Communications of the Association for Information Systems

Volume 52 Article 5

2023

Seeking Information Using Search Engines: The Impact of Negation on Judgments

Varol O. Kayhan
University of South Florida

Merrill Warkentin
Mississippi State University

Communications of the
Association for Information Systems

Accepted Manuscript

Seeking Information Using Search Engines: The Impact of Negation on Judgments

Varol O. Kayhan
School of Information Systems and Management
Muma College of Business
University of South Florida

Merrill Warkentin
Department of Management and Information Systems
College of Business
Mississippi State University

Please cite this article as: Kayhan, Varol O.; Warkentin, Merrill: Seeking Information Using Search Engines: The Impact of Negation on Judgments, *Communications of the Association for Information Systems* (forthcoming), In Press.

Communications of the Association for Information Systems

Research Paper ISSN: 1529-3181

Seeking Information Using Search Engines: The Impact of Negation on Judgments

Varol O. Kayhan
School of Information Systems and Management
Muma College of Business
University of South Florida

Merrill Warkentin
Department of Management and Information Systems
College of Business
Mississippi State University

Abstract:

We examine how online searches lead to judgment formation in two contexts where statements are negated differently. In one context, a statement (of the form "if P then Q") is negated using "not" (as in "if P then not Q"), while in the other context, it is negated using a different term or phrase (as in "if P then R"). We show that online searches to find information about the validity of this statement generate different results in these contexts. Our findings contribute to extant work on online searches by suggesting that when individuals test the validity of a statement using search engines, their searches lead to categorically different results contingent on the search context. From a practical standpoint, we make recommendations to search engine developers and content providers so online searches generate more inclusive results.

Keywords: Online Information Search, Search Context, Negation, Opposition, Experimental Design.

[Department statements, if appropriate, will be added by the editors. Teaching cases and panel reports will have a statement, which is also added by the editors.]

[Note: this page has no footnotes.]

This manuscript underwent [editorial/peer] review. It was received xx/xx/20xx and was with the authors for XX months for XX revisions. [firstname lastname] served as Associate Editor.] **or** The Associate Editor chose to remain anonymous.]

1 Introduction

According to Pew Research Center (2019a, 2019b), nine out of ten adults in the US use the Internet, three-quarters of households have broadband, and eight out of ten own a smartphone. Search engines are ingrained so much in daily life that individuals conduct online searches even for the most mundane tasks. However, finding the right information on the Web is not easy. For example, as part of an investigative report a Wall Street Journal journalist reported that two Google searches, one for "is milk good for you" and another one for "is milk bad for you," generated vastly different results (Nicas, 2017). To make matters worse, at the time of the report, the search results for "is milk bad for you" not only excluded any health benefits of milk in the result set, but also suggested a wrong and misleading top pick answer about the harms of milk.

These types of misleading search results are usually attributed to either search engines' design flaws in picking and promoting results (Nicas, 2017; Tenner, 2018) or individuals' biases in conducting searches (Kayhan, 2013; White, 2013). However, in this study, we show that there may be another factor in which the *search context* also might contribute to misleading search results. As we discuss later, searches conducted in one context might provide individuals with more confirming evidence about the validity of a statement compared to another context while all else are equal. Therefore, individuals might be more likely to confirm the validity of a statement in one context compared to another just because of the nature of that search context. To explain how this happens, we draw upon the theory of negation (Horn, 1989) and distinguish between two different types of contexts in which statements are negated differently. We show that this distinction causes search results to *include* disconfirming evidence in one search context and *exclude* it in the other even though individuals conduct the same type of online searches. This, ultimately, influences judgments about the information being searched, which may inappropriately influence important personal and professional decisions.

In terms of theory, this work advances our understanding of conditions under which online searches influence judgment formation and decision making. Specifically, we suggest that online searches cause individuals to form more confirming judgments in one type of search context despite the existence of ample disconfirming evidence. We further suggest that this happens because of the characteristics of this search context rather than some of the other factors already identified in extant work such as individuals' cognitive biases or search engines' prioritization or promotion of results. From a practical perspective, we make suggestions to search engine developers on how to generate more balanced search results, so they include disconfirming evidence (upon its existence) about a topic being searched. Further, we make recommendations to content providers on ensuring that their content is included in search results especially if they are providing opposing views about certain topics. Because opposing views can be considered disconfirming evidence depending on how search queries are formulated, and thus, eliminated from search results in certain contexts, ensuring that they are included in relevant search results is important to help individuals make more informed decisions about their online searches.

This paper is organized as follows. First, we review extant work on searching for information on the Web and the theory of negation. Then, we distinguish between the two types of contexts mentioned earlier. Next, we propose three hypotheses to explain why online searches might identify more confirming evidence in one of these contexts compared to the other. After, we describe our experimental setup to test these hypotheses, we conclude by presenting the results and discussing the theoretical and practical implications.

2 Background and Theory Development

2.1 Online Searches and the Theory of Negation

The process of conducting a search on the Web can be summarized in three high-level steps: 1) submitting a search query to a search engine, 2) analyzing the results, and 3) using the results to make a decision, form a judgment, or perform a task (Browne, Pitts, & Wetherbe, 2007; Hodkinson & Kiel, 2003; Kulviwat, Guo, & Engchanil, 2004; Lueg, Moore, & Warkentin, 2003; Marchionini & White, 2007; Sen, King, & Shaw, 2006). This process might unfold differently depending on the type of online search. For example, the process is straightforward for navigational searches that are geared toward finding a specific site's address. In this case, an individual submits a query using a set of keywords (such as "American Airlines") and retrieves the link of a website from the results (http://www.aa.com). These kinds of searches

are not of concern to this study. Neither are searches to learn about a specific topic. For example, submitting a set of keywords (such as “Covid-19 symptoms”) to find information and learn about a topic (i.e., what to expect if Covid-19 is contracted) is out of this paper's scope.

Rather, we are interested in searches by which individuals find and evaluate information about the validity of a statement, hypothesis, or rule (such as “coronavirus leads to brain fog”). In these kinds of searches, the search process depends on the set of keywords chosen by individuals. For example, consider that one wants to test the validity of a statement of the form “if P then Q” using a search engine. In this case, the individual might choose to use both P and Q in their search query because of multiple reasons. First, the framing used for the statement might induce individuals to subconsciously use P and Q as search terms. Prior research provides evidence for this by suggesting that the way information is framed, or presented, partly determines the types of keywords used in online searches (see Kessler & Guenther, 2017). Second, and more importantly, including P and Q in a search query might identify the most relevant result set. Given that search engines rely on keyword-matching (Aujla, Crump, Cook, & Jamieson, 2019; Patel, 2019), using both P and Q as keywords narrows the search space and generates the most relevant result set. Omitting one of these keywords might result in a broader result set and make it difficult as well as frustrating to find information about the statement. Therefore, if both P and Q are likely to be used in a search query, one should expect to retrieve information that satisfies the statement "if P then Q" in search results. These types of information are referred to as *confirming* evidence. Similarly, the use of P and Q in a search query might likely retrieve information that satisfies "if P then *not* Q" (upon its availability) in search results. These types of information are referred to as *disconfirming* evidence. Interested readers can refer to Hempel (1945) and Nickerson (1998) for more information about what constitutes confirming and disconfirming evidence.

Even though the nature of confirming and disconfirming evidence is straightforward, the theory of negation (Horn, 1989), dating back to Aristotle, suggests that disconfirming evidence can take other forms as well. For instance, natural languages lend themselves to four types of negation (or opposition): (1) correlation (e.g., double vs. half); (2) contrariety (e.g., good vs. bad); (3) privation (e.g., blind vs. sighted); and (4) contradiction (e.g., sunny vs. not sunny). For the purposes of this study, the semantic differences between these oppositions are immaterial, but their syntactic differences are important. From a syntactic perspective, the four types of oppositions can be categorized into two groups: *similar-term opposition* (which concerns contradiction only), where the opposition is created using “not,” and *dissimilar-term opposition* (which includes correlation, contrariety, and privation), where the opposition is created using a different word or phrase. It is this distinction that enables us to differentiate between two search contexts, which are discussed next.

2.2 Two Contexts: Similar- vs. Dissimilar-term opposition

The syntactic distinction between the two types of oppositions identified above is important because it determines the nature of confirming and disconfirming evidence regarding a statement. Consider the statement of the form “if P then Q.” In the case of similar-term opposition, any information that satisfies the statement of the form “if P then not Q” (for the original statement “if P then Q”) is considered disconfirming evidence. We refer to these types of contexts, where the distinction between confirming and disconfirming evidence is based on the existence of “not,” as *similar-term opposition contexts* hereafter.

Consider another statement of the form “if P then Q.” In the case of dissimilar-term opposition, any information that satisfies the statement “if P then R” (for the original statement “if P then Q”), where R is the opposite of Q, is considered disconfirming evidence. We refer to these types of contexts, where the distinction between confirming and disconfirming evidence is based on a different term or phrase, as *dissimilar-term opposition contexts* hereafter.

Please note that there could be a third context where disconfirming evidence can include both types of oppositions discussed above. For example, consider another statement of the form “if P then Q.” Also, consider that there could be two types of disconfirming evidence for this statement: one that satisfies “if P then not Q;” and another that satisfies “if P then R.” This type of context is out of this paper's scope. In the rest of this paper, we focus solely on the first two types of contexts where disconfirming evidence is only of one form.

As we discuss in greater detail below, search queries geared toward testing the validity of statements in these two contexts generate different results. In summary, search results include both confirming and disconfirming evidence in similar-term opposition contexts, whereas they include only confirming evidence

in dissimilar-term opposition contexts. The composition of search results, in turn, influences individuals' judgments. We discuss these and develop our hypotheses in the next section.

2.3 Hypotheses Development

As discussed before, when individuals test the validity of a statement using a search engine, they are likely to use keywords obtained from this statement in their search queries. For example, if one wants to test the validity of a statement of the form "if P then Q," they might choose to use both P and Q in the search query because doing so might identify the most relevant result set. Otherwise, the individual might risk obtaining a broader or non-relevant result set that might make it difficult and frustrating to find information about the statement. Further, the framing, or presentation, of this statement (i.e., "if P then Q") might subconsciously induce individuals to use both P and Q in their search queries (see Kessler & Guenther, 2017).

The two contexts, identified earlier, handle the same search query very differently because of the way search engines work. Popular search engines as well as those commonly used by scholars and scientists rely on keyword-matching (Aujla et al., 2019; Patel, 2019). Accordingly, search engines identify documents or sources that include the keywords submitted in queries. If individuals submit the keywords P and Q in their search queries, search engines will return only confirming evidence (for the original statement "if P then Q") in a dissimilar-term opposition context. This is because the search engine will identify only those sources that include both P and Q. Because disconfirming evidence satisfies the statement "if P then R" (for the original statement "if P then Q"), search engines will likely deem these types of evidence irrelevant for the search query (because R is not part of the search query). As a result, disconfirming evidence will be excluded from search results.

On the other hand, in a similar-term opposition context, the same search will return both confirming and disconfirming evidence. This is because both confirming and disconfirming evidence include the keywords P and Q (since any information that satisfies the statement "if P then not Q"—for the original statement "if P then Q"—is considered disconfirming).

Consider an example statement such as "business analysts are technical." In this case, P could be "business analysts" and Q could be "technical." If individuals are asked to test the validity of this statement, they might use both "business analysts" and "technical" in their search queries due to reasons discussed earlier. If this is a similar-term opposition context, where the disconfirming evidence satisfies the statement "business analysts are not technical" (i.e., "if P then not Q"), search results will likely include both confirming and disconfirming evidence because both types of evidence can be retrieved using the keywords "business analysts" and "technical."

However, consider another example statement such as "software development is easy." In this case, P could be "software development" and Q could be "easy." If individuals are asked to test the validity of this statement, they might use both "software development" and "easy" in their search queries. If this is a dissimilar-term opposition context, where disconfirming evidence satisfies the statement "software development is hard," search engines will likely exclude disconfirming evidence from top results because these types of information might be irrelevant when the keywords "software development" and "easy" are used in a search query. In this case, individuals must either use the keyword "hard" explicitly in the same or another search query or use a higher-level term (such as "difficulty") to increase the likelihood that both confirming and disconfirming evidence can be retrieved simultaneously. However, the cognitive costs of searching for oppositions and the effects of framing, established in extant work, will induce individuals to resort to "software development" and "easy" as their primary search query, which will exclude disconfirming evidence.

The peculiarity of the dissimilar-term opposition contexts and the exclusion of disconfirming evidence in these contexts can be observed in the real-world example discussed in the Introduction section. Recall, two Google searches, "is milk good for you" and "is milk bad for you," generated different results in an investigative report (Nicas, 2017). Here, one can consider "milk" as P and "good" as Q. However, the second query, which is the opposition of the first, is of the form "if P than R," where P is "milk" and R is "bad." Because this is a dissimilar-term opposition context, search results might not have included any health benefits of milk when the search query was of the form "is milk bad for you."

Therefore, we posit that the context determines the composition of search results generated by a search engine. In a dissimilar-term opposition context search queries are less likely to retrieve disconfirming evidence and more likely to retrieve only confirming evidence in search results. To the contrary, in a

similar-term opposition context search queries are more likely to retrieve both confirming and disconfirming evidence in search results. This leads to our first hypothesis.

Hypothesis 1: While testing the validity of a statement, individuals conducting searches in a dissimilar-term opposition context will identify more confirming evidence compared with individuals conducting searches in a similar-term opposition context.

We also posit that the context determines the type of evidence retrieved from search results such that individuals retrieve and read more confirming evidence from search results in dissimilar-term opposition contexts. The reason this happens is rooted in the composition of results generated by the search engine. As discussed for H1, search engines use keyword-matching to generate results. When individuals are presented with a statement of the form "if P then Q," they are more likely to use the keywords P and Q in their search queries. In a similar-term opposition context (where disconfirming evidence satisfies the statement of the form "if P then not Q"), this search query will generate a result set with both confirming and disconfirming evidence. Therefore, individuals will be more likely to retrieve and read both types of evidence. However, in a dissimilar-term opposition context (where disconfirming evidence satisfies the statement of the form "if P then R" for the original statement "if P then Q") the result set will mostly include confirming evidence and exclude disconfirming evidence (as posited in H1). In this case, individuals have no choice but to retrieve and read confirming evidence from the results. This leads to our second hypothesis:

Hypothesis 2: While testing the validity of a statement, individuals conducting searches in a dissimilar-term opposition context will read more confirming than disconfirming evidence compared with individuals conducting searches in a similar-term opposition context.

Last, we posit that individuals are more likely to indicate that a statement is valid when they test its validity in a dissimilar-term opposition context. Again, this is due to individuals' exposure to mostly, and sometimes solely, confirming evidence. Because search results in this type of context mostly exclude disconfirming evidence, individuals will retrieve and read confirming evidence (as posited in H2). As a result, they will be more likely to indicate that the statement is valid. It is worth noting that even if participants might indicate that the statement is valid, the underlying reason behind this is merely being exposed only to confirming evidence. This leads to the third and final hypothesis of this study.

Hypothesis 3: While testing the validity of a statement, individuals conducting searches in a dissimilar-term opposition context will be more likely to agree with the validity of the statement compared with individuals conducting searches in a similar-term opposition context.

To test these hypotheses, we conducted a rigorous experiment which is described next.

3 Experimental Design

For our experiment, we focused on a controversial issue in healthcare: coffee consumption. While certain studies report that coffee consumption leads to good health outcomes, others suggest that it leads to particularly bad outcomes including cancer (LaMotte, 2018). The link between coffee consumption and hypertension is of specific interest to this study because it lends itself to both similar- and dissimilar-term opposition contexts. This, in turn, allows us to conduct a controlled experiment using the same framing. For example, consider the statement "there is a link between *coffee consumption* and *hypertension.*" Any information that supports this link is considered confirming evidence. If this is a similar-term opposition context, any information that satisfies the statement "there is *not* a link between coffee consumption and hypertension" can be considered disconfirming evidence. By the same token, since the opposite of hypertension is *low blood pressure* (per the Systematized Nomenclature of Medicine: Clinical Terms (SNOMED CT)), a dissimilar-term opposition context can be created. Therefore, any information that satisfies the statement "there is a link between *coffee consumption* and *low blood pressure*" is considered disconfirming evidence in this context because hypertension and low blood pressure are at the opposite ends of the same spectrum, and coffee consumption cannot be associated with hypertension and low blood pressure at the same time.

This duality is observed in peer-reviewed studies too. Conducting a search through Google Scholar and PubMed identifies studies that report a link between coffee consumption and hypertension, coffee consumption and low blood pressure, as well as studies that fail to find a link between coffee consumption

and hypertension. This creates fertile ground to create these two contexts and test our hypotheses without manipulating framing.

We used several peer-reviewed studies identified through Google Scholar and PubMed to create three sets of abstracts (each set consisting of four bogus abstracts). We used the link between coffee consumption and hypertension as the reference point. Therefore, our first set of abstracts consisted of confirming evidence and included four abstracts that reported a link between coffee consumption and hypertension. The second set consisted of disconfirming evidence using four abstracts that reported no link between coffee consumption and hypertension. This allowed us to generate the similar-term opposition context. The third set of abstracts also consisted of disconfirming evidence using another four abstracts that reported a link between coffee consumption and low blood pressure. By combining these abstracts with the first set of abstracts that reported a link between coffee consumption and hypertension, we generated the dissimilar-term opposition context.

The development of these abstracts was an arduous task; Appendix A discusses their development including their realism and validity. Following the development of the abstracts, we created a separate webpage for each abstract. Each webpage had a title, abstract, bogus author details, and bogus journal details (see Appendix B, Figure B1 for an example webpage). We also created 30 additional webpages to act as noise. These did not concern any kind of link between coffee consumption and blood pressure but were abstracts of other peer-reviewed studies about either coffee consumption or blood pressure, but not both. All pages were hosted on a web server and included several PHP scripts to track participants' online activities.

To enable participants to conduct keyword searches, we used an open-source search engine provided by Open Search Server (http://www.opensearchserver.com/). This search engine is very similar to Google in every respect: it indexes webpages, helps users conduct keyword searches, and presents results the same way Google does. It is highly customizable and provides a Google-like experience. The settings used for the search engine are described in detail in Appendix C.

As a result, we generated two separate web servers (with each server having its own search engine). While the two servers indexed both confirming and disconfirming evidence, the type of disconfirming evidence indexed by these servers differed. As a result, one of the servers hosted a similar-term opposition context while the other hosted a dissimilar-term opposition context. The nature of confirming and disconfirming evidence indexed by two servers are summarized in Table 1.

Table 1. Web Servers Used in the Experiment

	Web Server 1 **(Similar-term opposition context)**	**Web Server 2** **(Dissimilar-term opposition context)**
Disconfirming evidence	Four abstracts that suggest there **is not** a link between **coffee consumption** and **hypertension**	Four abstracts that suggest there **is** a link between **coffee consumption** and **low blood pressure**
Confirming evidence	Four abstracts that suggest there **is** a link between **coffee consumption** and **hypertension**	Four abstracts that suggest there **is** a link between **coffee consumption** and **hypertension**
Noise	Thirty abstracts on either coffee consumption or blood pressure (but not both)	Thirty abstracts on either coffee consumption or blood pressure (but not both)

Please note that we only manipulate the content hosted by the two servers in this experiment. Therefore, a suitable manipulation check is determining whether each search engine will serve the related content when participants receive our experiment's instructions and submit search queries. A small-scale pilot study discussed in the Pretest 3 section of Appendix A demonstrates that search engines successfully served their content, including the confirming and disconfirming evidence, to participants in both search contexts even though participants did not know the nature or types of content indexed by these search engines. This provides further evidence that the manipulation of the content in these two search contexts, which is the focus of this experiment, worked as intended.

3.1 Control Variables

To test our hypotheses, we included individuals' familiarity with the topic as one of the control variables in this study. Familiarity has its roots in earlier work on decision making (Nickerson, 1998; Tversky & Kahneman, 1974). For example, Allen and Parsons (2010) show that individuals who have higher domain familiarity tend to write different statements than those who have lower domain familiarity when they are asked to search a database using Structured Query Language (SQL). Similarly, both Turner (2001) and Croskerry (2002) show that individuals exhibit different types of search strategies depending on their perceptions of the requirements of the search task as well as their approach to problem-solving. For example, an individual might choose to use an exhaustive search, while another might choose a heuristic-based search for the same search task. Therefore, in the context of online search, individuals might choose different search terms and construct different search queries depending on their familiarity with the statement or the search strategy they adopt for that statement.

Additionally, we use other control variables such as participants' age (six categories), gender (two categories), education (six categories), and time spent on the web in a typical day (five categories, by hours) to control for their approaches in completing the experimental task.

4 Participants and Procedure

Participants were recruited from Amazon Mechanical Turk (AMT). All participants were anonymous, and all study procedures were consistent with research ethics standards and were approved by the university Institutional Review Board. Participants received $1 for taking part in the study.

To recruit participants, we created a project in AMT and provided a link for the study. When participants clicked on this link, a script running at this link assigned them to one of the two web servers in a round-robin fashion. In the first screen, we captured each participant's initial familiarity with the link between coffee consumption and blood pressure (using a 7-point scale) before displaying any other instructions. The scale used to measure initial familiarity is provided in Appendix B, Figure B2. The instructions defined the term *hypertension* in case participants were not familiar with it. An example set of instructions is provided in Appendix B, Figure B3.

The instructions provided a link to the search engine to conduct searches regarding these statements (see Appendix B, Figure B4 for the search interface and Appendix B, Figure B5 for the presentation of results). The instructions did not indicate the number or nature of the webpages indexed by the search engine.

After conducting their searches, participants were asked to return to the instruction page to submit their answer ("valid," "invalid," or "other") and their level of confidence in their answer. The last step of the experiment captured the other control variables discussed above.

5 Results

5.1 Participants

A total of 105 participants were recruited across the two groups. However, usable answers were obtained from only 70 participants. Excluded participants either did not conduct any searches or did not read any abstracts. The number of participants in each group and their characteristics are presented in Table D1 of Appendix D.

5.2 Composition of Search Results

We used server-side scripting to capture the result set generated for each search query of each participant. In each result set we identified the number and nature of the links (i.e., whether confirming or disconfirming) returned. Then, we calculated the proportion of confirming links in each result set. Therefore, values closer to 1 represented search results consisting mostly of confirming links, and values closer to 0 represented search results consisting mostly of disconfirming links. Consequently, values closer to 0.5 represented balanced search results with equal number of confirming and disconfirming links. Note that this ratio did not consider the other abstracts that acted as noise.

Participants conducted a total of 135 searches (all usable). We conducted an ANCOVA, where the dependent variable was the proportion of confirming links, and the independent variable was the type of

context. The covariates were familiarity with the statement, age, gender, education, and time spent on the web. The level of analysis for this model was a single search. Since a participant could perform multiple searches, linear regression's assumption concerning the independence of errors could be violated. Therefore, we conducted a Durbin-Watson test on the first order effects. The analysis produced a result of 1.99, which is between the recommended range of 1.5 to 2.5. This suggests that the errors were independent in the analysis. We also conducted a heteroscedasticity test to check for randomness in residuals. Three tests of heteroscedasticity (White test, modified Breusch-Pagan test, and F test) were all non-significant ($p>0.05$) indicating that there was no evidence of heteroscedasticity.

The results, presented in Table 2, suggest that the composition of search results between the two groups is significantly different from each other. The proportion of confirming abstracts is 0.55 in the similar-term opposition context, whereas it is 0.90 in the dissimilar-term opposition context ($p<0.001$). None of the control variables are significant.

This suggests that there were equal number of confirming and disconfirming links in search results in the similar-term opposition context. For example, if participants received four disconfirming links, they also received four confirming links in the results. However, this value jumped to 90% in favor of confirming links in the dissimilar-term opposition context. Therefore, on average, 90% of the links returned in a search result consisted of confirming evidence. For example, if a search result returned four links, all of them confirmed the relationship between coffee consumption and hypertension. This supports Hypothesis 1.

Table 2. Results for composition of search results (significant effects are in bold)

ANCOVA: Dependent variable: proportion of confirming links in search results		
Variables	**Beta Coeff.**	**Sig.**
Type of context	0.34	<0.001
Initial familiarity	-0.00	0.95
Age	-0.00	0.93
Gender	0.03	0.52
Education	-0.02	0.54
Time spent on web	0.02	0.31
Adj. R-Sq = 0.39		

5.3 Information Read by Participants

Next, we examined the types of information, or evidence, read by participants by analyzing the types of links on which they clicked. We assumed that if a participant clicked on a link and retrieved the abstract, he/she read that abstract. Here, the level of analysis was a single participant.

We built another ANCOVA. The dependent variable was the proportion of confirming links clicked by participants. The independent variable was the type of context. The covariates were the same.

Table 3. Results for Information Retrieved by Participants (Significant effects are in bold)

ANCOVA: Dependent variables: proportion of confirming links clicked by participants		
Variables	**Beta Coeff.**	**Sig.**
Type of context	0.39	<0.001
Initial familiarity	0.01	0.62
Age	0.02	0.33
Gender	-0.08	0.72
Education	-0.01	0.59
Time spent on web	0.00	0.72
Adj. R-Sq = 0.24		

Results, presented in Table 3, show that type of context is significant while the control variables are not. Comparison of means shows that the ratio of confirming links clicked by participants is significantly different between the two groups (0.47 vs. 0.86, p<0.001). Therefore, in the similar-term opposition context, 47% of the links clicked by participants were confirming while the remaining 53% were disconfirming. This suggests that participants, on average, clicked on equal number of confirming and disconfirming links. However, this number jumped to 86% in the dissimilar-term opposition context. Therefore, on average, 86% of the links clicked by participants are confirming links in this group. This shows that participants in the dissimilar-term opposition context overwhelmingly retrieved and read confirming evidence. This supports H2.

5.4 Participants' Judgments

Finally, we evaluated participants' judgments by analyzing the answers they provided. To analyze the answers, we used a crosstab. The results, presented in Table 4, show significant differences between participants' answers (Chi-Sq=14.86, p<0.001).

Table 4. Agreement with the Statement (by type of opposition)

	Similar-term context	Dissimilar-term context
Valid	15 (42%)	28 (82%)
Not valid	13 (36%)	1 (3%)
Other	8 (22%)	5 (15%)
Total	36 (100%)	34 (100%)

As shown in Table 4, only 42% of the participants in the similar-term opposition context indicated that the statement was valid. In contrast, 82% of the participants in the dissimilar-term opposition context indicated that the statement was valid. Therefore, more participants indicated that the statement was valid in the dissimilar-term opposition context with statistical significance (p<0.001). This supports H3.

We also investigated whether participants' level of confidence in their answers differed across the two contexts. Therefore, we built another ANCOVA where the dependent variable was decision confidence, and the independent variable was the type of context. We also included the covariates into the analysis as before. Results showed that participants' level of confidence was not statistically different between the two groups. The means were 5.2 vs. 5.5 for similar- and dissimilar-term contexts respectively (p=0.44). The covariates did not have any statistical effects either.

5.5 Robustness of the Results

To establish the robustness of our findings, we repeated the experiment with the same settings except one. We enabled the autocomplete feature of the search engine so that it could make query recommendations in real time as participants typed their search queries. This feature was nearly identical to that of Google. Though, in this case, query recommendations were based on the keywords indexed by the search engine. This provided a Google-like search experience that is commonly observed in many mainstream search engines.

We recruited an additional 121 participants; however, we received usable responses from only 81 of them. We used the same two contexts and procedure discussed earlier. The results were same as previously discussed findings. Participants in the dissimilar-term opposition context received more confirming abstracts in their search results than those in the similar-term opposition context after accounting for the control variables (78% vs. 57% respectively, p<0.001). Similarly, participants in the dissimilar-term opposition context read more confirming links than those in the similar-term opposition context after accounting for the control variables (68% vs. 46% respectively, p=0.03). Consequently, more participants in the dissimilar-term opposition context indicated that the statement was valid than those in the similar-term opposition context (68% vs. 32% respectively, Chi-Square = 10.47, p=0.005). There was no statistical difference between the confidence levels of participants (5.40 for similar-term opposition vs. 5.79 for dissimilar-term opposition, p=0.28). Note that we also conducted one additional analysis by combining the data set obtained for this experiment with the first experiment. This allowed us to test whether autocomplete had a main effect or an interaction effect with the search context. This analysis

showed that autocomplete neither had any main or interaction effects in any of the analyses during hypothesis testing.

6 Discussion

6.1 Summary of Results

In this study, we examine how individuals test the validity of a statement in two different contexts where statements are negated differently. The results of our analyses show that in the dissimilar-term opposition context, where disconfirming evidence is identified using a different term or phrase in relation to the original statement being validated, search queries generate mostly confirming evidence (in support of H1). As a result, individuals retrieve and read more confirming evidence (in support of support H2) and are more likely to agree with the validity of the statement despite the existence of disconfirming evidence (in support of H3). Hypotheses of this study and the results are summarized in Table 5.

Table 5. Summary of Findings

Hypotheses		Support
1	While testing the validity of a statement, individuals conducting searches in a dissimilar-term opposition context will identify more confirming evidence compared with individuals conducting searches in a similar-term opposition context.	Yes
2	While testing the validity of a statement, individuals conducting searches in a dissimilar-term opposition context will read more confirming than disconfirming evidence compared with individuals conducting searches in a similar-term opposition context.	Yes
3	While testing the validity of a statement, individuals conducting searches in a dissimilar-term opposition context will be more likely to agree with the validity of the statement compared with individuals conducting searches in a similar-term opposition context.	Yes

6.2 Limitations

This paper is not without limitations. First, we conducted the experiment online without the presence of an experimenter. Therefore, it is possible that participants might have been distracted while completing the experimental task. However, the logs of the participants (obtained from both AMT and the web server used for the studies) indicate that participants did not take more than 10 minutes to complete the task. Therefore, the experimental task was not long or difficult enough to allow for distractions. Further, our findings are based on a relatively small number of participants (n=70). Even though we demonstrate the robustness of the results using a second sample (n=81), we urge future research to use a bigger sample size to extend or validate our work.

Second, we neither manipulated the task importance, nor offered any incentives for decision accuracy. Although this might raise concerns about participants' engagement with the experimental task, there is evidence that most participants took the experiment seriously: if we use the number of links clicked in a search result as a proxy for a participant's level of engagement, we see that nearly 68% of participants downloaded two or more documents from their search results, indicating a reasonable level of engagement.

Third, we did not use a mainstream search engine, such as Google, to conduct our experiments. This was because Google changed its terms and conditions at the time of this study and prevented us from building a custom search engine to index the pages used in the experiment. However, pilot tests conducted using Google (before it changed its terms and conditions) generated identical results (see Kayhan, 2015).

Fourth, we did not measure participants' initial beliefs about the statements used in this study. Even though we measured participants' initial familiarity with the statements, participants might still have had initial beliefs. However, it is also worth mentioning that participants had low initial familiarity with the statements. Therefore, it is likely that they might not have had strong initial beliefs about the statements.

Fifth, our study did not examine the context in which there could be two different forms of disconfirming information. For example, a statement of the form "if P then Q" can be disconfirmed using information that satisfies "if P then not Q" as well as information that satisfies "if P then R" at the same time. It is possible that search strategies employed in this context might generate the same results observed in the similar-

term opposition context because of the keyword Q that exists in both confirming and disconfirming evidence. We urge future work to examine searches in this context and determine whether search results differ from those reported in this study. Further, such an endeavor might shed light on how individuals form judgments when there are two different types of disconfirming evidence.

Sixth, our study does not examine contexts where there could be negations of disconfirming evidence. For example, consider a dissimilar-term opposition context, where a statement of the form "if P then Q" can be disconfirmed using information that satisfies "if P then R." It is possible that there can also be information that satisfies "if P then *not* R." Even though our study does not account for these types of contexts, the findings reported for H1 and H2 might still be applicable for them. This is because keywords needed to retrieve the negated disconfirming information (P and R) are needed to retrieve the disconfirming evidence for the original statement. Therefore, search results might still include only confirming evidence if individuals do not use the keywords P and R in their search queries. However, the results for H3, which concern judgments, might change. Depending on how individuals interpret the negated disconfirming evidence, these contexts might lead individuals to dismiss the disconfirming evidence. Therefore, H3 might change for the worse to suggest that individuals might form more confirmatory judgments.

Finally, one can criticize the realism of this study by suggesting that there might be very few dissimilar-term opposition contexts in the real world or that it is unlikely that real search contexts exclude opposition keywords. Even though these criticisms have some validity, one can argue otherwise. First, there are many dissimilar-term opposition contexts. For example, consider *introversion* and *extroversion*, which make a dissimilar-term opposition pair. One stream of research argues that a high level of attention is linked to introversion, while another argues that it is associated with extroversion (see Koelega, 1992). Similarly, consider an opposition created by *boy* versus *girl*. One camp posits that aggressiveness is more prevalent among boys, while the other suggests that it is more common among girls (Maccoby & Jacklin, 1976; Tieger, 1980). If these examples are presented using "if P then Q," disconfirming evidence will be "if P then R." Therefore, there is no shortage of dissimilar-term opposition contexts if one wants to find them. Second, it is possible that the titles or abstracts of peer-reviewed studies that are indexed by search engines commonly used by scientists or researchers might exclude opposing keywords. In fact, the abstracts used in this study were created based on the real abstracts of peer-reviewed studies which did not include the opposing keywords (see Appendix A for details). Third, and most importantly, there is no shortage of peer-reviewed studies that report conflicting and confusing findings. For example, between 1970 and 2020, coffee consumption has been linked to increased risk of heart attack, increased risk of urinary tract cancer, decreased risk of liver cancer, increased risk of lung disease, reduced risk of stroke, reduced risk of prostate cancer, reduced risk of heart failure, and reduced risk of heart disease (Brown, 2020; LaMotte, 2018). Therefore, the search context and the nature of disconfirming evidence are becoming more important in online searches.

6.3 Theoretical Implications

Findings of this study contribute to our understanding of judgment formation and decision making. Extant work suggests that when individuals are asked to conduct online searches about a statement, rule, or hypothesis, they might get misleading results either because of their cognitive biases (Kayhan, 2013; White, 2013) or search engines' promotion of wrong results (Nicas, 2017; Tenner, 2018). As a result, individuals might form more confirming judgments about the topic being searched. In this paper, we show that individuals might retrieve misleading search results because of the type of search context as well regardless of their biases or search engines' biases. This contributes to our understanding of judgment formation using online searches. When disconfirming evidence is identified using *not* (as in "if P then not Q" for the original statement "if P then Q"), online searches return both confirming and disconfirming evidence. However, when disconfirming information is identified using a different term or phrase (as in "if P then R," where R is the opposite of Q, for the original statement "if P then Q"), then online searches fail to identify disconfirming evidence and overwhelmingly return confirming evidence. This suggests that the two types of search contexts identified in this paper can be consequential for judgment formation in addition to the other factors already identified in extant work. Therefore, this study adds to extant work in online searches by suggesting that search engines, as the most salient IT artifact in online searches, must be able to account for these two types of search context. Instead of focusing solely on keyword matching, which can exacerbate the distinction between the two search contexts, developers of search engines must identify new mechanisms to deal with oppositions and negation so that both confirming and disconfirming evidence (if exists) can be incorporated to search results.

6.4 Practical Implications

The web is an incredibly large and rich content repository that is only as useful as the searches we conduct to find information. In this study, we show how searches conducted to validate the same statement steer individuals toward different paths because of the types of search contexts in which these searches are conducted. This, in turn, leads individuals to form different judgments about the validity of the statement. This is troublesome because it suggests that individuals are likely to form a judgment not because of their own merit or the availability of best information on the topic, but rather the type of search context.

Preventing this is not easy because the search context and the algorithms used by search engines are extraneous factors over which users have no control when they conduct online searches. Therefore, we must consider training users on how to conduct proper searches for hypothesis-testing tasks (Anderson, 1982; Ferreira, Garcia-Marques, Sherman, & Sherman, 2006; Lau & Coiera, 2009; Tweney et al., 1980). This can be accomplished by first understanding the ways in which individuals interact with search engines and then focusing on individuals' cognitive search techniques and learning styles (Hong, Thong, & Tam, 2004; Recker, Reijers, & van de Wouw, 2014; Riding & Sadler-Smith, 1997; Wang & Doong, 2010). We must also educate individuals on how to test the validity of a statement or hypothesis. Specifically, we must ensure that a statement or hypotheses should be tested by deliberate attempts to find disconfirming evidence about that statement or hypothesis (Klayman & Ha, 1987; Kuhn, 2012). We can do this not only for schoolchildren, but also for adults by developing user guides and best practices for search engines. As a result, we can change the ways in which individuals formulate search queries and seek information.

Search engine developers also can offer solutions to help in this regard. One possibility is the development of an advanced search interface that is specifically designed for hypothesis-testing tasks. Such an interface might allow users to enter the keywords that make up the hypothesis in separate fields. For example, a hypothesis of the form “if P then Q” could be validated by entering the keyword P in one field and the keyword Q in another field. Then, the search engine can not only identify all synonyms of P and Q, but also their negated forms to conduct a more comprehensive search. This way, search engines can make the search experience for dissimilar-term opposition contexts less exclusionary and more like the one observed for similar-term opposition contexts.

Alternatively, search engine developers can find new ways of searching for information, especially for hypothesis-testing tasks, without exclusively relying on keyword-matching. Some possibilities are using meta engines (Roussinov & Chau, 2008) or knowledge maps (Chung, Chen, & Nunamaker Jr, 2005; Dang et al., 2012). In this regard, healthcare-oriented dictionaries, thesauri, or semantic networks (such as those used in the Unified Medical Language System (UMLS)) can also provide more keywords to include in search queries even if users do not explicitly use them when they search for information. For example, when the term “hypertension” is searched within UMLS, the system returns 1) the synonyms of hypertension (such as “high blood pressure”); 2) the antonyms (such as “low blood pressure”); as well as 3) a hierarchical map of all related terms (where the term “blood pressure” is a parent term of both “hypertension” and “low blood pressure”). Therefore, a search engine that leverages a system such as UMLS can include all related terms in a search query even if individuals submit only one. This might generate a more comprehensive search and provide more (confirming or disconfirming) information about the statement, hypothesis, or relationship being investigated.

Finally, content providers can use certain strategies to ensure that information, especially those providing opposing views, are included in search results. By content providers, we do not mean content hosting platforms, but rather entities such as government agencies, non-profits, private businesses, and even private citizens that provide Web content to serve the needs of others. These entities can and should determine if and how their content can be conceptualized using "if P then Q," and thus, what kind of keywords can be used to create both similar-term and dissimilar-term oppositions. For example, consider a content provider who has information about the effects of coffee consumption on low blood pressure (e.g., "if P then Q"). Someone searching for the effects of coffee consumption on hypertension (e.g., "if P then R") might not be able to retrieve this content as shown in this paper. As a remedy, the content provider can incorporate keywords that include *not low blood pressure* (i.e., "not Q") as well as *hypertension* (i.e., "R") to the content so that a dissimilar-term opposition context can be converted to a similar-term opposition context using oppositions. As a result, content providers might wield more control over how their content is indexed by search engines. This could be particularly important for fighting misinformation, where malicious actors might use specific framing to steer online searches toward certain

content. Those who want to counter such misinformation can and should include oppositions of essential keywords so that searches conducted toward evidence that confirms misinformation can include disconfirming evidence as well.

7 Conclusion

Our goal was to understand how online searches lead to judgment formation in two contexts where statements are negated differently. We demonstrated that searches employed in a dissimilar-term opposition context, where a statement is negated using a different term or phrase, led to search results that include mostly, and sometimes only, confirming evidence. This, in turn, led individuals to retrieve and read mostly confirming evidence from search results, and ultimately, agree with the validity of the statement—despite the existence of evidence that disconfirmed the statement. On the other hand, identical searches conducted in a similar-term opposition context, where a statement is negated using "not," led to search results with both confirming and disconfirming evidence. This led individuals to retrieve and read both kinds of evidence from search results and to have less agreement with the validity of the statement. Our findings deepen our understanding of judgment formation in online searches.

References

Allen, G., & Parsons, J. (2010). Is query reuse potentially harmful? Anchoring and adjustment in adapting existing database queries. *Information Systems Research, 21*(1), 56-77.

Anderson, C. A. (1982). Inoculation and counterexplanation: Debiasing techniques in the perseverance of social theories. *Social Cognition, 1*(2), 126-139.

Aujla, H., Crump, M. J. C., Cook, M. T., & Jamieson, R. K. (2019). The Semantic Librarian: A search engine built from vector-space models of semantics. *Behavior Research Methods, 51*(6), 2405-2418.

Brown, J. (2020). Why coffee could be good for your health. *BBC.com*. Retrieved from https://www.bbc.com/future/article/20201028-the-benefits-of-coffee-is-coffee-good-for-health

Browne, G. J., Pitts, M. G., & Wetherbe, J. C. (2007). Cognitive Stopping Rules for Terminating Information Search in Online Tasks. *MIS Quarterly, 31*(1), 89-104.

Burke, V., Beilin, L., German, R., Grosskopf, S., Ritchie, J., Puddey, I., & Rogers, P. (1992). Association of lifestyle and personality characteristics with blood pressure and hypertension: a cross-sectional study in the elderly. *Journal of Clinical Epidemiology, 45*(10), 1061-1070.

Chung, W., Chen, H., & Nunamaker Jr, J. F. (2005). A Visual Framework for Knowledge Discovery on the Web: An Empirical Study of Business Intelligence Exploration. *Journal of Management Information Systems, 21*(4), 57-84.

Croskerry, P. (2002). Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Academic Emergency Medicine, 9*(11), 1184-1204.

Dang, Y., Zhang, Y., Chen, H., Brown, S. A., Hu, P. J.-H., & Nunamaker, J. F. (2012). Theory-Informed Design and Evaluation of an Advanced Search and Knowledge Mapping System in Nanotechnology. *Journal of Management Information Systems, 28*(4), 99-128.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology, 91*(5), 797-813.

Hempel, C. G. (1945). Studies in the Logic of Confirmation. *Mind, 54*(213), 1-26.

Hodkinson, C., & Kiel, G. (2003). Understanding Web Information Search Behavior: An Exploratory Model. *Journal of End User Computing, 15*(4), 27-48.

Hong, W., Thong, J. Y. L., & Tam, K. Y. (2004). The Effects of Information Format and Shopping Task on Consumers' Online Shopping Behavior: A Cognitive Fit Perspective. *Journal of Management Information Systems, 21*(3), 149-184.

Horn, L. (1989). *A Natural History of Negation*. Chicago IL: University of Chicago Press.

Jee, S. H., He, J., Whelton, P. K., Suh, I., & Klag, M. J. (1999). The effect of chronic coffee drinking on blood pressure: a meta-analysis of controlled clinical trials. *Hypertension, 33*(2), 647-652.

Kayhan, V. O. (2013). Seeking Health Information on the Web: Positive Hypothesis Testing. *International Journal of Medical Informatics, 82*(4), 268-275.

Kayhan, V. O. (2015). *Confirmation Bias: Roles of Search Engines and Search Contexts*. Paper presented at the Proceedings of the 36th International Conference on Information Systems (ICIS), Fort Worth, TX, USA.

Kessler, S. H., & Guenther, L. (2017). Eyes on the frame: Explaining people's online searching behavior in response to TV consumption. *Internet Research, 27*(2), 303-320.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211-228.

Koelega, H. S. (1992). Extraversion and vigilance performance: 30 years of inconsistencies. *Psychological Bulletin, 112*(2), 239-258.

Kuhn, T. S. (2012). *The structure of scientific revolutions* (4th ed.). Chicago and London: University of Chicago press.

Kulviwat, S., Guo, C., & Engchanil, N. (2004). Determinants of Online Information Search: A Critical Review and Assessment. *Internet Research, 14*(3), 245-253.

LaMotte, S. (2018). Health effects of coffee: Where do we stand? *CNN.com*. Retrieved from http://www.cnn.com/2015/08/14/health/coffee-health/

Lau, A., & Coiera, E. W. (2009). Can cognitive biases during consumer health information searches be reduced to improve decision making? *Journal of the American Medical Informatics Association, 16*(1), 54-65.

Lueg, J. E., Moore, R. S., & Warkentin, M. (2003). Patient Health Information Search: An Exploratory Model of Web-Based Search Behavior. *Journal of End User Computing, 15*(4), 49-61.

Maccoby, E. E., & Jacklin, C. N. (1976). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Marchionini, G., & White, R. (2007). Find what you need, understand what you find. *International Journal of Human Computer Interaction, 23*(3), 205-237.

Nicas, J. (2017). Google Has Picked an Answer for You—Too Bad It's Often Wrong. *WSJ.com*. Retrieved from https://www.wsj.com/articles/googles-featured-answers-aim-to-distill-truthbut-often-get-it-wrong-1510847867

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220.

Palatini, P., Dorigatti, F., Santonastaso, M., Cozzio, S., Biasion, T., Garavelli, G., Pessina, A. C., & Mos, L. (2007). Association Between Coffee Consumption and Risk of Hypertension. *Annals of Medicine, 39*(7), 545-553.

Patel, M. (2019). TinySearch -- Semantics based Search Engine using Bert Embeddings. *arXiv e-prints*, arXiv:1908.02451. Retrieved from https://ui.adsabs.harvard.edu/abs/2019arXiv190802451P

Periti, M., Salvaggio, A., Quaglia, G., & Di Marzio, L. (1987). Coffee consumption and blood pressure: an Italian study. *Clinical Science, 72*(4), 443-447.

Pew Research Center. (2019a). Internet/Broadband Fact Sheet. *pewresearch.org*. Retrieved from https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Pew Research Center. (2019b). Mobile Fact Sheet. *pewresearch.org*. Retrieved from https://www.pewresearch.org/internet/fact-sheet/mobile/

Recker, J., Reijers, H. A., & van de Wouw, S. G. (2014). Process model comprehension : the effects of cognitive abilities, learning style, and strategy. *Communications of the Association for Information Systems, 34*(9), 199-222.

Riding, R. J., & Sadler-Smith, E. (1997). Cognitive style and learning strategies: Some implications for training design. *International Journal of Training and Development, 1*(3), 199-208.

Roussinov, D., & Chau, M. (2008). Combining information seeking services into a meta supply chain of facts. *Journal of the Association for Information Systems, 9*(3), 175-199.

Sen, R., King, R. C., & Shaw, M. J. (2006). Buyers' Choice of Online Search Strategy and Its Managerial Implications. *Journal of Management Information Systems, 23*(1), 211-238.

Stensvold, I., Tverdal, A., & Per Foss, O. (1989). The effect of coffee on blood lipids and blood pressure. Results from a Norwegian cross-sectional study, men and women, 40-42 years. *Journal of Clinical Epidemiology, 42*(9), 877-884.

Tenner, E. (2018). Search Engines May Seem All-Knowing, But They're Not. Here's How to Get More Trustworthy Results. *TIME.com*. Retrieved from https://time.com/5318918/search-results-engine-google-bias-trusted-sources/

Tieger, T. (1980). On the biological basis of sex differences in aggression. *Child Development, 51*(4), 943-963.

Turner, C. W. (2001). Accountability demands and the auditor's evidence search strategy: The influence of reviewer preferences and the nature of the response (belief vs. action). *Journal of Accounting Research, 39*(3), 683-706.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology, 32*(1), 109-123.

Uiterwaal, C. S. P. M., Verschuren, W. M. M., Bueno-de-Mesquita, H. B., Ocké, M., Geleijnse, J. M., Boshuizen, H. C., Peeters, P. H. M., Feskens, E. J. M., & Grobbee, D. E. (2007). Coffee intake and incidence of hypertension. *The American Journal of Clinical Nutrition, 85*(3), 718-723.

Wakabayashi, K., Kono, S., Shinchi, K., Honjo, S., Todoroki, I., Sakurai, Y., Umeda, T., Imanishi, K., & Yoshizawa, N. (1998). Habitual coffee consumption and blood pressure: a study of self-defense officials in Japan. *European Journal of Epidemiology, 14*(7), 669-673.

Wang, H.-C., & Doong, H.-S. (2010). Online customers' cognitive differences and their impact on the success of recommendation agents. *Information & Management, 47*(2), 109-114.

White, R. (2013). *Beliefs and Biases in Web Search.* Paper presented at the Proceedings of the 36th international ACM SIGIR: Conference on Research and Development in Information Retrieval, Dublin, Ireland.

.

Appendix A: Development of the Abstracts

Using two databases (Google Scholar and PubMed), we searched for peer-reviewed studies that examined the relationship between coffee consumption and blood pressure. We identified four papers supporting the relationship between coffee consumption and hypertension (Burke et al., 1992; Jee, He, Whelton, Suh, & Klag, 1999; Palatini et al., 2007; Uiterwaal et al., 2007) and three supporting a relationship between coffee consumption and low blood pressure (Periti, Salvaggio, Quaglia, & Di Marzio, 1987; Stensvold, Tverdal, & Per Foss, 1989; Wakabayashi et al., 1998). Note that the abstracts of the papers supporting the link to hypertension did not include the phrase "low blood pressure," and the abstracts of papers supporting the link to low blood pressure did not include the term "hypertension" in them.

These studies were published in different journals over a span of 20 years and included samples from various countries and age groups, reported other disorders, and used inconsistent control variables. Our pilot studies revealed that these were a major source of confusion for participants, which confounded the experimental task. Therefore, we changed these abstracts (staying as faithful as possible to the original ones) by removing certain details about the study samples (such as ages and nationalities), making sure each abstract reported the same control variables, removing findings about other health problems unrelated to hypertension or low blood pressure, and changing paper details to make it look like they were published in the same year and in the same journal. We also created a fourth abstract supporting the relationship between coffee consumption and low blood pressure to balance the number of studies.

Further, we created another set of four abstracts using the abstracts that supported the link between coffee consumption and low blood pressure by replacing the phrase "low blood pressure" with "hypertension" and changing the tone of the corresponding sentences so that the new abstracts did not support the link between coffee consumption and hypertension. In the end, we had three sets of abstracts based on the original seven studies: 1) four abstracts indicating that there was a link between coffee consumption and hypertension; 2) four abstracts indicating that there was no link between coffee consumption and hypertension, and 3) four abstracts indicating that there was a link between coffee consumption and low blood pressure. To see whether the new abstracts were a good proxy for the original abstracts, we conducted a set of pretests, as follows.

Pretest 1

In Pretest 1, we focused on two of the newly created sets of abstracts: those indicating that there is a link between coffee consumption and hypertension and those indicating that there is no such link. Our goal was to see whether these two sets of abstracts were credible and whether their credibility was the same as or higher than that of the original abstracts. Using student participants, we measured (using a 7-point scale) participants' levels of confidence in the findings reported in each of these abstracts and compared them to levels of confidence in the original abstracts. To this end, six participants were given all four original abstracts on "coffee and hypertension," six participants were given all four new abstracts on "coffee and hypertension," and five participants were given all four new abstracts on "coffee and no hypertension."

The abstracts were provided to each participant in a random order to reduce the order effects. Comparison of mean confidence values revealed statistically significant differences ($F(2,14)=14.48$, $p<0.001$). The mean confidence value for the new abstracts about "coffee and hypertension" was higher than that for the original abstracts (5.75 vs. 4.42, respectively, $p<0.001$). In addition, the mean confidence value for the new abstracts about "coffee and no hypertension" was nearly the same as that for the new abstracts on "coffee and hypertension" (5.60 vs. 5.75, respectively, $p<0.60$).

There were no within-group differences among the new abstracts: a one-way ANOVA showed that the mean confidence value for each new abstract was not significantly different from the other means in the same set: $F(3,20)=1.23$, $p=0.33$ for abstracts on "coffee and hypertension" and $F(3,16)=1.67$, $p=0.21$ for abstracts on "coffee and no hypertension." Therefore, Pretest 1 showed that the new sets of abstracts were good proxies for the original abstracts.

Pretest 2

In Pretest 2, we focused on the credibility of the abstracts on coffee consumption and low blood pressure. We recruited another 13 student participants to measure their levels of confidence in the results reported

by the new abstracts. Six of the participants were given all three original abstracts, and seven were given all four new abstracts. We gave the abstracts to each participant in a random order to reduce the order effects.

The participants had a higher level of confidence for the new abstracts. The mean confidence values were 4.78 and 5.86 for the original and new abstracts, respectively ($F(1,11)=16.16$, $p=0.002$). Further, there were no within-group differences among the new abstracts; a one-way ANOVA showed that the mean confidence scores for the new eight abstracts were not significantly different from the other means ($F(3,24)=0.65$, $p=0.59$).

Pretest 3

In Pretest 3, we focused on the similarity between the original abstracts and the new abstracts on "coffee and low blood pressure." To this end, we conducted a study in a much smaller scale. We used a search engine that indexed the original seven abstracts, of which four were on "coffee and hypertension" and three were on "coffee and low blood pressure." Thirteen different student participants, who did not know the number or nature of the pages hosted on the server, took part in this pretest. Six participants were asked to test the validity of the link between coffee consumption and hypertension, and the other seven were asked to test the validity of the link between coffee consumption and low blood pressure. While participants submitted search queries to complete the experimental task, we captured their queries and the keywords used in each query. Later, we sent these queries to a search engine that indexed the eight new abstracts. Our goal was to compare the results obtained from the original abstracts with the results obtained from the new abstracts. If the queries returned the same set of pages, then the new abstracts would be considered similar to the original abstracts.

For the link between "coffee consumption and hypertension," all queries provided consistent results: if a specific query returned only confirming information from the original set of abstracts, it also returned only confirming information from the new set of abstracts. For the link between "coffee and low blood pressure," the new abstracts outperformed the original abstracts for certain queries. This was because one of the original abstracts reported an *inverse* relationship between "coffee consumption and blood pressure" and was not being included in the search results when participants used the phrase "low blood pressure" in their queries. On the other hand, the queries returned all four abstracts from the new set, providing better results. Overall, Pretest 3 showed that the new abstracts were good proxies for the original abstracts.

Appendix B

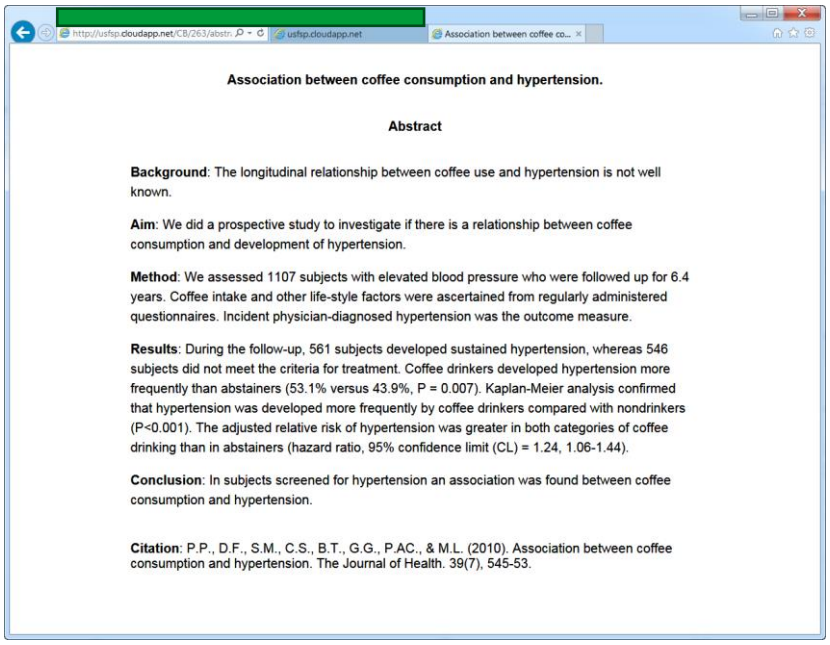

Figure B1. Sample Abstract

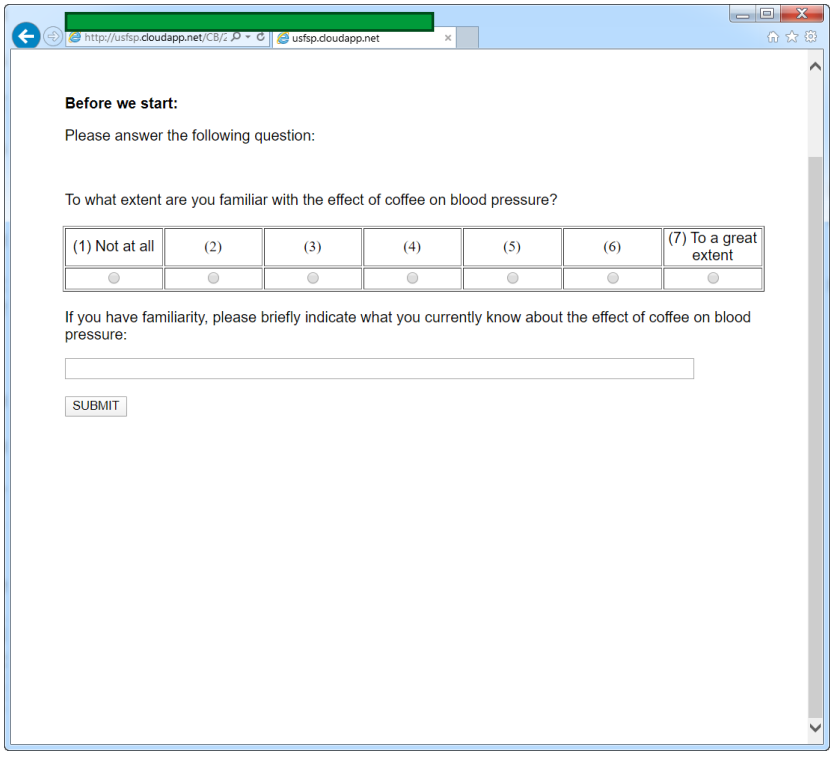

Figure B2. Initial Familiarity

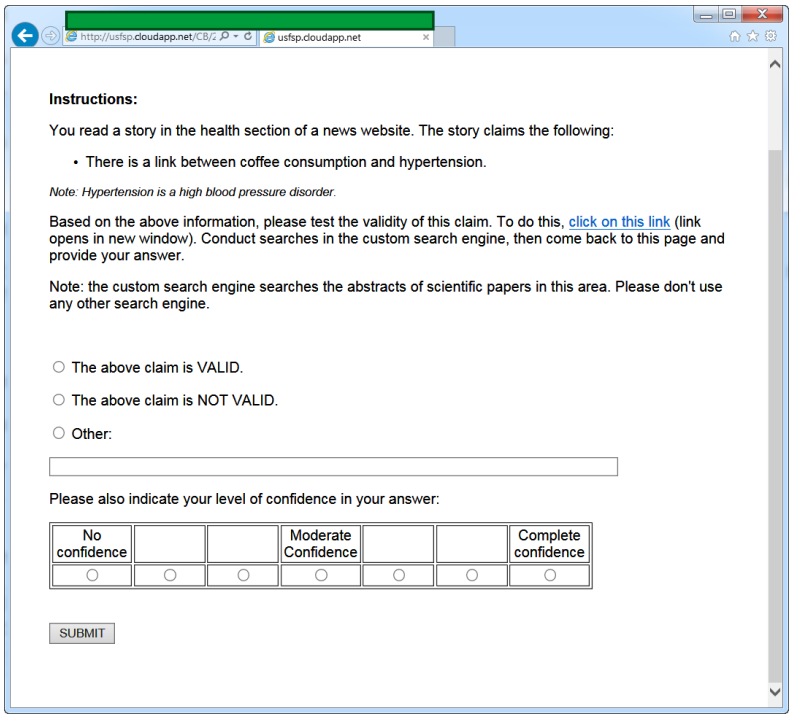

Figure B3. Instructions

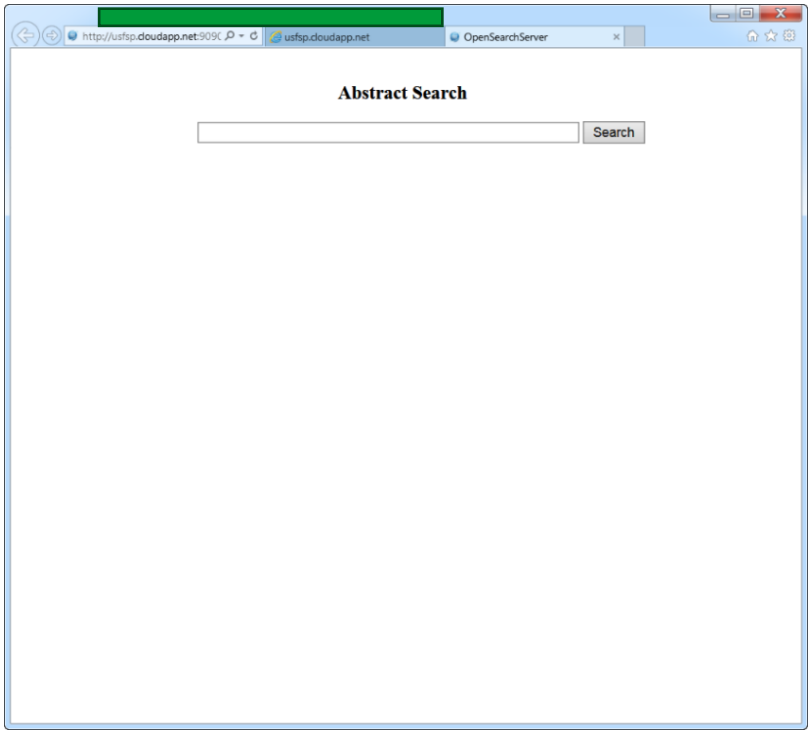

Figure B4. Search Page

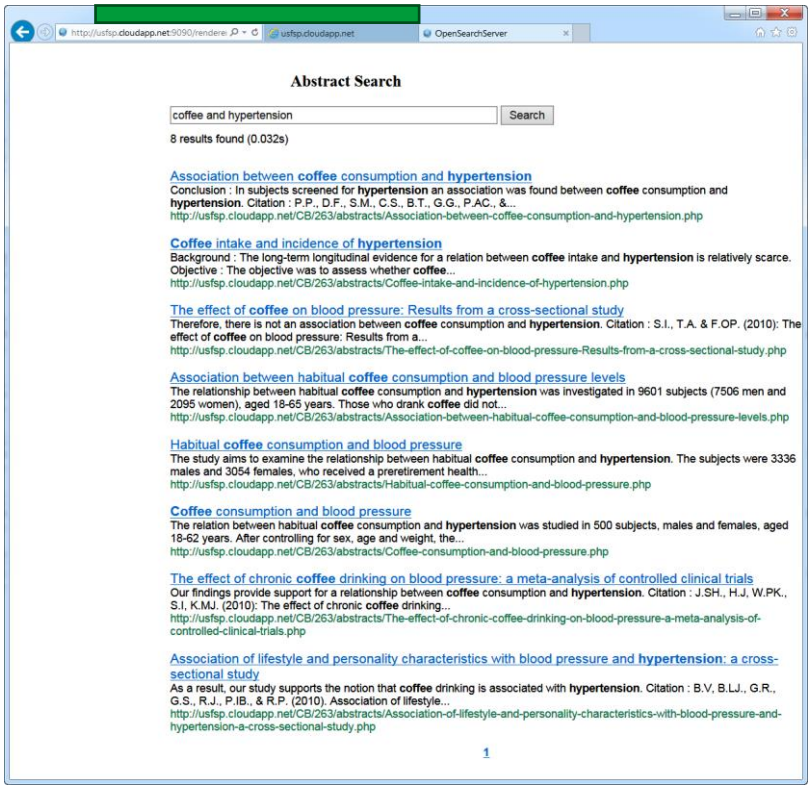

Figure B5. Presentation of Results

Appendix C

The Open Search Server software program allowed us to create the search engines used in this study. We created the search engines using the program's "web crawler" option. This option creates a search engine with default settings for an optimal search experience. However, our initial tests revealed that although this search engine worked as intended, it also searched English stop words (such as "a," "an," "the," "to," etc.) in its index if these words were used in queries. Therefore, we set the search engine to identify and eliminate English stop words in both queries and indexes by adding a new "filter" for stop words (called "StopFilter") to the search engine's English Text Analyzer.

The default settings provided a Google-like search experience. However, unlike Google searches, our search results did not include ads or sponsored links. Further, the "renderer" that generated the search results allowed us to insert scripts in the header and footer of the results to capture participants' online activities (such as the search queries they submitted and the links they clicked in search results).

The search engines were configured to index only the 38 pages discussed in the "Experimental Design" section of this paper. Autocomplete was disabled for the first experiment. To test the robustness of the findings, we repeated the experiment by enabling this feature of the search engine. When enabled, the list of words and phrases used in autocomplete was generated automatically by the search engine based on the patterns identified among the indexed pages.

Appendix D

Table D1. Participant Characteristics

	Autocomplete: On	Autocomplete: Off
Similar-term opposition	(N=37) Gender: Female 41% Male 59% Education: High school 16% Some college 41% Bachelor's 35% Master's 8% Doctorate 0% Age: 21-29 49% 30-39 30% 40-49 11% 50-59 8% 60+ 3% Time spent on the web: (hours) Less than 2 5% 2-4 19% 4-6 14% 6-8 24% 8+ 38%	(N=36) Gender: Female 39% Male 61% Education: High school 25% Some college 33% Bachelor's 36% Master's 3% Doctorate 3% Age: 21-29 50% 30-39 28% 40-49 8% 50-59 8% 60+ 6% Time spent on the web: (hours) Less than 2 3% 2-4 17% 4-6 25% 6-8 14% 8+ 42%
Dissimilar-term opposition	(N=44) Gender: Female 45% Male 55% Education: High school 16% Some college 39% Bachelor's 30% Master's 16% Age: 21-29 40% 30-39 36% 40-49 16% 50-59 4% 60+ 4% Time spent on the web: (hours) Less than 2 2% 2-4 25% 4-6 30% 6-8 7% 8+ 36%	(N=34) Gender: Female 53% Male 47% Education: High school 9% Some college 53% Bachelor's 32% Master's 6% Age: 21-29 46% 30-39 36% 40-49 13% 50-59 3% 60+ 3% Time spent on the web: (hours) Less than 2 6% 2-4 29% 4-6 6% 6-8 24% 8+ 35%

About the Authors

Varol Kayhan is an Associate Professor of Information Systems in the Muma College of Business at the University of South Florida. His research focuses on decision making, machine learning, and organizational knowledge management, and has appeared in *Communications of the ACM, Communications of the Association for Information Systems, Information & Management, Behavior Research Methods, Big Data*, *Journal of Computer Information Systems,* and others. He is the author of an eBook on data mining that is adopted as the formal course material by many universities. He teaches data analytics and machine learning courses in the undergraduate, graduate, and executive programs. He holds Ph.D. and MS degrees from the University of South Florida, and a BS degree from the Middle East Technical University in Turkey

Merrill Warkentin is a William L. Giles Distinguished Professor at Mississippi State University, where he serves as the James J. Rouse Endowed Professor of Information Systems in the College of Business. His research, primarily on the impacts of organizational, contextual, and dispositional influences on individual behaviors in the contexts of information security, privacy, and social media, has appeared in *MIS Quarterly, Journal of MIS, Journal of the Association for Information Systems, European Journal of Information Systems, Information Systems Journal, Decision Sciences, Information & Management,* and others. He is the author of over 110 peer-reviewed journal articles and the author or editor of seven books. Warkentin was the Program Co-Chair for the 2016 AMCIS Conference, and he serves (or has served) in editorial roles for *MIS Quarterly, Information Systems Research, Journal of the Association for Information Systems, Decision Sciences, European Journal of Information Systems, Information & Management,* and other journals. His work has been funded by NATO, NSF, NSA, DoD, Homeland Security, IBM, and others. He has been named an ACM Distinguished Scientist.