

The Effects of Credibility Cues on the Selection of Search Engine Results

Julian Unkel & Alexander Haas

LMU Munich

Author Note

Julian Unkel, Department of Communication Studies and Media Research, LMU Munich, and Munich Center for Internet Research, Bavarian Academy of Sciences; Alexander Haas, Department of Communication Studies and Media Research, LMU Munich.

This research was supported by the LMUexcellent Junior Researcher Fund. The authors would like to thank Nora Denner, Lena Storck, and especially Ramona Berger for their assistance with experiment supervision and coding, and the reviewers for their valuable suggestions.

Correspondence concerning this article should be addressed to Julian Unkel, Department of Communication Studies and Media Research, LMU Munich, Oettingenstr. 67, 80538 Munich, Germany. E-mail: unkel@ifkw.lmu.de

Abstract

Web search engines act as gatekeepers when people search for information online. Research has shown that search engine users seem to trust the search engines' ranking uncritically and mostly select top-ranked results. This study further examines search engine users' selection behavior. Drawing from the credibility and information research literature, we test whether the presence or absence of certain credibility cues influences the selection probability of search engine results. In an observational study, participants ($N = 247$) completed two information research tasks on preset search engine results pages, on which three credibility cues (source reputation, message neutrality, and social recommendations) as well as the search result ranking were systematically varied. The results of our study confirm the significance of the ranking. Of the three credibility cues, only reputation had an additional effect on selection probabilities. Personal characteristics (prior knowledge about the researched issues, search engine usage patterns, etc.) did not influence the preference for search results linked with certain credibility cues. These findings are discussed in light of situational and contextual characteristics (e.g., involvement, low-cost scenarios).

Keywords: credibility, selection, search engines, observation, quantitative methodology, heuristics

The Effects of Credibility Cues on the Selection of Search Engine Results

Mass media use requires a certain degree of selection by the recipient. This has been true for traditional media but applies even more to the use of digital media. Seventeen years ago, Wirth and Brecht (1999) already diagnosed an increased necessity for the recipient to (actively) select content, as compared to traditional media. Since then, the *amount* of online content available has increased drastically, which should further underscore the need to actively select content. Most web users rely on search engines to tame this proverbial information tide and find their desired information (Purcell, Brenner, & Rainie, 2012).

However, from the user's perspective, the online environment differs from traditional media not only by the number of applications and the amount of information available. Recipients are also confronted with a higher degree of *uncertainty* regarding the source and quality of the retrieved information. When users evaluate search engine results pages (SERPs), they face the following situation: Compared to traditional media, the number of content providers known by the user—and thus their motives, ethics, etc.—is substantially lower.¹ Financial and organizational requirements for creating online content are largely insignificant, while at the same time, control and filtering mechanisms are not as established online as in traditional media (cf. Rieh & Danielson, 2007). Assessing the credibility of online content thus becomes the responsibility of the recipient (Metzger, Flanagin, Eyal, Lemus, & McCann, 2003). Neglecting this responsibility is accompanied by the risk of getting exposed to wrong or even harmful information. We are thus interested in the role of credibility cues signifying expertise and trustworthiness of search results in search engine selection behavior.

Large-scale log file analyses (Agichtein, Brill, Dumais, & Ragno, 2006; DigitalRelevance, 2013) on search engine selection behavior lead to the conclusion that users are

not fully aware of their responsibility to evaluate online content on their own and are guided predominately by the search result ranking. Eye-tracking studies corroborate this finding: Even though users perform relevance judgments, the ranking still emerges as the strongest predictor of selection probability (Pan et al., 2007). The ranking furthermore influences time afforded to and the quality of these subjective judgments (Balatsoukas & Ruthven, 2012). Our study thus considers the overarching question of to what degree credibility cues are able to influence selection decisions when competing with the seemingly more dominant ranking cue.

In a first step, we analyze the effects of three systemically varied credibility cues—a reputation cue, a neutrality cue, and a social recommendation cue—on the selection behavior for two informational search tasks. We then examine whether interindividual differences in the propensity for selecting search results with specific credibility cues, based on personal characteristics such as prior knowledge, internet use, and search engine usage strategies, exist.

Literature Review

When users are searching information on the Internet, the first distinction to be made is whether they know a specific webpage where to (probably) find said information or not. In the first case, users can either navigate to the webpage directly by opening a bookmark or typing the uniform resource locator (URL), or employ search engines to navigate to a known source of information. Broder (2002) coined the term *navigational tasks* for search queries with an intent to reach a particular site, for example, searching for “nytimes” to get to the homepage of the New York Times. In the second case, if users do not know where to find the desired information, they likely perform an *informational task*, formulating a closed or open-ended search query. Tasks like this have led to search engines being considered new “gatekeepers” (Machill, Beiler, & Zenker, 2008, p. 591). Although their role is not identical to the traditional understanding of

journalistic gatekeepers (that is, deciding whether some content or information is published at all), they play a crucial role in determining whether said content can be found and accessed. Even well-known news sites such as NYTimes.com, WashingtonPost.com, and CNN.com receive about 15% to 30% of their traffic through search engines.² This gatekeeper role emphasizes the need to understand the selection criteria users employ when they use search engines.

Search Engine Rank as a Selection Criterion

In information foraging theory, information retrieval online is conceptualized as web users following the “information scent” (i. e., the perceived value of an information source; Pirolli & Card, 1999, p. 646) derived from proximal cues to distal sources of content. For example, the information present in a search result snippet (e. g., the headline and textual summary, the displayed URL) provides the user with information scent cues about the linked website (Pirolli, 2005). However, as the design of SERPs is generally rather sparse and simple, search results offer only a limited information scent (Schwarz & Morris, 2011). It is thus of no surprise that the most important factor influencing search result selection is the most dominant cue present on SERPs: the ranking. Most users select one of the top-ranked results, in many cases the first result at the top (Agichtein et al., 2006; Machill, Neuberger, Schweiger, & Wirth, 2004), as a higher rank signals higher relevance and also increases the chance of a search result being part of the “visible area” (the part of the SERP visible without scrolling down; Lewandowski, 2013, p. 188). Furthermore, most users select a result on the first SERP. A comparative meta-analysis of several studies on user behavior in search engines shows that this primacy effect has gotten even stronger over the years (Jansen & Spink, 2006). In 2000, almost three quarters of all search engine users never looked past the first SERP. This indicates that

users learned to trust search engines and assume that they will provide the most relevant results for search queries on the first page. Following this widely verified primacy effect, we propose:

H1: The higher the rank of the search result, the higher the probability of selecting it.

However, empirical results indicate that—despite the preference for top-ranked results—users employ other strategies to select search results. A survey by Beiler (2005) revealed several dimensions of selection strategies, showing that users may also select content by evaluating the whole SERP, by choosing a result randomly, or by following their intuition. Therefore, criteria in addition to the rank of the results may be relevant as well. Following our premise that perceived credibility should influence selection decisions, we will thus concentrate on credibility cues that can be observed on SERPs. After all, in most situations, if perceived as unbelievable, “the news is simply not worth reading, let alone exploring” (Sundar, Knobloch-Westerwick, & Hastall, 2007, p. 369).

Credibility Cues as Selection Criteria

The central role of credibility judgments in influencing strength (or generally, the *existence*) of media effects has been examined by scholars since the mid-20th century. Thus, the large body of work on credibility has produced “plentiful, contradictory, and confus[ing]” (Self, 2009, p. 435) conceptualizations and definitions. One common ground, though, is the conceptualization of credibility as a subjective characteristic perceived by the recipient, and not as a stable feature of a (media) object (Choi & Stvilia, 2015). For the recipient, credibility may function as a cognitive shortcut (O’Keefe, 2008): New information does not need to be extensively checked and evaluated but can be believed (or not believed) based on assumptions about the credibility of the source or the message. Thus, perceived credibility is a key factor in situations (a) that confront the recipient with information that cannot (or only with great effort)

be verified and (b) where a large amount of (maybe even contradicting) information is available. Both characteristics apply to the selection of online content.

The perception of credibility can be further differentiated by the reference object: Recipients may make assumptions about the *source* of information (e.g., the website where the information is found), the *message* that includes the information, and, more broadly, the *media* (e.g., the Internet) in general (Metzger et al., 2003). Furthermore, the perception of credibility is based on an evaluation of multiple dimensions. Focusing on source credibility, Hovland and colleagues (1953) identified expertise and trustworthiness as core dimensions of credibility. In this sense, credibility includes a factual and a moral component.³

Research on the *effects* of credibility online has been sparse and focused on the postcommunicative phase. In line with traditional persuasion research, the effects of credibility on attitudes and behavioral intentions was examined in the context of online health communication (e.g., Hu & Sundar, 2009). The precommunicative phase has received less attention thus far, which is why we focus on the effects of credibility on content selection. In this phase, credibility cues allow the user to perform predictive judgments about the credibility of the not-yet selected media object (Rieh, 2002). In the following sections, we present an overview of credibility cues that may influence such predictive judgments when users employ search engines. These cues are differentiated by the presumed strength of their effects on expertise or trustworthiness; however, most cues likely influence both credibility dimensions.

Credibility cues linked to perceived expertise. In general, the short snippets and uniform design of search results offer only few cues for credibility judgments (Yamamoto & Tanaka, 2011); however, many cues examined by online credibility studies at least may be present on SERPs. Out of 51 website elements and features analyzed by Fogg and colleagues

(2001), seven were related to the expertise dimension, including affiliation with a well-known, respected (offline) news organization, highlighting of author credentials, and a list of citations and references. Commonly, *information about the authors*—details about their backgrounds and qualifications as well as portraits—can lead to higher perceived expertise, and thus credibility (K. A. Johnson & Wiedenbeck, 2009). Although generally there is little room for such information in the deliberately simple and concise SERPs, Google, for example, offered website authors to show the author name and portrait picture next to the search result until the feature was shut down in 2014. Lists of citations and references are one component of *information completeness* (Dutta-Bergman, 2004), which includes annotations, methodological details, or relevant statistics for a news story. Although information completeness may be fully evaluated only by opening the result, the snippet may give a first impression. The most important factor when expertise was considered as voted by Fogg and colleagues' participants, however, was the affiliation with a well-known brand (2001). A site's *reputation* thus is a key factor of perceived credibility. Other scholars (Flanagin & Metzger, 2007; Greer, 2003; Meyer, Marchionni, & Thorson, 2010) came to the same conclusion: News sites of well-known organizations are perceived as more credible than other sites (e.g., blogs) that have not established an expert image. As all search results contain at least some information about the source via the displayed URL, this information has the potential to function as a reputation cue in every search scenario. We thus pose the following hypothesis:

H2: High reputation search results are more likely to be selected than low reputation ones.

Credibility cues linked to perceived trustworthiness. According to the literature (Fogg et al., 2001; Metzger et al., 2003), most cues that affected the perceived trustworthiness of a

source are located either at the target sites (e.g., *absence of advertisements, linking of sources*) or even at subpages of these sites (e.g., *ethics codices* on the imprint page). Another important criterion is the *website linked to the evaluated site*—in the case of selection on SERPs, the search engine itself. Thus, if the search engine is perceived as trustworthy, linked sites will likely be perceived as more trustworthy as well; the more so as some users assume that results are ranked by search engines according to their credibility (Nakamura et al., 2007). A higher rank on the SERP should thus lead to higher perceived trustworthiness (cf. Westerwick, 2013). Additional criteria that may influence the perceived trustworthiness on a source level and that are present on SERPs are the *name* and the *URL of the linked site*. Top-level domains such as .edu, .gov., or .org may refer to the noncommercial background of the linked site (Wathen & Burkell, 2002), while the name of the site might either already be known and thus linked with a certain trustworthiness, or might refer to noncommercial intentions or the unbiased background of the source.

Apart from these source indicators, trustworthiness might also be perceived at the message level. Conceptualizations of trustworthiness generally include criteria such as fair and unbiased reporting (Gaziano & McGrath, 1986; McCroskey & Teven, 1999). Thus, the *neutrality* of the linked site, as indicated by textual content of the snippet, should positively influence the perceived trustworthiness. Especially on controversial issues, messages are perceived as more fair and more trustworthy if they are two-sided (i.e., include arguments for both sides; Winter & Krämer, 2012). We thus propose as our third hypothesis:

H3: Neutral search results are more likely to be selected than biased ones.

As Self (2009) noted, recent research interest has shifted toward collaborative and interactive credibility evaluations of active users. With web tools, features, and technologies linked to the buzzwords *social media* and *Web 2.0*, users can include the experiences and

evaluations of third persons in the users' own credibility judgments (cf. Metzger, Flanagin, & Medders, 2010). This influence of third parties has been described as "reputed credibility" (Tseng & Fogg, 1999, p. 42). Research on how such *social recommendations* affect the selection of online content has been predominantly carried out in e-commerce contexts (cf. Park, Lee, & Han, 2007; Sun, 2012). Furthermore, the effect of social recommendations on selection decisions has also been found in news contexts (Leino, Rähkä, & Finnberg, 2011; Xu, 2013). However, social recommendations might also be relevant in search engines, as some have experimented with social features in recent years. Google, for example, enables site authors to include product ratings. First, such social recommendation cues provide additional visual anchors. Second, they are also able to enhance the information scent of a search result, as they provide some indication of the quality of the search result beyond the textual content of it (Sundar et al., 2007). An eye-tracking experiment showed that such social recommendations may indeed positively influence the selection probability of search results, but to a lesser degree than other factors such as thematic relevance (Terbeck, 2012). We thus propose as our fourth hypothesis:

H4: Search results with a high number of social recommendations are more likely to be selected than search results with a low number.

Additivity and perception of credibility cues. All three credibility cues (reputation, neutrality, social recommendations) might be present simultaneously in a search result. Thus, recipients are likely to perceive and process these cues at the same time. Sundar and colleagues (2007) adopted the additivity hypothesis from the psychological literature on credibility and proposed a cue-cumulation effect: Evaluation heuristics triggered by individual credibility cues would add up to increase the overall credibility perception of a reference object (in our case, the individual search result) if two or more cues were present at the same time. However, the authors

argued that one cue might also be sufficient for users' credibility judgments; the cue perceived as the most important would thus exclude the other cues from contributing to the credibility evaluation. We wanted to investigate this further and considered the following research question:

RQ1: Do credibility cues add up, thus leading to increased selection probability if two or more cues are present at the same time?

Because credibility is a subjective, perceived characteristic, personal factors can influence the likelihood of credibility cues being evaluated. Drawing from cognitive psychology literature, Metzger (2007) proposed a dual processing model of website credibility assessment. She found that individual motivation and ability moderate credibility evaluations, leading to either a systematic, central evaluation, a heuristic, peripheral evaluation, or no evaluation at all. In our case, central processing should lead to a stronger emphasis on the information background and content of the search results (i.e., the three credibility cues), while peripheral processing should increase the influence of surface cues, especially of the search results' rank. Personal factors that influence central or peripheral evaluation in the search engine context include prior knowledge on the respective issue: Web search users with a high domain expertise tend to invest more time and effort into search sessions than users with little prior knowledge (White, Dumais, & Teevan, 2009). In addition, the experience with the Internet in general as well as search engines in particular should moderate the ability to evaluate (in general, Internet expertise and domain knowledge lead to a more thorough and successful search; Wirth, Sommer, von Pape, & Karnowski, 2015). Beiler (2005) also showed that users employ different strategies when the users work with search engines. Although intuitive selection should lead to a peripheral credibility evaluation, evaluative selection is assumed to trigger central processing. Finally, since credibility is a subjective quality, users evaluate credibility cues in different ways. Individual

media use patterns, for example, have been shown to influence credibility evaluation. Although, generally speaking, news sites are rated as more credible, several studies have shown that frequent users of blogs attribute higher credibility to blogs than to traditional media (T. J. Johnson & Kaye, 2004; Schmierbach & Oeldorf-Hirsch, 2012). We thus pose as our second research question:

RQ2: How do Internet and search engine use, media use patterns, previous knowledge, and selection strategies influence the preference for search results with different credibility cues?

Method

To test the hypotheses and address the research questions, an observational study of search behavior was carried out in a research laboratory of a large German university. Following several introductory questions about Internet and search engine use, participants were instructed to gather information about an issue by using a preset search query. Afterwards, several questions about the issue had to be answered.⁴ For this purpose, a manipulated SERP of the search engine DuckDuckGo was included in the questionnaire via iFrame (see Figure 1). The participants navigated this SERP and clicked on the results as in any other browser; the only requirement was not to change the search query.⁵ Each participant had 5 min for his or her research, after which the questions about the research issue were displayed. All participants could also manually stop their research before the 5-min period ended. This research task was then repeated with a second issue, after which another questionnaire was displayed. All participants received 10€ for their participation.

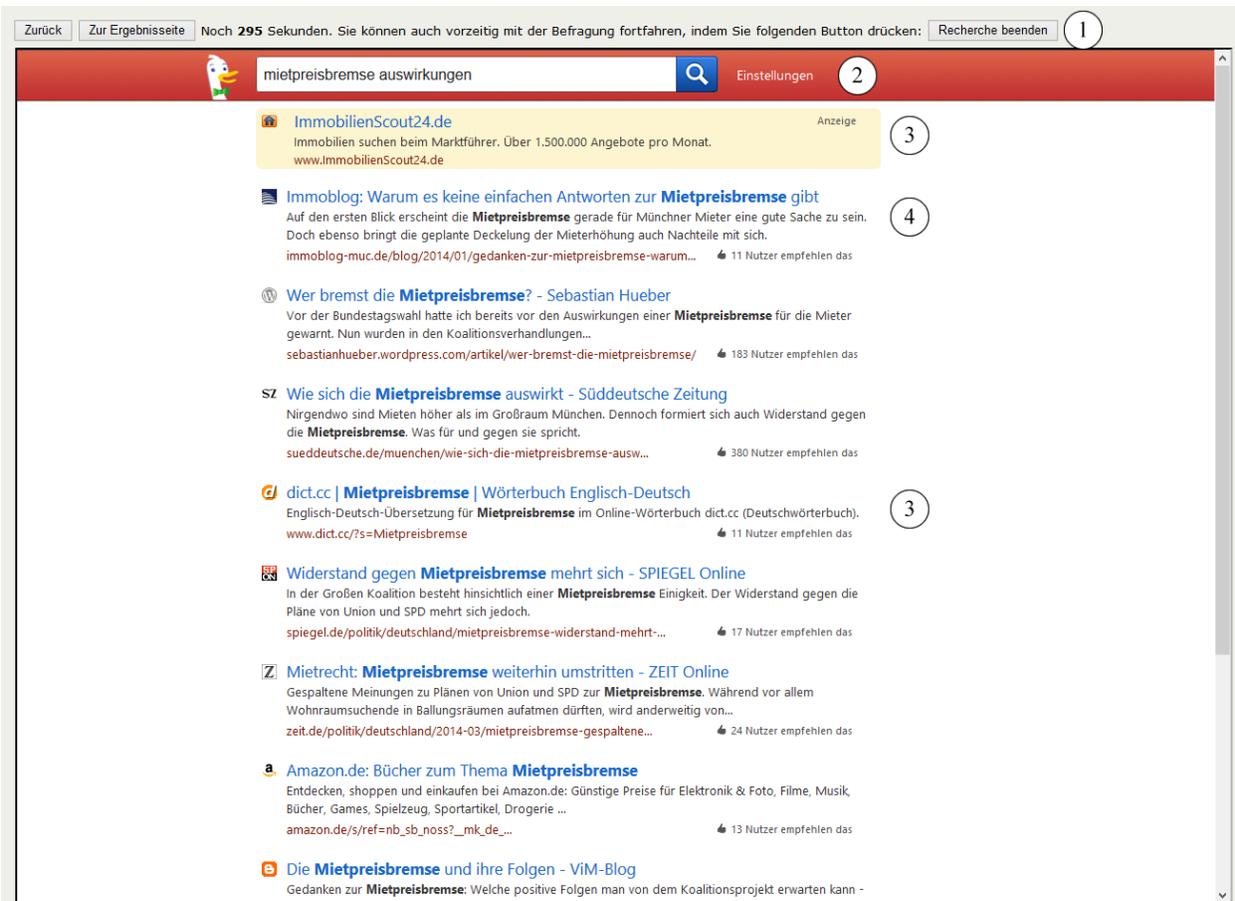


Figure 1. Manipulated SERP with (1) rudimentary browser controls and countdown timer, (2) preset search query, (3) advertisement and irrelevant search results, and (4) relevant, manipulated search results.

Stimulus

To create a natural information research situation, both SERPs as well as all target pages were built as fully functional web pages in HTML. Each SERP featured eight search results relevant to the research issue as well as one advertising link at the top of the SERP and two irrelevant results (e.g., a link to a shopping site) at ranks 4 and 8, respectively. In this setup, five relevant search results were fully visible in the visible area, with one result partly visible and two results in the invisible area.

The SERPs were systematically manipulated for both research tasks. First, the participants were randomly assigned to one of eight stimulus versions (i. e., SERPs), each of which presented a different *ranking* of the search results, with each result once at rank 1, once at rank 2, etc. Second, the credibility cues (a) *reputation*, (b) *neutrality*, and (c) *social recommendations* were varied on two factor levels (high/low) within each search result (see Table 1). Each factor level appeared four times, and each possible combination between the three factors appeared once on the SERP.

Table 1
Credibility cue variation

Credibility cue	Description	Indicated by	Sample wording – high condition	Sample wording – low condition
Reputation	Reputation of target site linked in search result; either a well-known news site or a personal website/blog	Site title named in headline of search result; displayed URL; displayed site favicon	Sueddeutsche.de, Zeit.de (well-known German news sites)	Immoblog-muc.de, serienorakel.de (little known personal websites or blogs)
Neutrality	Two-sided or one-sided argument on the search task’s issue (for each issue, all one-sided search results took the same stance toward the issue)	Distinct wording in headline and summary of the search result	“Online streaming of TV series: pros and cons”	“Who stops the rent control mechanism?”
Social recommendations	Three-digit or lower two-digit number of social recommendations	Displayed next to the search result	“380 users recommend this”	“17 users recommend this”

Real news and personal websites were used as templates for the target pages. All target pages featured news articles or blog entries with the same informational content. The issues selected for the research tasks were the effects of the proposed implementation of a rent control mechanism in Germany for the first research task (rent control, henceforth) and legal consequences of streaming TV series online for the second research task (streaming, henceforth).

Both issues were selected to improve external validity, since they are relevant to students in Germany.⁶

Usability, adequacy, and comprehensibility of the stimulus were pretested with nine participants who did not take part in the actual study. They were asked to think aloud while completing both research tasks. Changes made after the pretest included the addition of the two irrelevant search results, as two participants deemed the SERP as “too perfect”, and some wording refinements, as three participants noticed textual similarities between two target pages.

Measures

Observed research behavior. All research tasks were captured on video via screen capture software and coded by three student research assistants. All selected search results and the order in which they were selected were coded as well as the dwell time on the SERP before the first selection.

Prior knowledge. Prior knowledge on both issues was measured on a Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) consisting of four items: “I’ve extensively dealt with this topic before,” “I’ve already formed a strong opinion on this topic,” “I know the most important facts about this topic,” and “I could convince others of my stance on this topic” (rent control: $\alpha = .86$, $M = 2.16$, $SD = 0.88$; streaming: $\alpha = .90$, $M = 2.70$, $SD = 1.03$).

Selection strategies. The individual style of search engine use was measured using an shortened version of Beiler’s (2005) scale. A principal component analysis⁷ revealed three distinct factors: ranked selection (sample item: “I choose results that are ranked at the top of the list”), intuitive selection (e.g., “I choose results spontaneously”), and evaluative selection (e.g., “I deliberate about which result to click”).

Media use. Use of all six real news websites linked on the SERPs (two news sites were featured in both research tasks; mean index: $\alpha = .75$, $M = 2.59$, $SD = 0.81$) and of blogs ($M = 1.93$, $SD = 1.18$) was measured with the question “How often do you use the following websites to read up about recent events?” on a 5-point scale from 1 (*never*) to 5 (*daily*).

Internet and search engine use. Internet use per day was measured on a 6-point scale ranging from 1 (*less than 30 minutes*) to 6 (*more than 6 hours*). Seventy-four percent of the participants used the Internet more than 2 hours per day. Search engine use was measured on a 5-point scale ranging from 1 (*less than weekly*) to 5 (*several times a day*). Ninety-two percent of the participants used search engines at least daily.

Sample

A sample of 247 students participated in the study. One hundred seventy-one (69%) participants were female. On average, the participants were 22 years old ($Mdn = 21.00$, $SD = 3.50$). All participants were randomly assigned to one of eight groups, determining the stimulus versions for both research tasks. The groups did not differ by gender, $\chi^2(7, n = 247) = 5.49, p = .60$, or age, $F(7, 239) = 1.26, p = .27$. Except for one person, all participants used search engines at least multiple times per week. We thus based our stimulus on the little known search engine DuckDuckGo, so that the participants would not easily recognize our modifications. As expected, DuckDuckGo was not familiar to the majority (88%) of the participants.

Results

The following analyses include only clicks on relevant search results, that is, the respective eight search results with the manipulated credibility cues. Clicks on advertisement links and irrelevant search results (e.g., translation service) were excluded from these analyses.

Research Task Descriptives

The participants spent a brief period of time evaluating the SERP before they selected their first result ($M_{\text{rent control}} = 14.66$, $SD = 11.94$; $M_{\text{streaming}} = 9.35$, $SD = 7.56$). Most participants utilized the full 5 min given for their research tasks ($M_{\text{rent control}} = 280.24$, $SD = 52.71$; $M_{\text{streaming}} = 273.95$, $SD = 55.52$) and on average selected three search results per research task ($M_{\text{rent control}} = 3.18$, $SD = 1.16$, $M_{\text{streaming}} = 2.89$, $SD = 1.05$), with a maximum of seven selected results in both research tasks. Because the research tasks were in the same order for every participant, this might have led to a learning effect that explains the shorter selection time spans as well as the decrease in selections for the second research task.

As expected, the participants most often selected the search result displayed on rank 1, with percentages steadily decreasing towards rank 7 (see Table 2). The search results invisible (rank 7 and 8) or only partly visible (rank 6) without scrolling were selected considerably less often and only account for one-digit percentages.

Table 2
Selection of Results by Rank

	Rank ^a							
	1	2	3	4	5	6	7	8
Percentage of selected results	23.1	16.6	16.1	12.9	11.5	7.0	5.9	6.9

Note. $n = 1,500$; ^aRank on SERP, 1–8, from top to bottom

Table 3 shows the selected search results by individual factor combinations and factors, differentiating between the first selected results and the following selected results for each research task. Generally, high reputation search results were selected more often than low reputation search results. Considering the first selection decision only, more than three quarters of all selected search results had a high reputation (rent control: 77.3%; streaming: 75.7%). Of the following selection decisions, about two thirds of the clicks were still on high reputation

results. High-neutrality results were selected more often than low-neutrality results, especially during the first selection decision (64.4% and 59.9%, respectively). High social recommendation results were selected more often than low social recommendation results only during the first selection decision in the first research task (58.3%). During all other selection decisions, the high and low social recommendations results garnered about half of the selected results each.

Table 3
Selection of Results by Factors and Factor Combinations

Factor combination Rep / Neu / SoR	Percentage of selected results				Total (n = 1,500)
	Rent control task		Streaming task		
	First selection (n = 247)	Following selections (n = 538)	First selection (n = 247)	Following selections (n = 468)	
+ / + / +	34.0	18.0	19.4	16.7	20.5
+ / + / -	18.2	15.8	27.1	16.7	18.4
+ / - / +	11.7	14.1	22.3	15.2	12.9
+ / - / -	13.4	17.3	6.9	15.6	16.9
- / + / +	7.3	10.4	7.7	12.2	10.0
- / + / -	4.9	9.7	5.7	8.1	7.8
- / - / -	5.3	7.8	7.7	9.2	7.8
- / - / +	5.3	6.9	3.2	6.4	5.8
Factors					
High reputation	77.3	65.2	75.7	64.1	68.6
High neutrality	64.4	53.9	59.9	53.6	56.5
High SoR	58.3	49.4	52.6	50.9	51.9

Note. Rep = Reputation; Neu = Neutrality; SoR = Social recommendations. Percentages above 100 due to rounding.

Influences on Selection Decisions

However, these credibility cues do not appear isolated from each other, making it necessary to investigate the observed effects within one model. Furthermore, with each SERP providing eight viable options in the form of search results for the participants, the decision to open one result is at the same time a decision *not* to click on the other seven results. Therefore, the analysis of the effects of the credibility cues as well as the results' rank within the SERP also

must account for results (and their cues) that were not selected. This restructuring of data led to a binary dependent variable that differentiates between “search result selected” (1) and “search result not selected” (0). Thus, hierarchical logistic regressions were calculated based on the selection decisions for the first selection decision (i.e., each participant’s first selected result—and the seven results not selected while doing this—during each research task) and the following selections decisions (see Table 4). To analyze the individual influence of the predictors, we consider the regression coefficient b , the Wald statistic W and the odds ratio OR . Model fit was examined using Nagelkerke’s R^2 and classification statistics.

The model for the first selection decision had a good fit ($R^2 = .40$). Rank had the strongest influence on the selection: The higher the result was ranked on the SERP, the more likely it was selected ($b = -0.78$, $W = 413.08$, $p < .001$, $OR = 0.46$). Of the three credibility cues, only reputation still shows an effect, with high reputation results being selected more often ($b = 0.98$, $W = 19.23$, $p < .001$, $OR = 2.66$). The analysis also showed a (weak) interaction effect between reputation and neutrality ($b = 0.77$, $W = 9.61$, $p < .01$, $OR = 2.15$). To clarify this interaction, Step 2 in the regression model had to be considered, where reputation and neutrality show an effect. However, adding the interaction effect in Step 3 canceled out the individual effect of neutrality, thus indicating that the effect of neutrality on selection was present only when high reputation results were considered. In total, the model classified 89.3% of all search results correctly as either selected or not selected (32.6% of selected results and 97.5% of not selected results).

Table 4
Logistic Regression: Influences on Selection Decision

Model	First selection decision			Following selection decisions		
	<i>b</i>	<i>W</i>	OR	<i>b</i>	<i>W</i>	OR
Step 1						
Rank ^a	-0.71	409.81***	0.49	-0.16	102.56***	0.85
Constant	0.40	15.49***	1.49	-0.95	151.36***	0.39
Nagelkerke's R ²		.30			.03	
Step 2						
Rank ^a	-0.77	414.71***	0.46	-0.19	134.32***	0.82
Reputation ^b	1.66	175.98***	5.27	1.09	212.96***	2.98
Neutrality ^b	0.75	41.83***	2.11	0.37	25.53*	1.44
Social recommendations ^b	0.34	9.10**	1.41	0.12	2.73	1.26
Constant	-1.02	42.60***	0.36	-1.63	260.96***	0.20
Nagelkerke's R ²		.40			.09	
Step 3						
Rank ^a	-0.78	413.08***	0.46	-0.20	134.84***	0.82
Reputation ^b	0.98	19.23***	2.66	0.95	53.86***	2.60
Neutrality ^b	0.24	1.10	1.28	0.41	0.09	1.04
Social recommendations ^b	0.05	0.04	1.05	-0.16	1.27	0.85
Rep × Neu	0.77	9.61**	2.15	0.18	1.40	1.19
Rep × SoR	0.47	3.65	1.60	0.07	0.22	1.07
Neu × SoR	-0.05	0.04	0.96	0.44	9.24*	1.55
Constant	-0.54	6.83**	0.59	-1.41	113.75***	0.24
Nagelkerke's R ²		.40			.10	

Note. *b* = regression coefficient, *W* = Wald statistic (df = 1), *OR* = odds ratio; Rep = Reputation; Neu = Neutrality; SoR = Social recommendations.

n = 3,984 (first selection decisions), *n* = 6,256 (following selection decisions). ^aRank on SERP, 1–8, from top to bottom. ^bCategorical variable, 0/1-coded (1 = high characteristic).

* *p* < .05, ** *p* < .01, *** *p* < .001.

The analysis of the following selection decisions differed in two regards from the first model. First, in the first model, each participant was represented by two selection decisions (i.e., her or his first clicks during the first and second research task). However, the participants were free to open as many results as they wanted. Thus, in the following selection decisions model,

different participants might be represented by a varying number of selection decisions based on how many results they opened altogether during both research tasks after the first click. Second, the more results the participants opened, the more the number of available options decreased, provided that participants did not choose to open previously selected results again (as on most search engine, previously selected results were indicated by a different link color). Thus, in the following selection decisions model, previous selection decisions were considered with previously opened results excluded from the analysis (e.g., for the third click, only the five results not selected—that also were not selected during the first and second selection decision—are assigned to the selected result).

As in the first model, the participants preferred results that were ranked at the top of the SERP. Because most participants had chosen a top-ranked result during the first selection decision, the influence of the rank was comparably lower in the following selection decisions but remained the strongest predictor ($b = -0.20$, $W = 134.84$, $p < .001$, $OR = 0.82$). In turn, the influence of reputation increased, and high reputation results again were selected more often than low reputation results ($b = -0.96$, $W = 53.86$, $p < .001$, $OR = 2.60$). Taken individually, neutrality and social recommendations still had no meaningful influence. The interaction effect of reputation and neutrality observed in the first model disappeared in the following selection decisions model. However, there was an (albeit weak) interaction effect between neutrality and social recommendations ($b = -0.44$, $W = 9.24$, $p < .001$, $OR = 1.55$). In comparison with the first model, goodness of fit decreased considerably ($R^2 = .10$). In addition, in the second model the selection probability was never predicted to be more than the threshold value of 50%; thus all search results were classified as not selected. These results indicated that after a deliberate first selection, the following search results were selected randomly.

In total, these analyses supported H1 (rank) and H2 (reputation) but not H3 (neutrality) and H4 (social recommendations). Regarding a cue-cumulation effect (RQ1), we found weak interactions of neutrality with reputation (first selection decisions) and social recommendations (following selection decisions).

Interindividual Differences in Credibility Cue Preference

Thus far, the analyses focused on the effect of credibility cues on selection decisions. However, RQ2 asked about the effect of individuals' features on the preference for results with these credibility cues (e.g., high reputation results). Thus, the following analyses did not focus on selection decisions, but only on the selected results. We also concentrated on the first selection decision, since the previous analyses indicated that the following selection decisions were random. For each of the three credibility cues, a logistic regression was calculated that differentiated whether each case's first selected search result had a high (1) or low (0) characteristic of the respective credibility cue (see Table 5).

All three models have a very low goodness of fit (reputation: $R^2 = .05$; neutrality: $R^2 = .02$; social recommendations: $R^2 = .02$). The individual preferences for search results with high reputation, high neutrality, or high social recommendations were thus not predicted by individual characteristics such as Internet use, selection strategies, or prior knowledge. Regarding the selection of high reputation results, the expected influence of media use was at least partly found: Participants who reported they often used well-known news sites showed a greater tendency to select high reputation results (i.e., results that link to one of those well-known news sites; $b = 0.42$, $W = 7.76$, $p < .05$, $OR = 1.52$). However, high blog use did not correlate with a preference for low reputation results. All other personal characteristics did not show any meaningful effect in any of the three models.

Table 5
Logistic Regressions: Influences on Credibility Cue Preference

Model	Reputation			Neutrality			Social recommendations		
	<i>b</i>	<i>W</i>	<i>OR</i>	<i>b</i>	<i>W</i>	<i>OR</i>	<i>b</i>	<i>W</i>	<i>OR</i>
Internet use ^a	-0.05	0.24	0.95	0.04	0.15	1.04	-0.02	0.03	0.98
Search engine use ^b	0.05	0.09	1.06	0.04	0.06	1.04	-0.09	0.34	0.91
Selection by rank ^c	-0.13	1.35	0.87	0.09	0.74	1.09	-0.11	1.34	0.89
Spontaneous selection ^c	0.06	0.26	1.06	0.10	1.02	1.11	-0.05	0.25	0.95
Evaluative selection ^c	0.17	1.83	1.17	0.05	0.21	1.05	0.02	0.03	1.02
Blog use ^b	0.00	0.00	1.00	0.08	0.78	1.08	-0.05	0.44	0.95
News site use ^b	0.42	7.86*	1.52	0.17	1.71	1.18	0.20	2.40	1.22
Prior knowledge ^b	-0.16	2.02	0.85	-0.03	0.09	0.97	0.00	0.00	1.04
Constant	0.52	0.36	0.55	0.74	0.18	0.73	0.32	0.20	1.39
Nagelkerke's R ²	.05			.02			.02		

Note. *b* = regression coefficient, *W* = Wald statistic (df = 1), *OR* = odds ratio
n = 494; ^aScale from 1 to 6, higher values indicate higher use; ^bScale from 1 to 5, higher values indicate higher use/more knowledge. ^cNormally distributed factors, higher values indicate higher tendency to employ respective selection strategy.

* *p* < .05.

Discussion

Our results point to the importance of search engines for the visibility of online information. Only a few participants differed from the given order when they chose search results. This finding is consistent with those of previous studies (Agichtein et al., 2006; Pan et al., 2007), thus pointing to the external validity of our study. Of the three credibility cues, only reputation affected the probability of the selection of a search result. However, the effect was significantly lower compared to the effect of the ranking of the search results. These results can be interpreted in different ways.

First, the results can be seen as evidence of the secondary nature of credibility (cues) when search engine results are selected. Search engine users might not be aware that the assessment of the quality of online information lies within their responsibility (Metzger et al.,

2003). They seem to trust the given ranking uncritically, while credibility cues, such as the neutrality of the information or social recommendations, are not considered.

Second, users might reflect on the credibility of information sources very well but look for credibility cues other than those varied in this study. The participants in our study were, for the most part, experienced search engine users who had used such services regularly for several years. Thus, the participants might have developed a certain confidence in the given ranking over time (cf. Jansen & Spink, 2006). A high rank could thus be perceived by users as an indicator of high credibility (Hargittai, Fullerton, Menchen-Trevino, & Thomas, 2010; Westerwick, 2013) and may in its function be comparable to an implicit trustworthiness certification (Edelman, 2011), not least since search engines increasingly include automated credibility attributions in their ranking algorithms (Lewandowski, 2012; cf. Dong et al., 2015). However, Pan and colleagues conclude that solely trusting the ranking and neglecting other cues only leads to a successful search “some of the time” (2007, p. 817).

The absence of effects of neutrality and social recommendations warrants a closer inspection. Neutrality affects selection decisions only if the unbiased search result also links to a high-reputation site, while the selection of low reputation results (i.e., blogs and personal websites) is not influenced by whether the search result suggests one-sided or two-sided argumentation. This points to users attributing different credibility criteria to different sources: Although fair, unbiased reporting is expected from news sites, neutrality is not expected from opinionated blogs (T. J. Johnson & Kaye, 2004). However, the comparably small effects of neutrality might also be attributed to the operationalization of said cue via the snippet of the search result: Eye-tracking studies have shown that users skim snippets for certain catchwords rather than read the whole snippet (Lorigo et al., 2008). Furthermore, neutrality is harder to

assess than reputation: there are no “straightforward heuristics [...] for [...] objectivity” (Arazy & Kopak, 2011, p. 96), and criteria used to assess neutrality are expected to greatly vary inter-individually. Finally, the effects of social recommendations on selection decisions have been primarily examined and found in studies on e-commerce decisions (cf. Park et al., 2007). In this context, social recommendations are common to most users, while search engines only now are slowly adopting such cues.

Third, the participants’ self-reports on their use of search engines did not match the participants’ actual behavior. Some participants indicated that they selected search results spontaneously; others said that they strongly depended on the given ranking of results. However, participants also stated that they looked more closely at the search results and thought carefully about their selection. In contrast, our results showed nearly no connections between these self-reports and actual selection behavior in the observed search situations. This might be due to the use of a scale (Beiler, 2005) that was developed when search engines were not as ubiquitous in everyday Internet use behavior. Today, as usage statistics show, most users employ search engines extensively and should thus have developed habitual and routinized usage patterns.

Limitations

Limitations of this study primarily pertain to the external validity of the results. First, the use of a fairly homogenous college student sample may have led to the suppression of some effects. While search engines use is nowadays consistently high across all social strata (for Germany, see Frees & Koch, 2015), a more heterogeneous sample still should lead to more variance in individual characteristics such as prior knowledge, Internet use, and selection strategies. Such a sample may also offer more variance in the perception of credibility cues, as

demographics like age and gender have been shown to affect credibility perceptions of different online sources (T. J. Johnson & Kaye, 2010).

Second, the situational and contextual characteristics of this study remained constant. The instructions for the participants and the research tasks' topics possibly were not suited to create conditions in which the importance of credibility evaluation resembles real usage situations. This could lead to over- and underestimation of the role of credibility judgments on selection. Search result selection is a low-cost scenario, especially for monetary consequences. However, users normally are interested in finding relevant information as quickly as possible. Except out of curiosity, most users will not select a search result that they suspect offers unreliable or implausible information, if at the same time alternatives perceived as more expert and trustworthy are present. However, our participants might not have been highly involved in the research tasks, since personal consequences were nonexistent. Low involvement and habitualization favor the use of simple heuristics, such as selecting the top-ranked search result (Metzger, 2007). Highly involved users might have acted differently and focused more on credibility cues than solely the ranking. However, laboratory settings may lead participants to concentrate more on the assigned tasks than in real-life situations, since the participants are aware they are the subjects of a scientific study. This was indicated by the comparably long dwell times of our participants on the SERP before they selected the first search result. Thus, credibility cues may play even less of a role in real-life usage situations.

Third, this study looked at information research for at least somewhat politically charged issues. Thus, the results may not be fully adaptable for other domains as well. For example, it is quite likely that reputation has a stronger effect in the e-commerce context in that the presence of a major retailer overshadows all other cues, while users researching health information online

may place a stronger emphasis on both reputable and neutral search results. In future studies, we will thus further investigate the effects of credibility cues on the selection of search results in other domains as well.

Conclusion

In total, our results suggest that credibility has only secondary influence on the selection of online content or that general trust in search engines superimposes single credibility cues. The use of a simple primacy heuristic seems to guide selection decisions more than the evaluation of credibility cues. Although this is certainly the case in many real-life situations as well (in a log-file analysis of Bing users, only one fourth of all users seemed to evaluate the SERP thoroughly during non-navigational tasks; Buscher, White, Dumais, & Huang, 2012), several circumstances—high issue involvement, purchase intentions and thus monetary ramifications, or passing the searched information on to peers and thus social consequences—might lead to users being motivated to evaluate information credibility more thoroughly even *before* the selection. Future studies should thus examine the influence of situational and contextual characteristics, especially of different modes of information processing (cf. Metzger, 2007), on the perception and effects of credibility cues. Furthermore, this study focused on the role of credibility cues for predictive judgments. However, users make “predictive and evaluative judgments continuously until they complete the searches” (Rieh, 2002, p. 150). Evaluative credibility judgments performed on the target sites may confirm or contradict those initial predictive judgments, and thus influence further predictive judgments when users return to the SERP after selecting a search result. In our study, the explanative quality of the selection model vastly decreased for the following selection decisions. The interplay between predictive and evaluative judgments thus might offer valuable insights into selection decisions beyond the first click.

Footnotes

¹The top-ranked results are usually compromised of popular websites—especially encyclopedias (e.g., *Wikipedia*) and knowledge exchanges (e.g., *answers.com*) that provide answers for a large proportion of search queries, but also popular news sites. However, an analysis of hosts linked in SERPs for a large number of popular and rare search queries across four different search engines found that more than half of all hosts appeared only once in all investigated results pages (Höchstötter & Lewandowski, 2009). Especially less common search queries (e.g., multiple word queries used to address more open-ended search tasks) should lead to more unpopular (and thus most likely unknown to the user) content providers ranked on top.

²Reliable numbers for the exact percentages are rarely made public. The quoted percentages are based on analyses by the web analytics company SimilarWeb (www.similarweb.com).

³Expanding on the work of the Yale group, several other possible dimensions of credibility have been proposed, such as dynamism, sympathy, extraversion, etc. (cf. Kiouisis, 2001; Metzger et al., 2003). Recent studies on online credibility have also included technical and aesthetical dimensions into their conceptualizations (Chung, Nam, & Stefanone, 2012). However, as Rieh and Danielson argue, “it is often unclear whether the factors identified are mere predictors of source credibility or representative of an underlying dimension of the construct” (2007, p. 315). Thus, we focus on the often tested and verified dimensions of expertise and trustworthiness (for a review, see Choi & Stvilia, 2015).

⁴The instructions were as follows: “The search result page of the search engine ‘DuckDuckGo’ and the search query [‘consequences rent control’ / ‘series streaming download legal status’] is included on the next page of the questionnaire. Please inform yourself about the

topic [‘rent control mechanism’ / ‘legal status of streaming and download of TV series’] using this search results page. You will be asked to complete a short quiz about the topic afterwards. You may open as many or as little search results as you want. In total, you have got five minutes to do so. As we are interested in this search query in particular, we request you not to change the query.”

⁵Several control mechanisms prevented users from recognizing the manipulated nature of the stimulus or from “leaving” the setup. For example, URLs were not visible in the address bar or in the status bar when the user’s mouse hovered over the links. Opening links in another browser tab was prevented by a script. Changing the search query resulted in an error page that enabled the participants to return to the original results page. Users could, however, follow outgoing links on the manipulated target sites (which would then take them to the “real” site). A button above the iFrame then took them back to the manipulated SERP.

⁶The majority (63.2%) of the participants disclosed they used illegal streaming services themselves. Rental fees in the participants’ place of residence are ranked among the highest in Germany.

⁷Promax rotation; explained variance (cumulative) = 52.5%, KMO = .66; Bartlett’s test of sphericity: $\chi^2(36) = 235.54, p < .001$.

References

- Agichtein, E., Brill, E., Dumais, S., & Ragno, R. (2006). Learning user interaction models for predicting web search result preferences. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 3–10). New York, NY: ACM. <http://doi.org/10.1145/1148170.1148175>
- Arazy, O., & Kopak, R. (2011). On the measurability of information quality. *Journal of the American Society for Information Science and Technology*, 62, 89–99. <http://doi.org/10.1002/asi.21447>
- Balatsoukas, P., & Ruthven, I. (2012). An eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology*, 63, 1728–1746. <http://doi.org/10.1002/asi.22707>
- Beiler, M. (2005). Selektionsverhalten in den Ergebnislisten von Suchmaschinen. Modellentwicklung und empirische Überprüfung am Beispiel von Google [Selection behavior on search engine results pages. Model construction and empirical evaluation with the example of Google]. In M. Machill & N. Schneider (Eds.), *Suchmaschinen: Neue Herausforderungen für die Medienpolitik* [Search engines. New challenges for media politics] (pp. 165–189). Berlin: Vistas.
- Broder, A. (2002). A taxonomy of web search. In *SIGIR Forum* (Vol. 36(2), pp. 3–10). <http://doi.org/10.1145/792550.792552>
- Buscher, G., White, R. W., Dumais, S., & Huang, J. (2012). Large-scale analysis of individual and task differences in search result page examination strategies. In *Proceedings of the*

- 5th ACM international conference on Web search and data mining* (pp. 373–382). New York, NY: ACM. <http://doi.org/10.1145/2124295.2124341>
- Choi, W., & Stvilia, B. (2015). Web credibility assessment: Conceptualization, operationalization, variability, and models. *Journal of the Association for Information Science and Technology*, *66*, 2399–2414. <http://doi.org/10.1002/asi.23543>
- Chung, C. J., Nam, Y., & Stefanone, M. A. (2012). Exploring online news credibility: The relative influence of traditional and technological factors. *Journal of Computer-Mediated Communication*, *17*, 171–186. <http://doi.org/10.1111/j.1083-6101.2011.01565.x>
- DigitalRelevance. (2013, January). A tale of two studies: Establishing Google & Bing click-through rates. Retrieved from http://connect.relevance.com/a-tale-of-two-studies-establishing-google-bing-click_through-rates-digitalrelevance
- Dong, X. L., Gabrilovich, E., Murphy, K., Dang, V., Horn, W., Lugaresi, C., ... Zhang, W. (2015). Knowledge-based trust: Estimating the trustworthiness of web sources. *arXiv:1502.03519*. Retrieved from <http://arxiv.org/abs/1502.03519>
- Dutta-Bergman, M. J. (2004). The impact of completeness and web use motivation on the credibility of e-health information. *Journal of Communication*, *54*, 253–269. <http://doi.org/10.1111/j.1460-2466.2004.tb02627.x>
- Edelman, B. (2011). Adverse selection in online “trust” certifications and search results. *Electronic Commerce Research and Applications*, *10*, 17–25. <http://doi.org/10.1016/j.elerap.2010.06.001>
- 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. <http://doi.org/10.1177/1461444807075015>

- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., ... Treinen, M. (2001). What makes Web sites credible? A report on a large quantitative study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 61–68). New York, NY: ACM. <http://doi.org/10.1145/365024.365037>
- Frees, B., & Koch, W. (2015). Internetnutzung: Frequenz und Vielfalt nehmen in allen Altersgruppen zu - Ergebnisse der ARD/ZDF-Onlinestudie 2015 [Internet use: Frequency and diversity increase in all age groups - results of the ARD/ZDF online study 2015]. *Media Perspektiven, 2015*, 366–377.
- Gaziano, C., & McGrath, K. (1986). Measuring the concept of credibility. *Journalism & Mass Communication Quarterly, 63*, 451–462. <http://doi.org/10.1177/107769908606300301>
- Greer, J. D. (2003). Evaluating the credibility of online information: A test of source and advertising influence. *Mass Communication and Society, 6*, 11–28. http://doi.org/10.1207/S15327825MCS0601_3
- Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication, 4*, 468–494.
- Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences, 179*, 1796–1812. <http://doi.org/10.1016/j.ins.2009.01.028>
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. Westport, CT: Greenwood Press.
- Hu, Y., & Sundar, S. S. (2009). Effects of online health sources on credibility and behavioral intentions. *Communication Research, 37*, 105–132. <http://doi.org/10.1177/0093650209351512>

- Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, 42, 248–263. <http://doi.org/10.1016/j.ipm.2004.10.007>
- Johnson, K. A., & Wiedenbeck, S. (2009). Enhancing perceived credibility of citizen journalism web sites. *Journalism & Mass Communication Quarterly*, 86, 332–348. <http://doi.org/10.1177/107769900908600205>
- Johnson, T. J., & Kaye, B. K. (2004). Wag the blog: How reliance on traditional media and the Internet influence credibility perceptions of weblogs among blog users. *Journalism & Mass Communication Quarterly*, 81, 622–642. <http://doi.org/10.1177/107769900408100310>
- Johnson, T. J., & Kaye, B. K. (2010). Still cruising and believing? An analysis of online credibility across three presidential campaigns. *American Behavioral Scientist*, 54, 57–77. <http://doi.org/10.1177/0002764210376311>
- Kiousis, S. (2001). Public trust or mistrust? Perceptions of media credibility in the information age. *Mass Communication and Society*, 4, 381–403. http://doi.org/10.1207/S15327825MCS0404_4
- Leino, J., Rähkä, K.-J., & Finnberg, S. (2011). All the news that's fit to read: Finding and recommending news online. In P. Campos, N. Graham, J. Jorge, P. Palanque, N. Nunes, & M. Winckler (Eds.), *Human-Computer Interaction - INTERACT 2011* (Vol. 3, pp. 169–186). Heidelberg: Springer.
- Lewandowski, D. (2012). Credibility in web search engines. In M. Folk & S. Apostel (Eds.), *Online credibility and digital ethos: Evaluating computer-mediated communication* (pp. 131–146). Hershey, PA: IGI Global.

Lewandowski, D. (2013). Challenges for search engine retrieval effectiveness evaluations:

Universal search, user intents, and results presentation. In G. Pasi, G. Bordogna, & L. C. Jain (Eds.), *Quality Issues in the Management of Web Information* (Vol. 50, pp. 179–196). Berlin, Heidelberg, Germany: Springer.

Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., ... Pan, B. (2008).

Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59, 1041–1052.

<http://doi.org/10.1002/asi.20794>

Machill, M., Beiler, M., & Zenker, M. (2008). Search-engine research: A European-American

overview and systematization of an interdisciplinary and international research field.

Media, Culture & Society, 30, 591–608. <http://doi.org/10.1177/0163443708094010>

Machill, M., Neuberger, C., Schweiger, W., & Wirth, W. (2004). Navigating the Internet: A study

of German-language search engines. *European Journal of Communication*, 19, 321–347.

<http://doi.org/10.1177/0267323104045258>

McCroskey, J., & Teven, J. (1999). Goodwill: A reexamination of the construct and its

measurement. *Communication Monographs*, 66, 90–103.

<http://doi.org/10.1080/03637759909376464>

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. <http://doi.org/10.1002/asi.20672>

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. M. (2003). Credibility for

the 21st century: Integrating perspectives on source, message, and media credibility in the

- contemporary media environment. In P. J. Kalbfleisch (Ed.), *Communication Yearbook* (Vol. 27, pp. 293–335). New York, NY: Routledge.
- Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60, 413–439.
<http://doi.org/10.1111/j.1460-2466.2010.01488.x>
- Meyer, H. K., Marchionni, D., & Thorson, E. (2010). The journalist behind the news: Credibility of straight, collaborative, opinionated, and blogged “news.” *American Behavioral Scientist*, 54, 100–119. <http://doi.org/10.1177/0002764210376313>
- Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., ... Tanaka, K. (2007). Trustworthiness analysis of web search results. In *Research and advanced technology for digital libraries* (Vol. 4675, pp. 38–49). Berlin: Springer.
- O’Keefe, D. J. (2008). Elaboration likelihood model. In W. Donsbach (Ed.), *The international encyclopedia of communication* (pp. 1475–1480). Malden, MA: Blackwell Publishing.
- 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. <http://doi.org/10.1111/j.1083-6101.2007.00351.x>
- Park, D.-H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, 11, 125–148. <http://doi.org/10.2753/JEC1086-4415110405>
- Pirolli, P. (2005). Rational analyses of information foraging on the web. *Cognitive Science*, 29, 343–373. http://doi.org/10.1207/s15516709cog0000_20
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643.
<http://doi.org/10.1037/0033-295X.106.4.643#>

- Purcell, K., Brenner, J., & Rainie, L. (2012). *Search engine use 2012*. Pew Research Center's Internet & American Life Project. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf
- 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.
<http://doi.org/10.1002/asi.10017>
- Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. *Annual Review of Information Science and Technology*, 41, 307–364.
<http://doi.org/10.1002/aris.2007.1440410114>
- Schmierbach, M., & Oeldorf-Hirsch, A. (2012). A little bird told me, so I didn't believe it: Twitter, credibility, and issue perceptions. *Communication Quarterly*, 60, 317–337.
<http://doi.org/10.1080/01463373.2012.688723>
- Schwarz, J., & Morris, M. (2011). Augmenting web pages and search results to support credibility assessment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1245–1254). New York, NY: ACM.
<http://doi.org/10.1145/1978942.1979127>
- Self, C. C. (2009). Credibility. In D. W. Stacks & M. B. Salwen (Eds.), *An integrated approach to communication theory and research* (2nd ed., pp. 435–456). New York, NY: Routledge.
- Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, 58, 696–707. <http://doi.org/10.1287/mnsc.1110.1458>

- Sundar, S. S., Knobloch-Westerwick, S., & Hastall, M. R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology*, 58, 366–378. <http://doi.org/10.1002/asi.20511>
- Terbeck, V. H. (2012). Der Einfluss sozialer Empfehlungen auf das Selektionsverhalten in Suchmaschinen: eine Eyetracking-Studie der Google-Ergebnisseite [The influence of social recommendations on the selection behavior on search engines: An eye-tracking study of the Google results page]. *Info* 7, 27, 45–52.
- Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Communications of the ACM*, 42, 39–44. <http://doi.org/10.1145/301353.301402>
- Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology*, 53(2), 134–144. <http://doi.org/10.1002/asi.10016>
- Westerwick, A. (2013). Effects of sponsorship, web site design, and google ranking on the credibility of online information. *Journal of Computer-Mediated Communication*, 18, 80–97. <http://doi.org/10.1111/jcc4.12006>
- White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining* (pp. 132–141). New York, NY, USA: ACM. <http://doi.org/10.1145/1498759.1498819>
- Winter, S., & Krämer, N. C. (2012). Selecting science information in Web 2.0: How source cues, message sidedness, and need for cognition influence users' exposure to blog posts. *Journal of Computer-Mediated Communication*, 18, 80–96. <http://doi.org/10.1111/j.1083-6101.2012.01596.x>

- Wirth, W., Sommer, K., von Pape, T., & Karnowski, V. (2015). Success in online searches: Differences between evaluation and finding tasks. *Journal of the Association for Information Science and Technology*, Advance online publication. <http://doi.org/10.1002/asi.23389>
- Xu, Q. (2013). Social recommendation, source credibility, and recency: Effects of news cues in a social bookmarking website. *Journalism & Mass Communication Quarterly*, 90, 757–775. <http://doi.org/10.1177/1077699013503158>
- Yamamoto, Y., & Tanaka, K. (2011). Enhancing credibility judgment of web search results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1235–1244). New York, NY: ACM. <http://doi.org/10.1145/1978942.1979126>