Investigating the Effects of Google's Search Engine Result Page in Evaluating the Credibility of Online News Sources

Emma Lurie
Computer Science Department
Wellesley College
emma.lurie@wellesley.edu

Eni Mustafaraj
Computer Science Department
Wellesley College
eni.mustafaraj@wellesley.edu

ABSTRACT

Recent research has suggested that young users are not particularly skilled in assessing the credibility of online content. A follow-up study comparing students to fact checkers noticed that students spend too much time on the page itself, while fact checkers performed "lateral reading", searching other sources. We have taken this line of research one step further and designed a study in which participants were instructed to do lateral reading for credibility assessment by inspecting Google's search engine result page (SERP) of unfamiliar news sources. In this paper, we summarize findings from interviews with 30 participants. A component of the SERP noticed regularly by the participants is the so-called Knowledge Panel, which provides contextual information about the news source being searched. While this is expected, there are other parts of the SERP that participants use to assess the credibility of the source, for example, the freshness of top stories, the panel of recent tweets, or a verified Twitter account. Given the importance attached to the presence of the Knowledge Panel, we discuss how variability in its content affected participants' opinions. Additionally, we perform data collection of the SERP page for a large number of online news sources and compare them. Our results indicate that there are widespread inconsistencies in the coverage and quality of information included in Knowledge Panels.

CCS CONCEPTS

• **Information systems** → **Web search engines**; *Search interfaces*; • **Human-centered computing** → *Empirical studies in HCI*;

KEYWORDS

Google; search; news sources; credibility; user studies

ACM Reference Format:
Emma Lurie and Eni Mustafaraj. 2018. Investigating the Effects of Google's Search Engine Result Page in Evaluating the Credibility of Online News Sources. In *Proceedings of 10th ACM Conference on Web Science (WebSci '18).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3201064.3201095

1 INTRODUCTION

Researchers have been sounding the alarm for years that being born in the Internet age doesn't make one better at assessing content encountered online [4]. "Digital natives" are mostly a myth, as task-based studies have demonstrated [18, 19]. The 2016 large-scale study by the Stanford History Education Group (SHEG), which assessed 7804 middle school to college-aged participants on their ability to judge online information, discovered troubling results [9]. Students couldn't distinguish real news articles from promoted content, information by biased think tanks from peer reviewed research publications, and were deceived by domain endings such as `.org`. The authors of the SHEG study followed up with another study [32], in which they compared how fact-checkers and students approach the task of evaluating the credibility of an unknown online source. They observed that students spent most of the time on the website of the source, accessing different pages and trying to reason about its credibility. Meanwhile, fact checkers engaged in what the researchers call "lateral reading," leaving the site to google the organization, its associated members, and to gather information from other sources.

How effective is "lateral reading" for users who are not trained in fact-checking? Given the public calls for teaching media literacy [5], it's important to investigate the value of different proposed

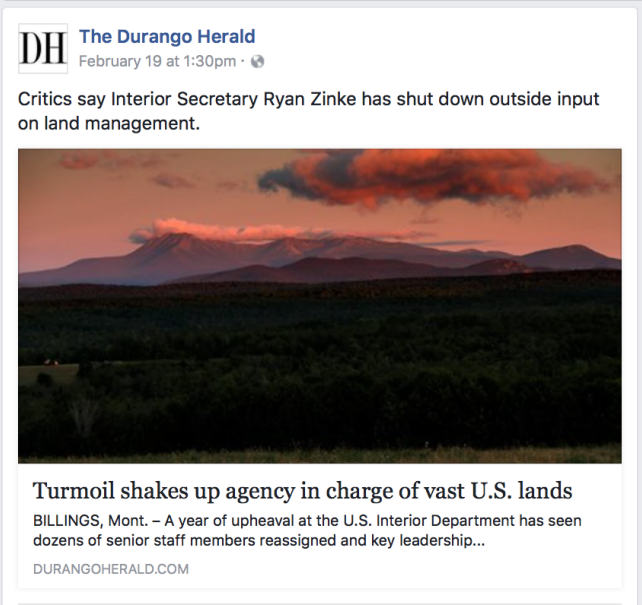

Figure 1: Screenshot of a Facebook news story card, the common way to display news in a user's News Feed. Such a card contains tidbits of the content as well as the name and URL of the source.

WebSci '18, May 27–30, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5563-6/18/05.
https://doi.org/10.1145/3201064.3201095

approaches. To answer the question of the effectiveness of googling as a literacy skill, we conducted a user study with 30 participants (age 18-22), in which they performed "lateral reading" to assess the credibility of three U.S.-based online news sources that were **unfamiliar** to them: *The Durango Herald*, *The Tennessean*, and *The Christian Times*. The scenario we imagined and shared with the participants is the following: most users encounter news stories in their social media feeds such as Facebook and Twitter in the form of "story cards", which contain a title, some text from the article, an image, the name of the website and the URL, as shown in Figure 1. Many of the fake news sites that were successful in spreading misinformation during the 2016 U.S. Election took advantage of Facebook's news cards feature, to make their stories look legitimate [14]. Thus, encountering cards with news headlines from unknown sources is a situation in which googling for the source is a logical step and is what literacy experts suggest [6]. Accordingly, we asked our study participants to google the three above-mentioned news sources and recorded how they used the Google search engine page result (SERP) to reason about the credibility of each website.

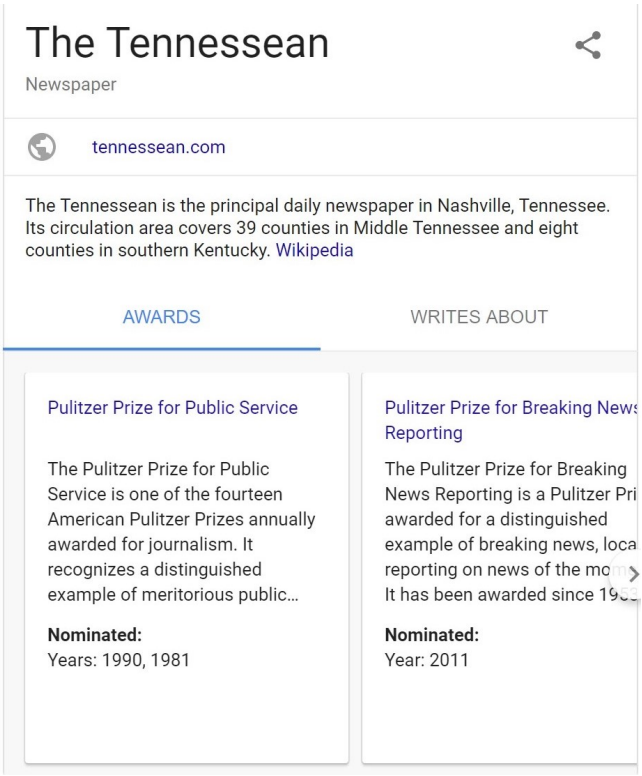

Figure 2: Partial screenshot of Google's SERP for The Tennessean. This box is known as a Knowledge Panel. In addition to a snippet retrieved from the Wikipedia page, it also contains two tabs titled "Awards" and "Writes About". Not all Knowledge Panels of newspapers contain these two tabs. This feature was announced on Nov 7, 2017 [33].

Our major findings from this study are the following:

(1) Participants find Knowledge Panels valuable in assessing credibility, especially when they contain the "Awards" tab. (Please refer to the Knowledge Panel displayed in Figure 2. We discuss Knowledge Panels later in the paper).
(2) Knowledge Panels are insufficient to make definitive credibility assessments, and sometimes generate more confusion.
(3) "Top Stories" and social media feeds provide meaningful signals for participants to incorporate into their assessment.

Given the importance that participants attributed to Knowledge Panels, we undertook a quantitative study of three different datasets of online news sources to investigate what information the SERP contains for each news source. We performed the data collection twice and compared the results. A major finding is that while in the January 2018 dataset the Knowledge Panel for some online sources contained a tab on "reviewed claims", presenting information from fact-checkers (see Figure 3), in February 2018, Knowledge Panels no longer contained "reviewed claims". This disappearance of useful information is worthy of notice and discussion. There is speculation in the media that Google bowed to political pressure[1] and removed the information from the Knowledge Panel.

To the best of our knowledge, this is the first research paper that studies how users perform on lateral reading tasks through Google, as well as the first paper that focuses on the role of Knowledge Panels for evaluating the credibility of an online news source.

The rest of the paper is organized as follows: we provide an overview of online credibility research in the past years, focusing on how young people evaluate sources. We describe in detail elements of the Google SERP for a news source, to show what information is available to users when they google for a source. Our user study is then explained in detail, with a discussion of both its findings and limitations. We report the results of the SERP data collection for three datasets of online news sources on two different dates and highlight the differences. Finally, we conclude with a discussion of Google's role in news literacy efforts, and the role that the research community can play in monitoring the quality of SERP information.

2 RELATED RESEARCH

Credibility is not an easy concept to pinpoint and it is studied in many research communities. The most recent literature survey on credibility in information systems [16], which extends previous work [13], identifies trustworthiness, expertise, quality, and reliability as its ingredients. In this section, we confine our review on credibility on the web, as well as on the inherent trust that users put on search engines like Google.

2.1 Credibility on the Web

Early research in communication studies defined credibility as the believability of a piece of information based on its content and source [22], and treated it as dependent on perceived expertise and trustworthiness [11]. Many studies emphasized that evaluating credibility on the Web is most successful when examining the following five criteria: accuracy, authority, objectivity, currency, and coverage [26]. These criteria are often part of checklists that are given to students by research librarians, who always advise

[1] https://www.poynter.org/news/blame-bugs-not-partisanship-google-wrongly-appending-fact-check-daily-caller

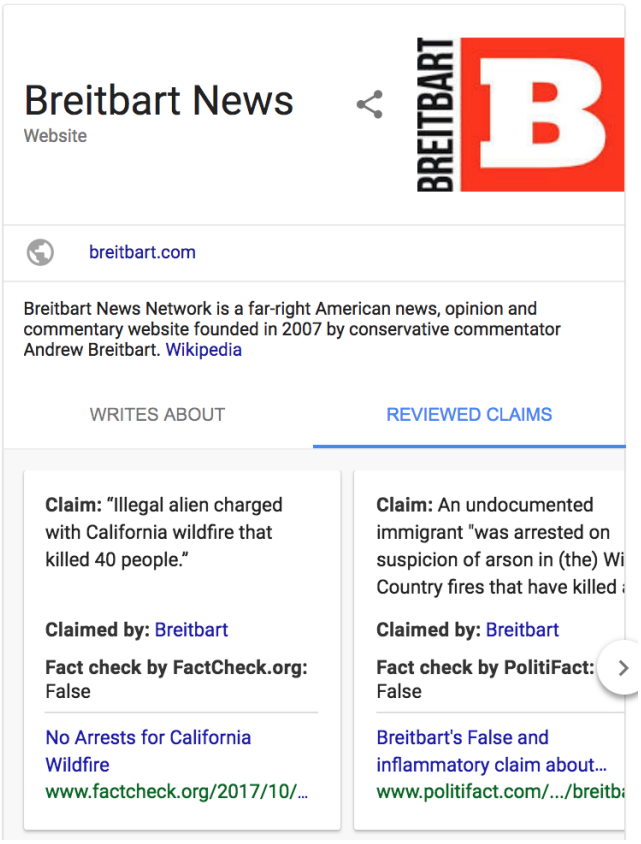

Figure 3: January 2018 screenshot for the Knowledge Panel for Breitbart.com, showing the "reviewed claims" section that contains fact-checked claims. As of February 2018, the "reviewed claims" section is not part of the panel anymore.

skepticism in dealing with sources. Current media literacy efforts have evolved from the "checklist" to "lateral reading", i.e. googling claims or stories to verify information [7, 32].

Especially in the 2000s, the public perception of the Internet was that of a source as credible as television, radio, and magazines, and Internet users reported rarely verifying what they read on the Web [11, 31]. Users are even less likely to perform time intensive source verification techniques, which are often the more deterministic measures of credibility [26]. Given this known user aversion to verification work, communication researchers such as [26] have proposed several tools such as credibility seals [12] or rating systems [28] to assist users in evaluating credibility.

When over 2,500 users were asked to describe the features they actually used to determine website credibility, almost half mentioned "site presentation" as a key factor [12]. In general, users seem to have a *positive evaluation bias*, thinking that sites are more credible than they actually are [1]. However, with expert suggestions on the credibility of Web sources at their disposal, users are able to make better decisions [1]. This idea of added context improving decision-making motivated our interest in investigating Knowledge Panels closely.

2.2 Search Engine Result Page

The Search Engine Result Page (SERP) has been studied by both academics and Internet marketers. In 2005, Internet marketing firm Enquiro coined the term "Google Golden Triangle" after noticing in an eye-tracking study that users' eyes naturally focus on the upper left of the SERP and travel in a small area down and around the SERP forming a triangle that only extends to a few search results [21]. However, as more heterogeneous SERPs began to emerge with the inclusion of images, videos, and the Knowledge Graph content on SERP, a 2014 study found that the "Golden Triangle" no longer existed, users now examine the SERP more vertically due to (1) mobile device scrolling habits and (2) top organic results are no longer always found in the upper left-hand corner due to the added elements on the SERP [8, 25].

Another line of research has illustrated that college students trust Google's ranking of SERP results [20] and are willing to click on the first couple of results, even when more relevant links were ranked towards the bottom of the SERP [27]. Such studies have been repeated with younger children, indicating that age is a factor in SERP behavior [17]. As more rich media snippets and Knowledge Graph information is found on the SERP, [24] explored the relationship between fewer results on the SERP and an increased attention to the higher ranked organic results. [30] explores the connection between high-quality Google SERPs and Wikipedia content.

Given the recency of Knowledge Panels, there is no current research that has assessed its effects on user behavior, something we are trying to partially address in this study.

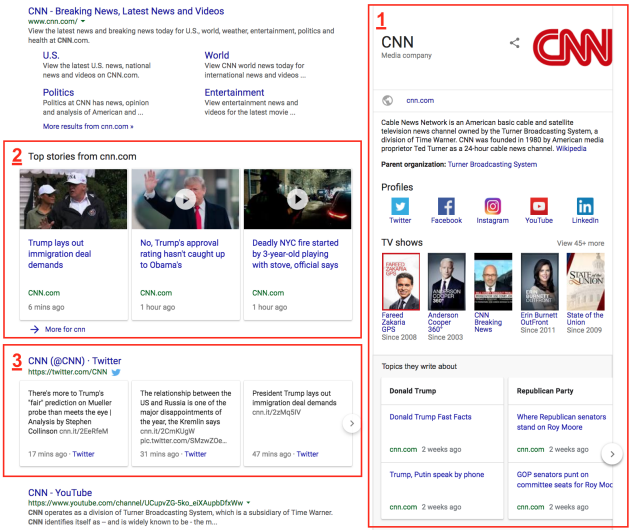

Figure 4: A screenshot of the Google search page results for CNN. It contains (1) the Knowledge Panel, (2) a section on top stories; and (3) a section of recent tweets. The panel "Topics they write about" contains stories from "2 weeks ago", indicating that topics are not "discovered" on the fly.

3 SERP: AN ANATOMY

Searching information on Google reveals the complex and increasingly intertwined behavior of three different actors:

(1) **Google's algorithms**, which, depending on the query, can display additional information on the SERP, e.g., Knowledge Panels which can be populated in different ways, top stories, tweets, direct answers, etc. (see Figure 4 for an example of some of these components);
(2) **users**, whose informational needs are revealed in the panel of "searches related to query phrase" (see Figure 5) commonly located at the bottom of a search page; or in a panel titled "People also ask" (see Figure 6), which occasionally appears either at the top or middle of a search results page.
(3) **online publishers**, whose links to relevant content appear in the search results. Such publishers include Wikipedia, content from which is displayed in the Knowledge Panels, as well as other publishers who are savvy in creating lists, tables, or answering common questions, often by using known search engine optimization techniques.

While for a long time the Google search page operated as a transactional "middleman": receiving a request (a query phrase) and displaying a list of organic results ranked from 1 to 10 (with ads on the side), things changed in 2012 with the introduction of the Knowledge Graph [29]. The Knowledge Graph pulls factual information about things in the real world from databases such as Wikipedia, Freebase, CIA World Factbook, etc., and shows them in separate panels in different parts of the SERP. Often, a user doesn't need to click on any of the links of a SERP, because these featured panels contain the answer to their query. Google regards the purpose of the SERP as both answering a query and a way of discovering other related things, therefore, the number of featured panels (also known as snippets) has increased over time [15]. We will discuss in the following two of the most prominent and sometimes controversial panels: the Knowledge Panel and the direct answer panel.

3.1 Knowledge Panels

Examples of the Knowledge Panel are shown in three figures. Figure 2 displays the panel for *The Tennessean*, Figure 3 shows the panel for *Breitbart News*, and the screenshot in Figure 4 contains the panel for *CNN*. These panels are different, with the CNN panel providing more information, including pictures of its TV anchors, and links to its social web presence. All three panels contain a section on "Writes About" which is populated with links of stories grouped in topics, though our figures only reveal the topics for CNN, for example, "Donald Trump", "Republican Party", etc. For several newspapers, such as *NY Times*, *Washington Post*, and other, the Knowledge Panel contains a tab dedicated to "Awards", which we also see in Figure 2 for *The Tennessean*. The existence of the "Awards" tab can be considered as a way to signal their authority in the field of journalism. For a two-month period, November 2017-January 2018, the Knowledge Panel also contained a tab on "Reviewed Claims." This tab included claims published by the source and reviewed by third-party fact-checkers such as Snopes or Politifact with a verdict of being found "True" or "False" (or something in between). An example can be seen in Figure 3.

Figure 5: A screenshot from a Google search page. At the bottom of most search result pages is a section titled "searches related to query phrase", in this particular case, the query is "best rated newspapers". This portion provides insights into users' informational needs.

Figure 6: A screenshot from a Google search page. Some result pages contain a panel titled "People also ask" that shows full questions asked by users in relation to the query. One can click on the down arrow on the right side and read the answer directly in the panel, without visiting the page.

3.2 Direct Answer

One of the featured snippets that has gained prominence in the past years is the so-called "direct answer", or "position zero" (because it is shown above the ranked results), see Figure 7 for an example. An annual study from the market research company Stone Temple, which collected the content of search page results for 1.4 million queries, found out that currently, 30% of queries contain a "direct answer", and the rate of direct answers is continuing to rise [10]. However, the practice of answering queries in this way has drawn criticism in the media [23], because it is prone to manipulation. One of the examples included in [23] displayed a featured snippet from a conspiracy website that claimed that President Obama is planning a communist coup d'etat before the 2016 election. The snippet was shown as direct answer to the query: "is Obama planning a coup."

In addition to potentially propagating misinformation, direct answer results can also be exploited by content creators who use search engine optimization techniques to bias Google search algorithms into considering their page as highly relevant. See Figure 7, for an example of a featured snippet that fails at providing a substantive answer to the question of newspaper quality, providing instead an answer about newspaper circulation. Although direct answers are not shown when looking up entities such as news websites, as an important part of SERP they need attention from researchers, something we plan to pursue in our future research.

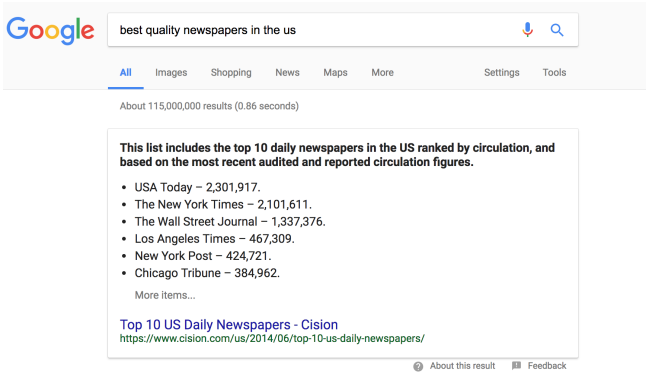

Figure 7: An example of a "direct answer", a featured snippet that is shown above the ranked search results. Notice that the content is from 2014 (the date in the URL of the article), although the query was performed on Jan 3, 2018.

4 USER STUDY: EVALUATING CREDIBILITY THROUGH SERP

Given that users might often encounter a news story from an unfamiliar website in their news feed, how would googling for the source help them believe or reject the story? More specifically, how would the composition and content displayed on the Google's SERP affect their decision making process for assessing the credibility of the source?

We designed a user study with in-person interviews and a think-aloud protocol to observe how users behave when asked to google unfamiliar news sources. Given that the one of the goals of lateral reading is to receive contextual information, we chose sources are not familiar to most people, but have a Knowledge Panel on their SERP. They are:

(1) **The Durango Herald** - a local newspaper in southwestern Colorado. The Knowledge Panel for this SERP has a detailed Wikipedia snippet that includes the date the newspaper was established as well as the region the paper primarily serves. *The Durango Herald's* Knowledge Panel also includes a "Topics they write about" section. A "Top Stories" and Twitter feed are also featured on The *Durango Herald's* SERP.

(2) **The Tennessean** - the principal daily newspaper in the Nashville, Tennessee region. The Knowledge Panel contains an "Awards" tab featuring three Pulitzer Prize nominations, as well as a "Writes About" tab. There is also a Wikipedia snippet, although it is noticeably less descriptive than the *Durango Herald's* Wikipedia excerpt. While there is a "Top Stories" on the SERP, there is no featured Twitter feed, although the newspaper has its Twitter account.

(3) **The Christian Times** - an online newspaper that is part of the Christian Media Corporation (CMC Group) that owns a number of Christian-issues focused online and print papers. The Christian Times does not have a Wikipedia page, however, Google still shows a Knowledge Panel for it from an unknown source, see Figure 8. It is worth noting here that Google is wrong about the parent company that owns the Christian Times. Further complicating the matters is the conflation of search results with the ones belonging to the former fake news site "Christian Times Newspaper." As a result, Snopes, Media Bias/Fact Check, and a CBS News' "Fake News Sites to Watch Out For" article are all featured on the SERP. While The Christian Times has a self-professed religious bias, it is not the fake news site that SERP indicates. This is a clear case in which a fake news site used a very similar name to a real source's website to receive legitimacy, and now that the fake news website is defunct, its footprints on the Web are inherited by the legitimate website, damaging its reputation.

Screenshots of all SERPs mentioned in the paper can be found online[2].

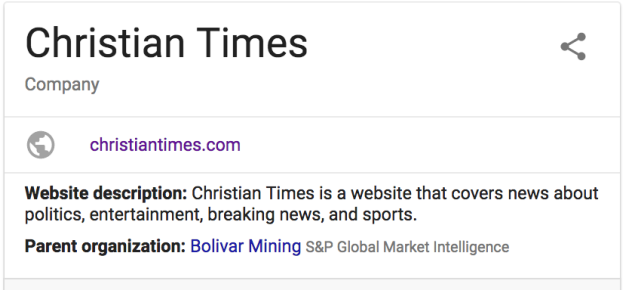

Figure 8: Excerpt from the Knowledge Panel for *The Christian Times*. The website description is generated automatically, and the Parent Organization information is wrong. Google doesn't provide the source for these pieces of information. Screenshot was captured on February 24, 2018. As of April 11, 2018, there is no longer a website description.

4.1 User Study Methodology

This study obtained IRB approval from our institution. Participation was voluntary, unpaid, and subjects signed a consent form before starting the interview. The following procedure was applied:

(1) Participants were asked to evaluate their ability to assess the trustworthiness of sources, describe their process for evaluating unknown online stories, and explain the factors that they believe to be important when assessing online content.

(2) Participants were asked to query Google on a desktop for the three online sources in a randomized order. After only viewing the SERP, they were asked to evaluate the credibility of each source. Participants were then asked what information they found compelling when determining credibility, or if they were unsure about the credibility of the source, what information was needed to make an accurate determination.

(3) Participants were then invited to visit the homepage of the site and any other pages on the SERP they wished to explore. They were then asked how their evaluation of the site's credibility changed.

[2] http://cs.wellesley.edu/~credlab/websci18/

(4) Finally, participants were asked which parts of the SERP were most useful when determining credibility and what they thought would improve the SERP.

4.2 Existing Techniques for Evaluating Credibility

At the beginning of the survey, participants were asked (1) how to test the reliability of an unknown website and (2) the most important factors to think about when determining the trustworthiness of a website. The most common response (53%) described a "lateral reading" approach that verifies claims based on similarly reported stories (typically through a Google search). Some participants (30%) also expressed a desire to only visit known, reputable sources to avoid all ambiguity. The most frequently mentioned techniques to determine the trustworthiness of an unknown news site were: the authors reputation and qualifications (30%), the domain ending of the site (.com, .org, .edu, etc.) (27%), and presentation of the page (23%) (site layout and spelling errors).

4.3 Credibility Signals Identified by Participants

By coding participants' responses using the qualitative data analysis software Atlas.ti, trends in participants responses emerged. In this section, we have highlighted the four most compelling: Knowledge Panels, organic results, the Top Stories panel, and social media-related results.

4.3.1 Knowledge Panel. When presented with the SERP of each of the sources, participants initially glanced at the Knowledge Panel. However, the information on the Knowledge Panel was often insufficient to make a credibility determination. According to participants, the most persuasive element when assessing the credibility of a source was the "Awards" tab (53%). The Wikipedia snippet was the most often-referenced piece of the Knowledge Panel (66%). Participants discussed the information in the Wikipedia description, especially the establishment date of *the Durango Herald* (33%) and the regional location of the local papers (23%). The "Awards" panel only appeared in one of the three Knowledge Panels, and was talked about in 60% of interviews, but the "Topics they write about" section appears on all three Knowledge Panels, but only a fraction of participants (20%) referenced it. Even fewer (10%) found it useful.

Beneath the title of a source on the Knowledge Panel, there is a label describing the entity. For the *Durango Herald* and *the Tennessean* the label is "newspaper," but for *the Christian Times* the label is "company" (see Figure 8). Some participants (20%) commented on this discrepancy as a reason why *the Christian Times* is not credible.

4.3.2 Organic Results. Even while acknowledging the value of the "Awards" section of the Knowledge Panel of *the Tennessean*, 93% of participants still assessed other parts of the SERP before making a credibility evaluation. All participants surveyed at least the first two results on the SERP for at least one of the three sites. On the *Durango Herald* and *the Tennessean's* SERP, the first result was the website itself. For the *Christian Times*, the website was the second result[3], decreasing its credibility for some participants (20%).

[3]This ranking was true for the time period of the study. It might change in the future.

All participants identified the *Christian Times* as not credible. On the SERP, the first result, a Snopes archive link, did not contain the term "fake news," but the first sentence of the snippet was "Maryland Lawmaker's Aide Fired for Creating Fake News Site." Nine participants discussed "fake news" being mentioned in the first result as a factor in believing the Christian Times as not credible. While this is 30% of our sample, we expected this number to be higher given that typically users do not look further than the first three results [21]. In our experiment, participants kept scrolling, identifying links further on the SERP that explicitly called into question the credibility of the *Christian Times*, such as "Christiantimes.com - Fake news sites to watch out for," "How Christian Times Traded Its Good Name-Twice," and "Christian Times Newspaper - Media Bias/Fact Check." 13 participants did not reference the Snopes result specifically, but either mentioned generally that results on the SERP indicate *the Christian Times* as unreliable or referenced other specific fact-checking links lower on the SERP.

Additionally, some participants were impressed by the sitelinks (30%), the links under a Google result that enhance page navigation) of the *Durango Herald* and the *Tennessean*.

4.3.3 Top Stories. The "Top Stories" panel on the *Durango Herald* and the *Tennessean's* SERP was persuasive to 53% of the participants. They referred to the "Top Stories" feed as a resource that provided recently published headlines and the publishing dates of the stories. Participants used the headlines as a signal to determine the scope of the source, for example, the *Durango Herald* primarily publishes information about southwestern Colorado, as well as the perceived bias of the source. Some participants also commented that they were more willing to trust sources that had published more recently and with greater frequency (33%), but other participants (17%) were also concerned that they either had not previously heard of the source or believed that papers covering local issues are less reliable. However, when offering a rationale for the credibility of a source based on the "Top Stories", participants were more likely to reference the source's frequency of publication than the headlines' biases.

4.3.4 Social Media. Participants were also interested in the social media pages of the news sources (53%). *The Durango Herald's* Twitter feed was featured on the SERP and the Facebook pages for *the Tennessean* and an imitator site of *the Christian Times* (Christian Times Magazine) displayed ratings on the SERP. Participants were interested in seeing if sources were Twitter verified, and one participant became skeptical of the *Durango Herald* because of its only three-star Facebook rating visible on the Facebook result's rich snippet. Two other participants mentioned the social media rating of the news sources in the rich snippet. For one participant, the absence of a featured Twitter feed was a factor in labeling the *Christian Times* as unreliable.

4.4 Discussion of Results

Our experiments revealed several interesting results. In this section, we have chosen to discuss Knowledge Panels, "Top Stories", organic search results, SERP layout, and the limitations of domain knowledge. We also discuss the differences in claimed participant credibility evaluation behavior and actual behavior.

Table 1: Results of the credibility assessment of *the Durango Herald*, *Tennessean*, and *Christian Times* based on their SERP, following step (2) in the study procedure.

Online source	Is credible?	Is not credible?
Durango Herald	21	9
The Tennessean	23	7
Christian Times	0	30

4.4.1 Knowledge Panel and "Top Stories". A revealing result that requires future inquiry is that participants did not know how to evaluate the credibility of the content on the Knowledge Panel itself. While not all of them mentioned this concern, some participants believed that the news source curated its own Knowledge Panel, while others speculated that Wikipedia produced the Knowledge Panels. No participant explicitly mentioned that Google had developed the Knowledge Panel, but for some it was possibly inferred. The Knowledge Panel is only useful as a signal if participants value the information it provides.

However, participants reacted positively to the contextual information, most notably the Wikipedia description and "Awards" tab. A surprise was how poorly the "Topics They Write About" section performed. The think-aloud protocol makes it difficult to know why participants did not think about a given topic, but one participant remarked that she was dubious of the credibility of the "Topics they write about" panel because the stories in that section are over a month old. Comparing the stale articles to the content on the "Top Stories" panel which contains current news stories, seems to be a plausible answer to why the "Topics they write about" section was not valued more, but further research is required.

4.4.2 Organic Search Results and the Christian Times. The search results themselves are important to participants for assessing the credibility of the source. Beyond the content on the page, the ranking of the news source itself on the SERP was very valuable, too.

The *Christian Times* was the second result on the SERP, providing a negative signal of the perceived credibility of the source. Even when participants did not know that the first result, Snopes, was a fact-checking organization, they still were dubious of the authority of a source that was ranked second for its own name. Additionally, the bold and explicit title from a CBSNews article proclaiming "christiantimes.com - Fake news sites to watch out for" was a big hit with the participants. The irony is that `christiantimes.com` has never been a fake news site. While the current website is clearly not a fake news site, a thorough evaluation of the Wayback Machine page archives revealed that the website never looked the way the CBSNews article [4] claimed. The other articles on the SERP were referring to the far-right conspiracy site `christiantimesnewspaper.com`, active during the 2016 U.S. presidential election.

This is not an isolated example of SERP results from legitimate sources and their sound-alike fakes being combined on the same SERP. Searches for the now defunct fake news site the *Boston Tribune* produce a Knowledge Panel for the reputable *Boston Herald*.

[4]https://web.archive.org/web/20180412003913/https://www.cbsnews.com/pictures/dont-get-fooled-by-these-fake-news-sites/12/

4.4.3 Effects of SERP Layout. Another notable result of our study is the role that the layout of the SERP itself (what blocks it contains and how they are organized) plays in the assessment of credibility. As discussed in the related research section, previous research has identified the importance of a site's layout [12] or has commented on how non-HTML link content changes the behavior of users on SERP [8]. However, no studies have investigated how the site layout of the Google SERP itself changes perceptions of credibility. Observations from our study appear to suggest that enhanced social media presence on the SERP improves the perceived credibility of sources. Therefore, further research is needed to explore the link between a strong social media presence and perceived credibility for news sources on SERP. This makes the errors of Google on generating the SERP for a news source even more costly. The *Tennessean* newspaper has a Twitter account, (@Tennessean), but the Twitter feed is not shown when one searches for "the tennessean". The inconsistency of the SERP content for slightly differing query phrases might have a big impact on decisions that users make about the credibility of sources.

4.4.4 Challenges of Domain Knowledge. 23% of participants evaluated a news source that was nominated for multiple Pulitzer Prizes as not credible (see Table 1). What made the award-winning Tennessean an unreliable site for some users? Of the 7 incorrect classifications, 4 directly involved a distrust of local journalism. This is an important example of the limits of lateral reading for users with limited knowledge of journalism. Regardless of the information Google could place on the SERP, people can make incorrect credibility assessments, because they lack domain knowledge to make sense of the provided information. Concretely, we suspect these participants were not aware that a Pulitzer Prize rewards high-quality journalism.

Similarly, even though the first result for *the Christian Times* was a Snopes article fact-checking the *Christian Times Newspaper*, most participants were unfamiliar with Snopes (90%). So, even though the Google algorithm has helpfully prioritized the fact-checker as the first result, the value of this algorithmic decision is lost on most participants who are unfamiliar with Snopes.

Both the Pulitzer Prize and Snopes examples illustrate that adding context does not empirically increase the ability of web users to accurately access sources.

4.4.5 Contrasting Claimed and Actual Credibility Evaluation Techniques. While the focus of our study is the behavior of web users when performing a lateral reading task, an additional finding is that there is a gap between the claimed and actual behavior of participants with regards to which credibility signals they find important. The top five most frequent responses from the pre-task survey were: "lateral reading" techniques (i.e. see what other sources are reporting the story), only search within reputable sites, find journalist credentials, glance at the end domain name, and examine the page layout. These findings are generally in line with other user studies discussed in the related research section.

In this study, we set participants up to perform lateral reading strategies by instructing them to google the news sources, thereby encouraging the behavior. For example, many participants (67%) commented on what the other search results on the page said about *the Christian Times*.

The second most frequently "claimed" strategy, only using reputable sites, was not possible for this experiment with unfamiliar news sources. However, the language that participants used to describe these reputable sources, included terms such as "trustworthy", "reliable", and "objective" indicating that participants had a solid understanding of the definition of credibility.

Investigating the author or journalist's credentials (30%) was the next most common response; however, when participants were permitted to examine the site itself, only two (7%) commented on the journalists or suggested that Knowledge Panels include journalist information. No participants commented on the fact that all three sources were ".com" sites. Using end domain suffixes to test the credibility of news sources is ineffective, and observed behavior of participants illustrates that it is not actually attempted. The importance of site layout of the SERP has already been discussed, but participants were referring to the actual website layout in this exercise. This finding is replicated in our study with 80% of participants mentioning page layout when viewing one of the three sites' home pages.

While the gap in self-reported and observed behavior is not unique to this experiment, it does prompt the question: how do people select which elements of the SERP they value most? How does the SERP that Google serves to participants alter what they think is a valuable credibility signal? These questions should be explored further in future studies.

4.5 Limitations of the Study

Several factors limit the power of this study. The first one is that this was a study with a sample of convenience: 30 undergraduate students of ages 18-22, who skewed heavily female. Because of their education level, these students are most likely more familiar with more recent web literacy approaches than the general public. A bigger sample with more diversity in demographic traits needs to be interviewed.

Our interviews lasted between 10-20 minutes providing several minutes for participants to examine the credibility of news sources. Users typically do not spend 3-5 minutes on the SERP evaluating the credibility of sources. As a result of the lack of a time constraint, participants were frequently observed scrolling through the entire SERP, a practice that [21, 25] indicate is not typical online behavior.

The ever changing nature of the SERP also created challenges, since our study was conducted over a six-day period. Originally, the prompt for *the Tennessean* was to search "tennessean", but on day five, it was discovered that the Knowledge Panel belonging to the newspaper had disappeared for that query, so that one participant was not able to use the Knowledge Panel to make a credibility assessment. In subsequent interviews, the query "*the* tennessean" produced the expected Knowledge Panel on the SERP.

The Christian Times was selected because its Knowledge Panel did not contain the word fake news, but the other results on the SERP claimed that it was a "fake news" site (due to a mix-up with a similarly named fake news website). However, nine participants (30%) explained that because the source name contained the word "Christian", they were less likely to deem the site as credible. When pressed what about the word "Christian" makes the site unreliable, they clarified that Christian didn't mean unreliable, but at the very least meant biased.

Finally, it is important to mention that the interviewers never defined credibility for the participants. In fact, we used the terms credibility, reliability, and trustworthy interchangeably. When prompted by the study participants about the definition of credibility, interviewers responded with the "story card" scenario described earlier. While the presence of the *Christian Times* website led to interesting discoveries about the inaccuracy of the SERP, as well as participants' prior biases, in future studies we will make sure to choose sources that force participants to do more thinking of what is important to them given the contextual information on the SERP.

5 ANALYSIS OF KNOWLEDGE PANELS

Although the participants in the study were not sure about the provenance of information in the Knowledge Panel, they nevertheless valued the contextual information that it provided about the sources. Given this interest, Google's decision to provide Knowledge Panels that summarize information about an entity in one place seems very helpful. But how consistent and how accurate is the information shown in a Knowledge Panel? We set out to test this for a large number of news sources. Concretely, we tested three different datasets:

(1) The top 100 news sources from the Amazon Web Service Alexa Top Sites, which lists the highest-performing websites globally according to the Alexa Traffic Rank algorithm (AlexaRank). The AlexaRank of a site is calculated from the site traffic over the past three months and is a measure of how many pages a user visits on that site. The "news" category that was used for this analysis features sites such as CNN and The New York Times, as well as news aggregators such as Reddit and Google News.
(2) The USNPL (United States Newspaper List)[5], which is a database of US-based local newspapers, TV, and radio stations broken down by state (n = 7269). We chose to use this list instead of each state's Wikipedia list of newspapers, because we found the USNPL list to be more evenly distributed state-to-state and equally, if not more, complete.
(3) A Buzzfeed News list of highly partisan news sites (n = 677) [6] that includes sites such as MSNBC as well as classic examples of fake news such as `100percentfedup.com`. Many of these sites are of low-authority, the kind of sources users would need support to learn more about.

Through an automatic script, we searched the names of the sources on Google Search, for example, "fox news", and recorded whether the Knowledge Panel existed or not. For the first two datasets we checked whether the panel title corresponded to the site name, to avoid spurious results. This was not possible for the third dataset, for which we used the provided domain names, for example "yesimright", instead of the site name "Yes I'm Right".

Table 2 indicates that the majority of popular sites have the Knowledge Panel, but roughly a third of sources for the two other datasets do. This is problematic. Users are already familiar with

[5] http://www.usnpl.com
[6] https://github.com/BuzzFeedNews/2017-08-partisan-sites-and-facebook-pages/tree/master/data

Table 2: The occurrence and composition of Knowledge Panels in the three datasets on January 3, 2018.

	Knowledge Panel	Writes About	Reviewed Claims
{1} Alexa (n= 100)	96 (96%)	63 (63%)	3 (3%)
{2} USNPL (n= 7269)	2702 (37%)	698 (10%)	1 (0%)
{3} BuzzFeed (n= 677)	230 (34%)	114 (17%)	145 (21%)

Table 3: The occurrence and composition of Knowledge Panels in the three datasets on February 24, 2018.

	Knowledge Panel	Writes About	Reviewed Claims
{1} Alexa (n= 100)	96 (96%)	63 (63%)	0 (0%)
{2} USNPL (n= 7269)	2784 (38%)	1120 (15%)	0 (0%)
{3} BuzzFeed (n =677)	239 (36%)	128 (19%)	0 (0%)

popular sites. It is the least known ones, such as local newspapers or online sources pretending to be legitimate local or national sources, about which users want to learn more.

For the sources that have a Knowledge Panel, their information comes usually from Wikipedia. While some established partisan and "fake news" sites have full Wikipedia pages, many of the sites that we anticipate users should be googling to evaluate their credibility do not rise to the level of notability warranted for a Wikipedia page. Google has recognized the need to still provide supplemental information about these sites and has developed an alternate format (see Figure 9) that replaces the Wikipedia snippet with a a summary of the topics from the "Writes about" section. Based on our observations, it appears that the "Writes About" section is periodically created from topic models [2] of articles the site has published. For the websites of the datasets {1} and {2} we see no examples of this alternative format for describing a website.

We were also curious whether different panels of the search results page co-occur. Concretely, did the presence of a "Top Stories" panel (see Figure 4) increase the likelihood for the existence of a Knowledge Panel? Of the 2835 pages from {2} that have a "Top Stories Panel ", 1307 do not have a Knowledge Panel. Here lies an opportunity for improvement on Google's part: if a source is already recognized as a content publisher in the "Top Stories" panel, there should be an accompanying Knowledge Panel providing context for the source. Table 3 summarizes the data for the repeat data collection in February 2018.

5.1 SERP from January 2018 to February 2018

Google's search algorithm is incredibly dynamic with 500-600 changes being implemented annually [7]. As discussed in the introduction, the most meaningful change from January to February was the removal of the "Reviewed claims" section. While we do not have results from our user study emphasizing the importance of this section for credibility assessment, based on users desire for explicit fact-checks in organic results, it is likely that the "Reviewed Claims" would be perceived as valuable.

[7] moz.com/google-algorithm-change

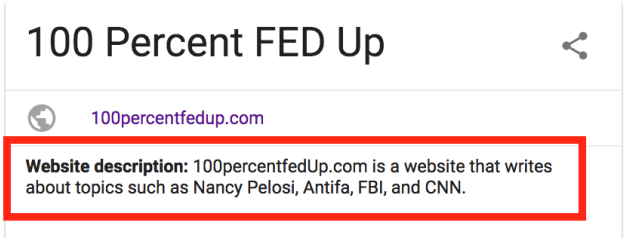

Figure 9: A website description without a Wikipedia citation. This website is part of the Buzzfeed dataset of partisan or fake news websites. The Website description portion is generated automatically by an algorithm. As of April 2018, this description isn't available anymore.

For all three datasets, Knowledge Panel numbers stayed relatively constant. The one notable change was the expansion of the "Writes About" section. In the USNPL dataset, there were only 82 Knowledge Panels added between January and February, but 422 "Writes About" sections added. This increase indicates that Google is actively expanding its "Writes about" section. This is an interesting observation given that many of the participants in our study preferred the "Top Stories" panel.

5.2 Wikipedia's Role

Our analysis reveals that in the majority of cases, the information in the Knowledge Panel is lifted directly from the Wikipedia page. For media companies, such as Fox News, CNN, ABC, the Knowledge Panel also contains information that Google algorithms are inserting from other sources: links to social media presence, or list of anchors and flagship programs. Because the information is coming directly from Wikipedia, this increases the pressure on Wikipedia to both maintain accuracy of information, as well as increase the coverage.

Since less than half of all news sites have Knowledge Panels, we were particularly interested in what separated the sites that have Knowledge Panels from the ones that did not. The best predictor of a knowledge panel was the presence of a Wikipedia page for that site. There are only 373 sites in dataset 2 that have a Knowledge Panel but not a Wikipedia page.

However, not all Wikipedia pages are created equal. An examination of 1043 Wikipedia pages for U.S. newspapers from all 50 states shows several inconsistencies in the infobox panel. We found 79 unique fields for the infobox, fields such as editor, format, owner, etc., but frequently, only a few of these contain information. There is much to be done to provide more information in Wikipedia too.

6 DISCUSSION AND FUTURE WORK

In our digital age, news lives on the web, so the credibility of news sources and their claims should be verified through the web. Often, the first stop in such a verification process is performing "lateral reading" through Google. "I've taken to generally Googling things just to try to get a concept of it."—said one of the participants in a recent study about news consumption [3]. What users look for and what they find when searching on Google influences their decision making and their news literacy. Google's SERP has become

an arena where algorithms, humans, and publishers with good or not-so-good intentions meet and try to influence one another. While most approaches for news literacy focus on the responsibility of the users to evaluate sources, one cannot discount that online platforms, such as Google, which users already trust, play a crucial role in supporting decision making when it comes to deciding which source to trust and why. In our study, the distorted SERP for *the Christian Times* led all participants to label it a not credible source. How could other actors influence and support news literacy efforts? Here are some ideas to consider:

Wikipedia editors could contribute pages for all recognized news sources (e.g., the USNPL database) while also taking care of being consistent in what information they provide. They should also be alert to possible manipulation. It is to be expected that as Knowledge Panels become more prominent, there will be efforts to modify existing pages with incorrect information or to create pages for non-existing sources.

Researchers could build new algorithms to automatically monitor the quality of search results about news and media related queries and point out to Google what is doing wrong. Similarly to how the computer security research community is always watching out for possible failures and weak points in various hardware and software systems, we should treat the information ecosystem enabled by Google Search as something that requires constant monitoring. Researchers in the Web Science community may be well-suited to take the lead on this task.

Educators and literacy organizations could write meaningful web content that provides background and domain knowledge about what makes news sources credible. For example: explain the value of local journalism and its long tradition, the value of recognition by a third party such as a Pulitzer Prize, but also how easy it is to fake Facebook ratings or have a Twitter feed that is constantly tweeting (signals that our participants used to assign credibility).

One thing is clear: we all have a lot of work to do. In the near future, we plan (1) to perform similar experiments on a more representative sample and (2) to examine the credibility signals on mobile devices rather than desktops. The long-term goal of our research is to model how users reason about the credibility of online sources.

REFERENCES

[1] Katarzyna Abramczuk, Michał Kakol, and Adam Wierzbicki. 2016. How to Support the Lay Users Evaluations of Medical Information on the Web?. In *International Conference on Human Interface and the Management of Information*. Springer, 3–13.

[2] David M Blei. 2012. Probabilistic topic models. *Commun. ACM* 55, 4 (2012), 77–84.

[3] Pablo Boczkowski. 12/2017. The Rise of Skeptical Reading. *NiemanLab Predictions for Journalism 2018* (12/2017). http://www.niemanlab.org/2017/12/the-rise-of-skeptical-reading/

[4] Danah Boyd. 2014. *It's complicated: The social lives of networked teens*. Yale University Press.

[5] Monica Bulger and Patrick Davison. 2018. The Promises, Challenges, and Futures of Media Literacy. (2018).

[6] Mike Caulfield. 07/03/2017. How to Find Out If a "Local" Newspaper Site Is Fake (Using a New-ish Google Feature). *Hapgood: A personal blog* (07/03/2017). https://hapgood.us/2017/07/03/how-to-find-out-if-a-local-newspaper-site-is-fake-using-a-new-ish-google-feature/

[7] Mike Caulfield. 2017. *Web Literacy for Student Fact-checkers*. https://webliteracy.pressbooks.com/

[8] Danqi Chen, Weizhu Chen, Haixun Wang, Zheng Chen, and Qiang Yang. 2012. Beyond Ten Blue Links: Enabling User Click Modeling in Federated Web Search. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12)*. ACM, New York, NY, USA, 463–472. https://doi.org/10.1145/2124295.2124351

[9] Brooke Donald. 2016-11-22. Stanford researchers find students have trouble judging the credibility of information online. *https://ed.stanford.edu/news/stanford-researchers-find-students-have-trouble-judging-credibility-information-online* (2016-11-22).

[10] Eric Enge. 05/24/2017. Featured Snippets: New Insights, New Opportunities. *Stone Temple Company Blog* (05/24/2017). https://www.stonetemple.com/featured-snippets-new-Insights-new-opportunities/

[11] Andrew J Flanagin and Miriam J Metzger. 2000. Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly* 77, 3 (2000), 515–540.

[12] BJ Fogg, Cathy Soohoo, David R Danielson, Leslie Marable, Julianne Stanford, and Ellen R Tauber. 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*. ACM, 1–15.

[13] BJ Fogg and Hsiang Tseng. 1999. The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 80–87.

[14] Sarah Frier. 12/17/2017. He Got Rich by Sparking the Fake News Boom. Then Facebook Broke His Business. *Bloomberg* (12/17/2017). https://www.bloomberg.com/news/articles/2017-12-12/business-takes-a-hit-when-fake-news-baron-tries-to-play-it-straight

[15] Michael Galvez. 12/05/2017. Improving Search and discovery on Google. *Google Blog* (12/05/2017). https://blog.google/products/search/improving-search-and-discovery-google/

[16] Alexandru L. Ginsca, Adrian Popescu, and Mihai Lupu. 2015. Credibility in Information Retrieval. *Foundations and TrendsÃĆÂő in Information Retrieval* 9, 5 (2015), 355–475. https://doi.org/10.1561/1500000046

[17] Jacek Gwizdka and Dania Bilal. 2017. Analysis of Children's Queries and Click Behavior on Ranked Results and Their Thought Processes in Google Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (CHIIR '17)*. ACM, New York, NY, USA, 377–380. https://doi.org/10.1145/3020165.3022157

[18] Eszter Hargittai. 2003. The digital divide and what to do about it. *New Economy Handbook* (2003), 821–839.

[19] Eszter Hargittai. 2010. Digital na(t)ives? Variation in internet skills and uses among members of the "net generation". *Sociological inquiry* 80, 1 (2010), 92–113.

[20] Eszter Hargittai, Lindsay Fullerton, Ericka Menchen-Trevino, and Kristin Yates Thomas. 2010. Trust online: Young adults' evaluation of web content. *International journal of communication* 4 (2010), 27.

[21] Gord Hotchkiss and Steve Alston. 2005. *Eye tracking study: An in depth look at interactions with Google using eye tracking methodology*. Enquiro Search Solutions Incorporated.

[22] Carl I Hovland, Irving L Janis, and Harold H Kelley. 1953. Communication and persuasion; psychological studies of opinion change. (1953).

[23] Adrianne Jeffries. 03/05/2017. Google's Featured Snippets are Worse than Fake News. *The Outline* (03/05/2017). https://theoutline.com/post/1192/google-s-featured-snippets-are-worse-than-fake-news

[24] Diane Kelly and Leif Azzopardi. 2015. How many results per page?: A Study of SERP Size, Search Behavior and User Experience. In *SIGIR*.

[25] R Maynes and I Everdell. 2014. The evolution of Google search results pages and their effects on user behaviour. *USA: Mediative* (2014).

[26] Miriam J Metzger. 2007. Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the Association for Information Science and Technology* 58, 13 (2007), 2078–2091.

[27] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. 2007. In google we trust: UsersâĂŹ decisions on rank, position, and relevance. *Journal of computer-mediated communication* 12, 3 (2007), 801–823.

[28] Joshua J Seidman. 2006. The mysterious maze of the World Wide Web: How can we guide consumers to high-quality health information on the Internet. *Internet and health care: Theory, research and practice* (2006), 195–212.

[29] Amit Singhal. 05/16/2012. Introducing the Knowledge Graph: things, not strings. *Google Official Blog* (05/16/2012). https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html

[30] Nicholas Vincent, Isaac Johnson, and Brent Hecht. 2018. Examining Wikipedia With a Broader Lens: Quantifying the Value of Wikipedia's Relationships with Other Large-Scale Online Communities. (2018).

[31] Stanley Wilder. 2005. Information literacy makes all the wrong assumptions. *The Chronicle Review* 51, 18 (2005).

[32] Sam Wineburg and Sarah McGrew. 2017. Lateral Reading: Reading Less and Learning More When Evaluating Digital Information. *Stanford History Education Group Working Paper No. 2017-A1* (2017). https://ssrn.com/abstract=3048994

[33] Ranna Zhou. 11/07/2017. Learn more about publishers on Google. *Google Blog* (11/07/2017). https://www.blog.google/products/search/learn-more-about-publishers-google/