F_s < 1.77$).

Furthermore, we analyzed the reading times (square-root-transformed because of a strong positive skew that is typical for time data) spent on the different page types. Again, the repeated-measures ANCOVA showed no significant main effect of SERP type, $F(1, 64) = 1.82, p = .18$, but of Web page type, $F(1.77, 113.10) = 18.73, p < .001$, partial $\eta^2 = .23$ (Greenhouse-Geisser corrected). This effect was qualified by a significant interaction with SERP type, $F(1.77, 113.10) = 12.89, p < .001$, partial $\eta^2 = .17$ (Greenhouse-Geisser corrected). In accordance with Hypotheses 1a and 1b, Fisher LSD *post hoc* tests showed that participants in the reversed SERP group spent significantly more time on the irrelevant pages, which in this

case were located at the top of the list, than did those in the normal SERP group, for whom these pages were at the bottom ($p = .001$), and significantly less time on the relevant-complete pages, which were located at the bottom of the reversed SERP, than did the normal SERP group, for whom these pages were at the top ($p = .05$). The time spent on the relevant-incomplete pages, which were located in the middle of the list on both SERPs, did not differ between SERP types ($p = .10$).

In addition, there was a significant main effect of background knowledge, $F(1, 64) = 11.81, p = .001$, partial $\eta^2 = .16$, with a higher amount of background knowledge being associated with longer reading times for all three types of Web pages.

In addition, we tested whether the 'top link' heuristic also holds for the navigation to the relevant-trustworthy page (i.e., position 3 in the normal SERP and position 8 in the reversed SERP, respectively). For this purpose, we performed a logistic regression analysis to examine whether or not the participants visited the relevant-trustworthy page, with SERP conditions (normal and reversed), and background knowledge (z-standardized), entered as predictors. The results showed that SERP conditions and participants' background knowledge did not influence whether the participants visited the relevant-trustworthy page or not, $b = 0.67, SE = 0.44, p = .12$ and $b = 0.52, SE = 0.41, p = .20$, respectively. All but eight (i.e., $N = 59$) participants visited the relevant-trustworthy page (94.1% of the participants in the normal SERP group and 81.8% of the participants in the reversed SERP group). Similarly, an ANCOVA with SERP condition as a between-subjects factor and background knowledge (z-standardized) as a continuous factor also showed no effect of SERP type, $F < 1$ for the time spent on this page. That is, irrespective of the link position, participants visited the relevant-trustworthy page to a similar extent, which contradicts the 'top link' hypothesis. In addition, as for the other page types (see above), there was a significant effect of background knowledge, $F(1, 64) = 4.29, p = .04$, partial $\eta^2 = .06$, with a greater amount of background knowledge being associated with longer reading times for the relevant-trustworthy page.

4.2 Information use (page bookmarking)

Hypothesis 2 stated that when searching for conflicting topics, students would bookmark relevant Web pages, irrespective of their position in the SERP, that is, even if these pages were located further down the list. Therefore, for each condition (normal and reversed SERP) we computed how many of the bookmarked pages were relevant and complete pages (relevant-complete), relevant, but incomplete, pages (relevant-incomplete), or rather irrelevant pages (irrelevant). In accordance with the hypothesis, the results of a repeated-measures ANCOVA with SERP type as a between-subjects factor, Web page type as a within-subjects factor, and background knowledge (z-standardized) as a continuous factor, showed a significant main effect of Web page type, $F(1.14, 71.29) = 71.47, p < .001$, partial $\eta^2 = .53$ (Greenhouse-Geisser corrected). Approximately two thirds of the bookmarked pages were relevant and complete pages (67.91%), which numbered significantly more ($p < .001$) than the relevant, but incomplete, pages, which accounted for approximately one third (30.60%) of the bookmarked pages. These, in turn, were bookmarked significantly more often ($p < .001$) than the irrelevant pages, which were rarely bookmarked (1.49%, i.e., one participant in each group), despite the fact that they were at the top of the list on the reversed SERP. In addition, there were no other significant effects (all $F_s < 1$ or $F(1.14, 71.29) = 2.23, p = .14$ for the interaction between Web page type and background knowledge, respectively). See Table 3 for descriptive statistics (corrected for the influence of background knowledge) as a function of SERP type.

- Insert Table 3 about here -

Furthermore, predictions based on the limited-resources hypothesis (Bråten et al., 2011) claimed that participants would bookmark highly trustworthy pages and use page trustworthiness as a criterion more often if they were in the normal SERP group than if they were in the reversed SERP group (Hypotheses 3a and 3b), or if they possessed higher background knowledge (Hypotheses 4a and 4b).

With regard to the bookmarking of the relevant-trustworthy page, of the 59 participants who had visited the page, 59.3% (N=35) also bookmarked it (56.3% in the normal SERP and 63.0% in the reversed SERP group, respectively). A logistic regression analysis showed no effect of SERP condition, $b = -0.12$, $SE = 0.29$, $p = .68$, which did not support hypothesis 3a.

Background knowledge was also not a significant predictor, $b = 0.36$, $SE = 0.30$, $p = .24$, but there was a significant interaction between the two factors, $b = 0.63$, $SE = 0.30$, $p = .04$. In the normal SERP group, the relationship between the participants' background knowledge and the bookmarking of the relevant-trustworthy page was significantly positive ($r = .41$, $p = .02$). That is, the higher the participants' background knowledge, the more likely they were to bookmark the relevant-trustworthy pages. In contrast, in the reversed SERP group, background knowledge was not related to the bookmarking of this page ($r = -.13$, $p = .51$).

Finally, we compared the criteria the participants used to bookmark the pages, that is, 1) the link position, 2) the topic relevance, or 3) the trustworthiness of the information in the Web page. For the three criteria used as dependent variables, we performed three logistic regression analyses with SERP type (normal and reversed), the type of Web page that was bookmarked (i.e., a relevant-untrustworthy page, the relevant-trustworthy page, or a relevant-incomplete or irrelevant page¹), and background knowledge (z-standardized), entered as predictors (as well as the interaction between the factors in the case of heterogeneous regression slopes). Table 4 shows the percentage of participants that based their bookmarking of a certain Web page type on each respective criterion.

With respect to page relevance, all but three participants reported that they considered this criterion in their bookmarking. Therefore, we refrained from computing any analyses for this variable. With respect to link position, the regression analysis showed significant main effects of both SERP type and background knowledge, $b = 0.77$, $SE = 0.22$, $p < .001$ and $b = -0.78$, $SE = 0.24$, $p = .001$, respectively. In the normal SERP group, significantly more participants

¹ We did not differentiate between relevant-incomplete and irrelevant pages, because in both the normal SERP group and the reversed SERP group only one participant had bookmarked an irrelevant page.

reported that they had based their bookmarking on the link position of the page than in the reversed SERP group. Furthermore, the higher the participants' background knowledge, the less likely they were to base their bookmarking on this criterion. Web page type had no significant effect on the consideration of the link position, $b = 0.31$, $SE = 0.28$, $p = .27$. Prior knowledge significantly influenced whether or not participants considered page trustworthiness as a criterion for bookmarking, $b = -1.33$, $SE = 0.47$, $p = .01$, as expected by hypothesis 4b. This effect was qualified by a significant interaction with Web page type, $b = 1.01$, $SE = 0.37$, $p = .01$. Although the relationship between background knowledge and the consideration of page trustworthiness as a criterion was significantly positive for the relevant-trustworthy page ($r = .35$, $p = .04$), this relationship was negative for the relevant-incomplete or irrelevant pages ($r = -.39$, $p = .01$). That is, the higher the participants' background knowledge, the more likely they were to consider page trustworthiness as a criterion for the bookmarking of the relevant-trustworthy page, and the less likely they were to consider it as a criterion for the relevant-incomplete or irrelevant pages. For the relevant-untrustworthy pages, background knowledge was not significantly related to the consideration of page trustworthiness ($r = -.20$, $p = .15$). Contrary to what was expected, SERP type had no significant effect on the reported use of page trustworthiness as a criterion, $b = 0.16$, $SE = 0.26$, $p = .55$.

5. Discussion

The results of our study reveal that the participants employed a 'top link' heuristic while searching Web pages for a conflicting socio-scientific topic, but that they engaged in more systematic processing to assess the topic relevance and trustworthiness of Web pages when deciding which information to bookmark for later study. Furthermore, their background knowledge only played a prominent role in assessing the trustworthiness of Web pages' when the SERP displayed more topically relevant pages at the top of the list, and less relevant pages at the bottom (i.e., the normal SERP).

In the following sections, we initially discuss how our results emphasize the need to distinguish between Web navigation and information use. Next, we discuss the page and user factors that play a role in the bookmarking of Web pages for further study. Finally, we comment on the potential instructional applications of our results.

5.1 Navigation versus use of Web pages of conflicting topics

Previous studies have emphasized that users follow the order of Web pages provided in SERPs while searching the Web ('top link' heuristic). Pan et al. (2007) argued that information seekers trust search engines highly, implying that users would not only navigate the Web pages in the order suggested by the SERP, but would also use the information on the top pages, accordingly.

Our results concur with previous research, in that the participants used a 'top link' heuristic to guide their navigation through a SERP (Cutrell & Guan, 2007; Granka et al., 2004; Lorigo et al., 2006; Pan et al., 2007). Most participants begin by visiting the Web page displayed at the top of the SERP, both when the SERP included more topically relevant pages at the top of the list (normal SERP), or when it displayed less topically relevant pages at the top (reversed SERP). Similarly, the data show that the students visited topically relevant and complete Web pages more often, and for a longer amount of time, when they were located at the top of the SERP than when they were located at the bottom, whereas the reverse was found for the time spent on rather irrelevant pages. Finally, the participants visited the relevant and trustworthy page from the United Nations agency, which was located nearer to the middle of the SERP (3rd position in the normal SERP, 8th position in the reversed SERP), to a similar degree and for a similar amount of time in both SERP groups. This result does not concur with the 'top link' heuristic view, because, in the reversed SERP, that particular link was located towards the bottom of the list. This suggests that if the SERP unexpectedly provides fewer relevant pages at the top of the list during navigation, users may as well activate more systematic evaluative processes to navigate.

Furthermore, our results show that the participants' navigation behavior did not necessarily correspond to their actual information use. Indeed, in the context of searching for socio-scientific conflicting topics, such as the effectiveness of potential solutions to reduce greenhouse gas emissions, applying a 'top link' heuristic when bookmarking pages would reflect an uncritical acceptance of the recommendations of the search engine. Contrary to this view, our results show that the participants, although apparently navigating by using a 'top link' heuristic, turned to a more systematic processing of other Web page features to decide which pages to bookmark for further study. Specifically, the data show that across the SERP types (normal and reversed) most of the participants reported that they bookmarked Web pages for further study on the basis of their topic relevance, and that they also bookmarked primarily relevant and complete Web pages, while rarely choosing rather irrelevant pages. The fact that users do not bookmark Web pages at the top of the SERP that are superficially relevant (because their link titles include the search terms), but are actually irrelevant, suggests that they engage in rather systematic processing to bookmark pages (cf. Rouet et al., 2011).

Although bookmarking pages for later study is a rather common activity in learning scenarios, we should be cautious of overgeneralizing from this task to other information use tasks on the Web. By bookmarking a Web page, students are making an explicit decision regarding the utility of that page for a later study activity, such as writing an essay on the topic it covers. Interestingly, this measure represents a trade-off between information and simplicity. Therefore, while a bookmark clearly represents what students believe is a relevant Web page, it is simple enough not to be dependent on other, more complex, abilities of the student, such as writing or argumentative skills. Nevertheless, one could argue that bookmarking captures only a partial view of more complex Web tasks, because students might in fact decide not to incorporate information from an already bookmarked page in a final assignment. Therefore, future research should explore the extent to which bookmarked Web pages are actually used in later tasks, such as essay writing (e.g. Wiley et al., 2009).

In summary, these results point to the need to clearly distinguish between students' Web page searching and their actual Web page use, such as when bookmarking pages for further study. Conclusions based only on navigation data could result in misleading views. From the point of view of a dual-processing model, students' navigation behavior might be described as a heuristic process that is to a certain extent characterized by the use of superficial relevance cues, such as order on the SERP. Conversely, a description of students' Web page use better fits with systematic processing, in which deep relevance cues are considered. The extent to which the use of deep relevance cues, such as topic relevance and trustworthiness, is dependent on students' background knowledge is discussed in the next section.

5.2 The role of background knowledge in trustworthiness evaluations

Among the Web page features that influence page bookmarks, three played an important role in our study: topic relevance of the Web page, link position, and trustworthiness of the page information. Most participants reported that they considered topic relevance when bookmarking Web pages on a conflicting topic, and they actually bookmarked relevant pages more often, as discussed above. However, the use of link position and trustworthiness as relevance criteria was moderated by the participants' background knowledge on the topic. Specifically, participants with high background knowledge reported that they bookmarked Web pages more often on the basis of the trustworthiness of the Web page information, and less often on the basis of link position, as compared to those with little background knowledge. This pattern of results suggests that the participants with high prior knowledge relied on deep, rather than on superficial relevance, cues in their assessment of Web pages. This concurs with the limited-resources hypothesis, which states that knowledgeable users are not overwhelmed by the task of assessing a page's topic relevance, as is the case with less knowledgeable users, so the former can devote more cognitive resources to assessing page trustworthiness (Bråten et al., 2011).

When it comes to the actual decision of bookmarking a page, the results reveal a more complex pattern. Actual bookmarks indicate that when the SERP included more topically relevant pages

at the top of the list and less relevant pages at the bottom (normal SERP), participants' background knowledge was related to a more systematic bookmarking of Web pages. In such a scenario, the participants with higher background knowledge more often bookmarked a topically relevant and trustworthy Web page than did those with lower background knowledge. This pattern of results concurs with the limited-resources hypothesis, in that students with high background knowledge might have additional cognitive resources available to assess not only the topic relevance of a page, but also its trustworthiness, which ultimately leads to the selection of a highly trustworthy page (Bråten et al., 2011).

Nevertheless, the results related to the bookmarking of the relevant and trustworthy page in the reversed SERP do not fit with our interpretation of this hypothesis. In a reversed SERP, in which the results are listed in a highly unexpected order, we predicted a greater involvement of resources to assess the topic relevance of pages. In this situation, students are expected to devote the most time and cognitive resources to assessing the topic relevance of pages, which should reflect in lower selections of not only a relevant, but also a trustworthy, page, as compared to a normal SERP. The data do not support this prediction. Alternatively, it could be expected that in a challenging reversed SERP, only individuals with a great amount of prior knowledge should be capable of devoting some time and resources to assessing the trustworthiness of Web pages. Again, this prediction is not supported by the data.

A possible explanation for this lack of effect of prior knowledge is that a reversed SERP might not have been sufficiently challenging for a sample of university students, in which case most of them would have been capable of bookmarking a relevant and trustworthy web page, independent of their background knowledge. Alternatively, we might consider that only students that have background knowledge and searched using the normal SERP would have sufficient cognitive resources available to judge the trustworthiness of the web page. In the reversed SERP condition, the non-normal order of results adds to the cognitive load, which might have been too

challenging for all participants, regardless of their background knowledge ². If this was the case, the fact that many participants selected the relevant and trustworthy web page from the reversed SERP would be due to other unexpected factors, such as it appearing first on the list before the other relevant, but not trustworthy, pages. To clarify this issue, future studies could employ additional measures to capture students' cognitive load, such as subjective ratings of the latter during a web search (e.g. Madrid, van Oostendorp, & Puerta Melguizo, 2009). Alternatively, a dual-task paradigm to experimentally relate students' processing load with evaluations of trustworthiness could be employed.

In summary, our findings indicate that students with a higher background knowledge more often based their bookmarks on the trustworthiness of the information on a page, and that in a normal SERP they actually more often chose a relevant and trustworthy web page, compared to students with little background knowledge. But what are the cognitive mechanisms underlying the evaluations of trustworthiness made by knowledgeable students? At least two possibilities are plausible: a strategic mechanism and a familiarity mechanism. Users with high prior background knowledge might be aware of the varying reliability of information sources dealing with the conflicting topic, and so might be more eager to strategically assess the pages' trustworthiness before deciding which page to bookmark. This mechanism corresponds to the behavior of experts in a field, as identified by Wineburg (1991): Experts initially search for information regarding the authors' credentials and consider the authors' points of view, while reading a set of documents on a conflicting topic. A second possibility might consider that users with high background knowledge are aware of a greater number of Web sources and are familiar with the positions and roles they play with respect to the conflicting topic, and so might rely on that source knowledge to decide on their bookmarks. Our future studies aim to clarify this issue.

5.3 Instructional applications

² We thank an anonymous reviewer for suggesting this alternative interpretation.

Our study relates to current efforts to understand students' Web literacy skills in complex learning scenarios, such as those involving searching for conflicting socio-scientific topics (Rouet & Britt, 2011). In accordance with this, our results offer new insights for the understanding of two key aspects of web literacy: how students assess topic relevance and the trustworthiness of Web pages when searching for information regarding conflicting topics, and when bookmarking this information for further study.

As an important step in Web searching, students must appropriately assess the topic relevance of Web pages provided by a SERP. Recent studies have suggested that middle school students might use rather superficial cues to assess a Web page's relevance, such as the presence of keywords from the search query in the page title provided by the SERP (Rouet et al., 2011). The pattern of results for bookmarking in our study reveals that undergraduate students are not attracted by such superficial cues of topic relevance to decide on their use of Web information. Two possible explanations for this discrepancy are the difference in the educational levels (middle school vs. university students), and the distinct tasks used (searching for simple facts vs. searching for conflicting topics to write a class report). Many middle school students may still not be completely fluent in comprehending complex expository texts, and may need to devote additional cognitive resources to decode word meanings (Perfetti, 2007). In this case, just comparing keywords in the Web page title with those in the search query could be interpreted as a tradeoff between a rather low investment of cognitive resources and a rather high probability of success. Although this explanation is tentative, it points to the need to increase our understanding of how students' Web literacy skills interact with other skills, such as reading comprehension abilities (Salmerón & García, 2011), and how they develop across educational levels (Rouet et al., 2011).

Conversely, we could assume that the difference between superficial or deep assessment of relevance cues is related to the students' task: while students might consider that superficial relevance cues are sufficiently predictive in searching for simple facts (such as in previous 'top link' heuristic studies, including Cutrell & Guan, 2007; Granka et al., 2004; Lorigo et al., 2006;

Pan et al., 2007), they might find the keyword approach risky to cope with the demands of searching for complex and conflicting topics. If this is the case, teachers could confront young students with the challenges inherent in searching conflicting socio-scientific topics on the Web. In this manner, teachers could raise students' awareness of the perils of assessing the topic relevance of Web pages on superficial cues only (Braasch, Bråten, Anmarkrud, & Strømsø, 2012). Although the explanation based on task differences is plausible, we must acknowledge that it is rather tentative, given that our experiment was not designed to directly address this issue, which is therefore open for future research.

A second relevant Web literacy skill considered in our study is trustworthiness evaluation. Our results show that while most participants reported they have used that criterion on which to base their bookmarking, only half actually bookmarked a highly topical, relevant and trustworthy page. This trend was particularly pronounced for participants with lower levels of background knowledge (cf. Gerjets et al., 2011; Walraven, Brand-Gruwel, & Boshuizen, 2009; Wiley et al., 2009). According to the limited-resources hypothesis (Bråten et al., 2011), a major problem is that most of the cognitive resources of students with little background knowledge are required to just comprehend the content of a Web page, therefore these individuals cannot simultaneously consider its trustworthiness. Different technological and instructional approaches could be considered to support students with little background knowledge in their use of trustworthy Web pages.

On the technological side, several solutions based on collective intelligence, that is, other users' behaviors, could be used to overcome the potential limitations of students with little background knowledge to enable them to select trustworthy information. Several Internet services, such as Reddit or Digg, include users' page bookmarks or votes in their SERP algorithms. Specifically, users' bookmarks are considered to be an implicit cue of the relevance and trustworthiness of the Web page, so the most bookmarked pages would have a higher probability of appearing among the top results. Negative votes may presumably go to less trustworthy pages, which would ultimately get a lower position in the SERP. Such algorithms might help students to

concentrate only on finding relevant information from a set of trustworthy pages provided on top of the SERP. Similarly, a few Web programs with the purpose of alerting users whenever they visit a page that includes a fact that has been challenged by a trustworthy source have been released. One of the most notorious of these programs is Dispute Finder (Ennals, Trushkowsky, & Agosta, 2010). The program functions as a browser extension, and highlights any statement in the pages being visited that resembles a conflicting claim, as shown by existing Web sites providing databases for such claims, as well as from other users' negative votes about such claims. By clicking on the highlighted statement, the user can access a popup window, in which that individual can read articles arguing for and against the claim.

A number of instructional interventions on how to evaluate the trustworthiness of information sources during Web searches have recently been developed, such as consideration of an information provider's expertise or motives, the source type, or the currency of the information (e.g., Walraven, Brand-Gruwel, & Boshuizen, 2010; Weber, Derrico, Yoon, & Sherwill-Navarro, 2009; Wiley et al., 2009). These interventions usually contain a mix of theoretical descriptions, and worked examples, as well as interactive exercises to apply the newly learned criteria and associated feedback. The first empirical findings reporting whether instructional measures have positive effects on users' evaluation of trustworthiness processes during a Web search for conflicting topics are promising. Alternatively, the limited-resources hypothesis (Bråten et al., 2011) proposes that instructors should look for methods of reducing the cognitive resources required during students' Web searching. A solution for instructors might be to provide students with some basic background information about an unfamiliar issue before allowing them to conduct a Web search on the topic. Preliminary evidence in support of this solution in middle school students has been provided by Rouet et al. (2011).

5.4 Limitations and future research

Finally, it should be acknowledged that the present study comes with certain limitations, which must be considered when interpreting the results. First, in the bookmarking task we explicitly

instructed participants to select the two pages that best covered the topic and to indicate whether or not they had considered a given set of criteria for their bookmarking. These self-reports might be prone to social desirability biases, and so might differ from their normal bookmarking and evaluation behavior (cf. Flanagin & Metzger, 2007). Second, for the sake of experimental control, we introduced some constraints in our study. That is, the participants conducted their Web search in a laboratory setting for a single predefined search topic with a finite set of only 10 Web pages. Moreover, the setting did not allow the participants to enter their own search terms in the search engine. Therefore, further research is required to extend the findings of the present study to other search topics and to more natural search situations. Finally, the present study did not examine what the participants had learned from their Web search, or to which information and sources they would refer when writing an essay about the searched topic. This should be addressed in future studies.

6. References

- Amadiou, F., Tricot, A., Mariné, C. (2011). Comprendre des documents non-linéaires: quelles ressources apportées par les connaissances antérieures? [Comprehension of non-linear documents: what is the role of prior knowledge?] *Année Psychologique*, *111*, 359-408.
- Amadiou, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction*, *19*, 376-386.
- Braasch, J. L. G., Bråten, I., Anmarkrud, Ø., & Strømsø, & H. I. (2012). Promoting high school students' evaluation of source features of multiple documents. *Paper presented at the annual meeting of the Society for Text and Discourse*, Montreal, Quebec.
- Braasch, J. L. G., Lawless, K. A., Goldman, S. R., Manning, F. H., Gomez, K. W., MacLeod, S. M. (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.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, 53, 1207-1217.

Bråten, I., Strømsø, H., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction*, 21, 180-192.

Cho, J., & Roy, S. (2004). Impact of search engines on page popularity. In S. Feldman, & M. Uretsky (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.

Eastin, M. S. (2001). Credibility assessments of online health information: the effects of source expertise and knowledge of content. *Journal of Computer-Mediated Communication*, 6(4).

Eisenberg, M. B., & Johnson, D. (2002). *Learning and teaching information technology*. Syracuse, NY: ERIC Clearinghouse on Information & Technology (No. ED465377). Retrieved from <http://www.gpoaccess.gov/eric/200212/ed465377.pdf>

Ennals, R., Trushkowsky, B., & Agosta, J. M. (2010). Highlighting disputed claims on the web. *Proceedings of WWW '10 Proceedings of the 19th international conference on World wide web* (pp. 341-350).

Evans, J.S. (2008). Dual-processing accounts of reasoning, judgment and social cognition. *Annual Review of Psychology*, 59, 255–278.

- 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: The MIT Press.
- Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How do users evaluate the credibility of Web sites? A study with over 2,500 participants. *In Proceedings of the 2003 Conference on Designing for User Experiences (DUX '03)* (pp. 1-15). New York, NY: ACM Press.
- 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.
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H.I. (2010). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology*, 31, 30-68.
- 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.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In *The Cambridge handbook of thinking and reasoning*. In K. Holyoak & R. G. Morrison (Eds.) (pp. 267-294). Cambridge: Cambridge University 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.
- Lewandowski, D. (2011). The influence of commercial intent of search results on their perceived relevance. In *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.
- Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., & Gay, G. (2006). The influence of task and gender on search and evaluation behavior using Google. *Information Processing and Management*, *42*, 1123-1131.
- Madrid, R. I., van Oostendorp, H., & Puerta Melguizo, M. C. (2009). The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order. *Computers in Human Behavior*, *25*, 66-75.
- 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.
- OECD. (2011). *PISA 2009 results vol VI: Student on line. Digital technologies and performance*. Paris: OECD. DOI: 10.1787/9789264112995-en

- 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.
- Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin, 92*, 382-402.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383.
- Petty, R. E., & Cacioppo, J.T. (1981). *Attitudes and persuasions: Classical and contemporary approaches*. Dubuque, IA: William C Brown.
- Rieh, S. Y., Kim, Y. M., Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing and Management, 48(6)*, 1136-1150.
- Rouet, J.-F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M.T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Greenwich, CT: Information Age Publishing.
- Rouet, J.-F., Ros, C., Goumi, A., Macedo-Rouet, M. & Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction, 21*, 205-219.
- Salmerón, L., & García, V. (2011). Comprehension skills and children's navigation strategies in hypertext. *Computers in Human Behavior, 27*, 1143-1151.
- 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.
- Savolainen, R., & Kari, J. (2006). User-defined relevance criteria in web searching. *Journal of Documentation, 62*, 685-707.

- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2009). How students evaluate sources and information when searching the World Wide Web for information. *Computers & Education, 52*, 234-246.
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H.P.A. (2010). Fostering transfer of web searchers' evaluation skills: A field test of two transfer theories. *Computers in Human Behavior, 26*, 716-728.
- Weber, B. A., Yoon, S., Derrico, D., & Sherwill-Navarro, P. (2009). Educating patients to evaluate web-based health care information: The GATOR approach to healthy surfing. *Journal of Clinical Nursing, 19*, 1371-1377.
- 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.
- Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*, 79.
- 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.

Table 1. Overview of the ten Web pages

Web page link title	Measures proposed	Site link	Source type	Words
Observatory - reducing emissions of greenhouse gases	Improve efficiency of petrol-based industry, improve efficiency of electrical appliances, and use of renewable energies.	www.cambioclimatico.org	Blog	688
GreenFacts : 6. How could we reduce emissions of greenhouse gases?	Improve efficiency of petrol-based industry, sustaining forests, and raising social awareness.	www.greenfacts.org	NGO	713
Reducing emissions of greenhouse gases	Gas emissions trade, promotion of carbon dioxide capture and storage, and sustaining forests.	www.un.org	Institution	636
Gloobal - Kyoto protocol to reduce emissions of greenhouse gases	Gas emissions trade.	www.gloobal.net	NGO	725
Activity 3: Reducing emissions of greenhouse gases	Sustaining forest.	www.tematea.org	Blog	630
Biodiesel could reduce emissions of greenhouse gases	Renewable energies: biodiesel.	www.biodisol.com	Commercial	607
Reducing emissions of greenhouse gases from hog waste	Renewable energies: biomass.	www.mma.es	Institution	672
ISTAS: Reduction of greenhouse gas emissions incompatible with road construction	Labor union denounces that road construction is against plans to reduce greenhouse emissions.	www.istas.net	Labor union	624
DNV verifies the reduction of greenhouse gas emissions in Kia Motors	Company verifies plans for slashing greenhouse gases from a motor car company.	www.dnv.es	Commercial	710
Certification of the reduction of greenhouse gas emissions	Spain obtains certificates for greenhouse gas emission reductions.	www.noticias.info	Newspaper	660

Table 2

Means (and standard errors) of the navigation behavior variables for the different Web page types as a function of SERP type.

Navigation data	SERP type	
	Normal	Reversed
Time (in s) spent on relevant-complete pages	79.86 (10.55)	52.52 (10.71)
Time (in s) spent on relevant-incomplete pages	38.67 (7.25)	58.19 (7.36)
Time (in s) spent on irrelevant pages pages	17.29 (6.35)	51.76 (6.44)
Number of relevant-complete pages visited (in %)	89.54 (4.41)	76.43 (4.48)
Number of relevant-incomplete pages visited (in %)	70.88 (5.30)	64.10 (5.38)
Number of irrelevant pages visited (in %)	57.90 (6.95)	68.62 (7.05)
Time (in s) spent on the relevant-trustworthy page	24.78 (4.24)	25.11 (4.30)

Table 3

Mean percentages (and standard errors) of bookmarks of a certain Web page type as a function of SERP type.

	SERP type	
	Normal	Reversed
% bookmarks of the relevant-complete pages	66.18 (5.56)	69.70 (5.64)
% bookmarks of the relevant-incomplete pages	32.35 (5.41)	28.79 (5.49)
% bookmarks of the irrelevant pages	1.47 (1.48)	1.52 (1.50)

Table 4

Percentage of participants that based their bookmarking of a certain Web page on the respective criterion.

Bookmark criteria	Normal SERP			Reversed SERP		
	RT	RU	RI+Irr	RT	RU	RI+Irr
Link position in the list	70.37%	33.33%	26.08%	7.14%	22.22%	25.00%
Page trustworthiness	81.48%	94.44%	91.30%	78.57%	83.33%	85.00%
Page relevance	96.30%	100%	91.30%	100%	100%	100%

Note 1. RU = relevant-untrustworthy pages, RT = relevant-trustworthy page, RI = relevant-incomplete pages, Irr = rather irrelevant pages

Note 2. We did not differentiate between RI and Irr pages, because in both the normal SERP and the reversed SERP only one participant had bookmarked an Irr page.

Note 3. The *N* for these analyses was 134 (68 in the normal SERP and 66 in the reversed SERP, respectively), because students bookmarked two pages in each scenario.

RUNNING HEAD: Searching the Web for conflicting topics



The image shows a screenshot of a search engine results page (SERP) for the query "REDUCCION DE EMISIONES DE GASES DE EFECTO INVERNADERO". The search bar at the top contains the query and a "Buscar" button. Below the search bar, there are options to search in "la Web", "páginas en español", and "páginas de España". The results section shows "Resultados 1 - 10 de 1.020.000 de REDUCCION DE EMISIONES DE GASES DE EFECTO INVERNADERO". The first ten results are listed, each with a title, a brief description, and a URL. The results include links to an observatory, GreenFacts, a UN report, a Kyoto Protocol page, a World Heritage site activity, a biodiesel article, a news item about reducing emissions, an ISTAS report, a DNV verification news item, and a government announcement about certified emissions reduction.

Resultados 1 - 10 de 1.020.000 de REDUCCION DE EMISIONES DE GASES DE EFECTO INVERNADERO

Observatorio - Reducir las emisiones de gases de efecto invernadero
Conseguir que la eficiencia sea rentable * Conseguir más electricidad, medios de transporte y producción industrial con menos carbón, petróleo o gasolina es ...
www.cambioclimatico.org/node/30 - 24k -

GreenFacts - 6 ¿Cómo podrían reducirse las emisiones de gases de ...
Desde 1995, los progresos técnicos para reducir las emisiones de gases de efecto invernadero han sido mayores de lo previsto. Las opciones durante los ...
www.greenfacts.org/es/cambio-climatico/n-2/cambio-climatico-6.htm - 47k

Reducción de las emisiones de gases de efecto invernadero
La mayor parte del progreso inmediato que se puede conseguir para reducir las emisiones de gases de efecto invernadero consiste en utilizar los combustibles ...
www.un.org/spanish/climatechange/background/ghg.shtml - 14k -

Global - Protocolo de Kyoto para reducir los gases de efecto invernadero
Aquellos países que han ratificado el Protocolo, y que han tratado de reducir las emisiones de gases de efecto invernadero incluso antes de entrar en ...
www.global.net/epala/global/fichas/ficha.php?entidad=Textos&id=790 - 21k -

Actividad 3: Reducir las emisiones de gases de efecto invernadero ...
Alentar a los administradores de sitios de Patrimonio Mundial a reducir las emisiones de gases de efecto invernadero a nivel de estos sitios: WHC Decisión ...
www.tematea.org/spanish/?q=node/2774&PHPSESSID=f6 - 44k -

El Biodiesel Podría Reducir las Emisiones de Gases de Efecto ...
Un estudio confirma que emplear biodiésel puro, o mezclado con combustible convencional, puede reducir las emisiones de gases con efecto invernadero ...
www.biodisol.com/biocombustibles/el-biodiesel/ - 61k

Reduciendo las emisiones de los gases de efecto invernadero de los ...
3 May 2007 ... La versión en inglés de "Reduciendo las emisiones de los gases de efecto invernadero de los desperdicios de cerdo" ("Slashing Greenhouse ...
www.mma.es/noticias/cambio_climatico/may07/hog0507.es.htm - 38k -

ISTAS: La reducción de emisiones de gases de efecto invernadero ...
En opinión de Comisiones Obreras, a España le va a resultar muy difícil reducir las emisiones de gases de efecto invernadero si el modelo de transporte ...
www.istas.net/web/index.asp?dpagina=3340 - 70k -

DNV verifica la reducción de emisiones de gases de efecto invernadero
DNV, la primera compañía acreditada por Naciones Unidas para validar y verificar la reducción de emisiones de gases de efecto invernadero, ...
www.dnv.es/certificacion/news/DNVreduccionemisionesgasefectoinve... - 50k -

Certificados de Reducción de Emisiones de gases de efecto ...
Primeros Certificados de Reducción de Emisiones de gases con efecto invernadero. /noticias.info/ - El Gobierno español se beneficiará de dichos certificados ...
www.noticias.info/archivo/2005/200510/20051024/20051024_111581.shtm - 33k -

Figure 1. Screen capture of the mock SERP