Received: 29 July 2022 | Revised: 31 January 2023 | Accepted: 7 March 2023
DOI: 10.1002/acp.4058

SPECIAL ISSUE ARTICLE

WILEY

# Understanding why searching the internet inflates confidence in explanatory ability

**Emmaline Drew Eliseev** | **Elizabeth J. Marsh**

Department of Psychology & Neuroscience, Duke University, Durham, North Carolina, USA

**Correspondence**
Elizabeth J. Marsh, Department of Psychology & Neuroscience, Duke University, Durham, NC 27708, USA.
Email: emarsh@psych.duke.edu

**Funding information**
Charles Lafitte Foundation Program for Research in Psychology and Neuroscience at Duke University; Google Faculty Research Award, Grant/Award Number: R2_255

## Abstract

People rely on the internet for easy access to information, setting up potential confusion about the boundaries between an individual's knowledge and the information they find online. Across four experiments, we replicated and extended past work showing that online searching inflates people's confidence in their knowledge. Participants who searched the internet for explanations rated their explanatory ability higher than participants who read but did not search for the same explanations. Two experiments showed that extraneous web page content (pictures) does not drive this effect. The last experiment modeled how search engines yield results; participants saw (but did not search for) a list of hits, which included "snippets" that previewed web page content, before reading the explanations. Participants in this condition were as confident as participants who searched online. Previewing hits primes to-be-read content, in a modern-day equivalent of Titchener's famous example of a brief glance eliciting false feelings of familiarity.

**KEYWORDS**
illusion of explanatory depth, internet, metacognition, online searching

## 1 | INTRODUCTION

People often rely on the internet for information, for example, Google processes more than 3.5 billion search queries a day (Internet Live Stats, 2021). The internet provides easy access to information well beyond what can be stored in an individual's memory—making it the ultimate transactive memory partner (see Marsh & Rajaram, 2019; Sparrow et al., 2011; Ward, 2013). Here, we focus on people's awareness of the boundaries between what they know (and have stored in their memories) and the information that they find online. Several recent demonstrations show that internet searching inflates people's estimates of what they know (e.g., Fisher et al., 2015), consistent with a larger literature showing people overestimate their knowledge in a variety of situations (Fischhoff et al., 1977; Sanchez & Dunning, 2018). We replicate and extend these findings, with the goal of understanding why internet searching leads to an illusion of explanatory ability.

Successful internet searching inflates people's estimates of what they know, regardless of whether people assess their abilities in general or their learning of specific facts. For example, people who used Google to answer 10 trivia questions gave themselves higher marks on a measure of Cognitive Self-Esteem (sample item: *I have a better memory than most people*) as compared to people forced to answer the questions without access to the internet (Ward, 2021). People who search the internet for explanations (e.g., why golf balls have dimples) later overestimate their ability to answer new questions about unrelated domains (i.e., weather; Fisher et al., 2015). In addition, participants who searched for information on topics like photosynthesis predicted they would do better on an upcoming fact-based multiple-choice quiz on that topic, as compared to participants who read but did not search for the information (Fisher et al., 2022). In all cases, the increased confidence in the search condition is considered an overestimate; there is no reason to believe that participants randomly assigned to the search condition are more knowledgeable about topics, like weather or photosynthesis, than participants in other conditions. In fact, the participants who searched for facts online actually performed worse on a multiple-choice test probing these facts than did participants who read the same information (Fisher et al., 2022).

  

To understand why searching inflates estimates of knowledge, we briefly review the mechanisms underlying other illusions of knowledge. For example, some work shows that people have expectations (i.e., schemas) about what they *should* know, leading people who claim expertise in a domain (e.g., finance) to purport knowledge of seemingly related concepts that do not exist (e.g., "pre-rated stocks"; Atir et al., 2015). And many illusions of knowledge likely reflect source monitoring errors, with information from some external source being misattributed to the one's internal knowledge. Specific examples include cryptomnesia (unconscious plagiarism; Brown & Murphy, 1989) and the false fame effect (mistakenly attributing prior exposure to a non-famous name as evidence of fame; Jacoby et al., 1989). Particularly relevant here is the finding that people overestimate their own abilities after receiving assistance from an external collaborator, like a human teammate or AI algorithm, so long as there is ambiguity about who is responsible for the team's success or failure (Fisher & Oppenheimer, 2021).

Source misattributions are part of the *community-of-knowledge hypothesis*, which posits that people do not make a sharp distinction between what they know and what others know. What differentiates the community-of-knowledge hypothesis from other source monitoring errors is the nature of the to-be-confused source: not only is it external, but it represents a larger group of people who share knowledge with the individual. In one study supporting this hypothesis, participants claimed better understanding of a fictional natural phenomenon (e.g., a newly discovered rock) when told that scientists fully understood it than when told that scientists had yet to explain or understand it (Sloman & Rabb, 2016). However, this effect disappeared when participants were told that the communal knowledge (i.e., the scientists' explanations) was classified and thereby inaccessible to them (Sloman & Rabb, 2016). Applying these ideas to the present question, the internet is almost always available, and thus people may not draw a strong boundary between their knowledge and the information available on the internet.

The internet's near-constant accessibility is not its only property likely to have consequences for people's metacognitive judgments. The internet also returns search results quickly, mimicking a well-established metacognitive cue. That is, well-known information normally comes to mind quickly, allowing retrieval speed to be interpreted as evidence of knowing (Benjamin & Bjork, 1996). But reliance on this heuristic is problematic when information is quickly retrieved for reasons other than stored knowledge. For example, prior exposure to a list of answers explicitly labeled as incorrect primes the later retrieval of those answers, and this expedited retrieval in turn increases confidence in those answers (Kelley & Lindsay, 1993). To the extent that internet searches return hits quickly, confidence will likely be inflated. Consistent with this claim, participants' search times are negatively correlated with their estimates that they will later remember the information (Stone & Storm, 2021). Critically, the effects of internet searching disappear when search speed is artificially slowed. When Google search results were delayed for 25 s, people who had used Google to answer 10 trivia questions no longer rated themselves higher on the measure of Cognitive Self-Esteem than did people forced to answer the questions without access to the internet (Exp 6; Ward, 2021). That is, once internet searching no longer mimicked the cue of fast retrieval, people's estimates of their own knowledge and abilities were more accurate.

Here, we explore another aspect of the internet that may contribute to inflated estimates of knowledge: the type of content that is returned in response to search queries. In many studies, subjects either study a question-answer pair or receive the question and then search the internet for the answer (e.g., Stone & Storm, 2021). But the participant in the internet condition is exposed to much more information while searching than the person who is simply told that "Sudan" is the answer to the question "What country borders Egypt to the south?" The internet searcher will see many hits, and the search results may include maps, images, and other information. Seeing photos, even uninformative ones, can increase people's self-perceived knowledge and their estimations of how well they understand scientific concepts; for example, seeing a picture of a rainbow increases belief in one's understanding of how rainbows form, despite gleaning no additional explanatory knowledge from the image itself (Cardwell et al., 2017). Consequently, it is possible that online searching inflates confidence because it exposes searchers to additional web page elements like photos. After testing (and rejecting) this hypothesis in two experiments, our last experiment models the way that internet search engines typically yield results; participants saw (but did not search for) a list of search hits, which included "snippets" that previewed the content of relevant web pages, before reading the explanations. Strikingly, participants in this condition were as confident as participants who searched online directly.

In all studies, we focus on people's estimates of their explanatory ability, as people are more likely to overestimate their ability to explain natural and mechanical processes (*illusion of explanatory depth*), as compared to their knowledge of facts, narratives, or procedures (Rozenblit & Keil, 2002). In Experiment 1, we replicate past findings showing that searching for explanations leads to higher estimates of explanatory ability. We extend past work by adding a final test to assess whether this inflated confidence is justified or not. To date, few studies have measured downstream consequences for learning following internet searches. In the relevant studies that we know of, searching for facts increased confidence but was associated with worse performance on a multiple-choice test probing these facts (Fisher et al., 2022). However, that study was focused on recognition of facts, a much simpler task than producing detailed explanations. We include a final explanation test as a measure of participants' learning in all of our experiments; however, because the critical comparison (read vs. search) was only significant in two of the four experiments, we conducted a mini meta-analysis and discuss the accuracy results together in the general discussion.

In sum, we conducted four experiments examining the effects of internet searching on estimates of explanatory ability and on learning. After establishing the basic effect in Experiment 1, two experiments investigated the role of pictorial content, and the last experiment explored previewing content via search result snippets. All studies contained a measure of learning to compare participants' confidence in their explanatory ability with the accuracy of the explanations they generated at test.

## 2 | EXPERIMENT 1

Experiment 1 was designed to replicate the basic effect, namely that internet searching leads to higher estimates of explanatory ability than does reading the same explanations. Experiment 1 extended past work by asking participants to produce the explanations, in order to evaluate whether the higher estimates in the searching condition are justified or not.

### 2.1 | Method

#### 2.1.1 | Participants

Three-hundred and twelve Amazon Mechanical Turk workers located in the United States with an approval rating above 90% participated online for compensation. The sample size was determined based on a power analysis that suggested a sample size of 303 would be sufficient to detect a small to medium sized effect with a power of 80%. Seventeen participants were excluded for not following instructions (i.e., they reported using the internet to look up answers even though they were instructed not to), so data were analyzed with the remaining 295 participants (144 women, 150 men, 1 non-binary; *M* age = 34.94 years). We also excluded 63 individual responses (3.05% of the total) where the wrong URL was provided,[1] because we could not confirm that these participants had read the same explanation materials as others on these specific trials. This exclusion criterion was applied to both the analyses of confidence ratings and explanation accuracy scores.

#### 2.1.2 | Design

This experiment had three between-subject conditions: control, read, search.

#### 2.1.3 | Materials

We selected seven explanatory knowledge questions (e.g., Why are there leap years?) from Fisher et al. (2015).[2] The questions targeted familiar concepts (e.g., zippers, moon phases), so that most participants could offer a partial explanation without looking up the complete answer (see Appendix A for the full set of questions). The order of questions was randomized in all conditions.

For each question, we selected a specific article from a reputable website (e.g., history.com, scientificamerican.com) and tasked participants in the search condition to find these target articles. Each web page appeared at or near the top of Google search results and explained the answer to the question fully. To create explanation texts for participants in the read condition, we copied and pasted the text (minus phrases like "click here to learn more") from the same target websites that participants in the search condition were asked to find. Thus, participants in the read and search conditions read the same explanations.

#### 2.1.4 | Procedure

After giving informed consent, participants completed the *rating phase* where they were asked to rate how well they could explain the answer for each of the seven questions on a scale of 1 (*very poorly*) to 7 (*very well*). Before rating their explanatory ability for each question, participants in the search condition were asked to conduct a Google search in a separate browser window to find the target article from a pre-specified website (e.g., history.com) to confirm the details of the explanation to the question (e.g., why are there leap years?). In order to confirm that participants found the correct article, they were asked to copy and paste the URL of the specific web page they found.[3] Participants in the read condition were asked to read a text before rating their explanatory ability for each item. These explanation texts were copied and pasted from the same target websites that participants in the search condition were asked to find. Control participants were simply asked to evaluate how well they could explain the answer to each question without using any outside sources—just their own knowledge.

After making ratings of explanatory ability for all seven questions, participants completed the *explanation phase*. Participants were asked to explain the answer to each question they saw in the previous *rating phase*. They were instructed to provide as much detail as possible without using any outside sources. After completing the *explanation phase*, participants were asked to report if they used the internet to look up the answers to any questions in either the *rating phase* or the *explanation phase*.

### 2.2 | Results

The alpha level was set at .05 for all statistical tests.

#### 2.2.1 | Metacognitive judgments of explanatory ability

We conducted a one-way between subjects ANOVA on participants' ratings of explanatory ability by condition (control, read, search) collapsing across individual questions. The relevant data appear in Figure 1. Homogeneity of variance was violated, as assessed by Levene's Test of Homogeneity of Variance ($p = .007$). Confidence in explanatory ability significantly differed by condition, Welch's $F(2, 191.83) = 65.75$, $p < .001$. Participants in the search condition ($M = 5.45$, $SD = 0.99$) gave significantly higher ratings of explanatory

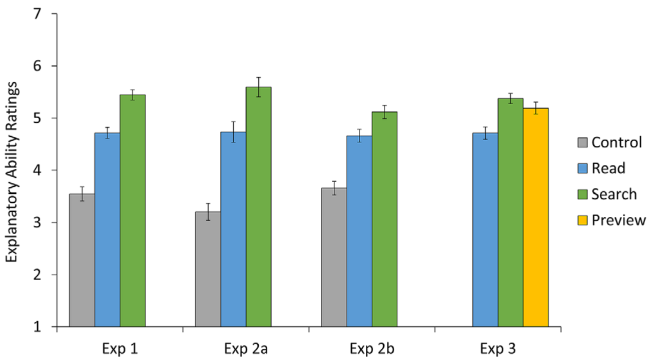


**FIGURE 1** Mean ratings of explanatory ability by condition for Experiments 1, 2a, 2b, and 3. Error bars reflect standard error of the mean. Note that Experiment 3 included a preview condition and no control condition.

ability than participants in the read condition ($M = 4.71$, SD $= 1.07$, $t(196) = 5.00$, $p < .001$, $d = 0.71$) and those in the control condition ($M = 3.55$, SD $= 1.30$, $t(179.12) = 11.48$, $p < .001$, $d = 1.65$). Participants in the read condition also gave higher ratings of explanatory ability than those in the control condition, $t(185.39) = 6.86$ $p < .001$, $d = 0.98$.

### 2.2.2 | Learning: Explanation accuracy

Two trained, independent coders scored participants' explanation responses for accuracy and completeness on a 3-point scale: 0 (no credit), 1 (half credit), 2 (full credit). Interrater reliability was good, $\kappa = .85$, and all scoring discrepancies were resolved through discussion. An accuracy score (0 to 1) was calculated for each participant based on the proportion of total possible points (14) they received across their responses to the seven questions. Homogeneity of variance was violated, as assessed by Levene's Test of Homogeneity of Variance ($p < .001$). A one-way between-subjects ANOVA on participants' accuracy scores revealed accuracy significantly differed by condition, Welch's $F(2, 190.45) = 70.37$, $p < .001$. Participants in the read condition ($M = 0.59$, SD $= 0.20$) produced more accurate responses than participants in the search condition ($M = 0.42$, SD $= 0.23$, $t(196) = 5.47$, $p < .001$, $d = 0.78$). Accuracy scores in the control condition ($M = 0.28$, SD $= 0.16$) were significantly lower than the read condition, $t(184.77) = -11.86$, $p < .001$, $d = -1.69$, as well as the search condition, $t(173.48) = -4.91$, $p < .001$, $d = -0.70$.

## 3 | EXPERIMENTS 2a AND 2b

In Experiment 1, participants in the read condition received the text that was copied and pasted from the original online article. Participants in the search condition, however, were exposed to target web pages that contained more content than just text; they, like most web pages on the internet, included photos (both relevant and not), links to other websites, colorful headers and footers, and eye-catching ads. As photos, even uninformative ones, can increase people's self-perceived knowledge and their estimations of understanding complex concepts, like scientific processes (Cardwell et al., 2017), it is possible that the search condition's confidence ratings in Experiment 1 were inflated by exposure to these additional web page elements. Additionally, when reading an online article, the target information is not typically the only content on web page; rather, it appears in the context of other relevant information that allows for potential elaboration of the target information.

To eliminate these confounds in Experiments 2a and 2b, participants in the read condition were asked to read full-page screenshots of the target web pages, so they would be exposed to the same explanation article and many of the same web page elements as participants in the search condition (with the exception of differences in personalized ads or other customizations based on an individual's browsing history and settings). Experiments 2a and 2b were identical; 2b was a replication with a larger sample size.

### 3.1 | Method

#### 3.1.1 | Participants

In both studies, participation was limited to Amazon Mechanical Turk workers located in the U.S. with an approval rating above 90% who participated online for compensation. One-hundred and nineteen subjects participated in Experiment 2a and 288 participated in Experiment 2b. The sample size for Experiment 2a was calculated with a power analysis based on the results of Experiment 1. The power analysis suggested that a sample size of 93 would be sufficient to provide 80% power to replicate the medium to large sized effect of the search-induced confidence boost observed in Experiment 1. In order to account for excluded participants, we collected data until we had a usable sample size over 93. With hindsight, we suspected that Experiment 2a might have been underpowered to detect a difference in explanation accuracy as observed in Experiment 1 between the search and read conditions, so we ran Experiment 2b as a higher-powered replication. We determined the sample size for Experiment 2b based on a power analysis to detect a medium sized effect with 95% power, which indicated a minimum sample size of 252. Again, anticipating exclusions, we collected data until we had a usable sample size over 252. In Experiment 2a, 25 participants were excluded for not following instructions on more than half of the questions (i.e., 16 participants reported looking up answers with an unauthorized outside source, 9 participants gave non-compliant responses that suggested the questions were not read[4]), so data were analyzed with the remaining 94 participants (39 women, 53 men, 1 nonbinary, 1 unreported; $M$ age $= 35.95$ years). We excluded 20 individual ratings (3.04% of the total) from our analyses in cases where we could not confirm that participants had read the same explanation materials as others (i.e., 17 cases in which the correct URL was not reported, three cases in which the participant reported they did not follow their condition's instructions by using an unauthorized source). In Experiment 2b, 35 participants were excluded for not following instructions on more than half of the questions (i.e., 29 participants reported looking up answers with an unauthorized outside source, 6 participants gave non-compliant responses that suggested the questions were not read), leaving 253 participants in the analysis (113 women, 136 men, 2 nonbinary, 2 unreported; *M* age = 35.11 years). As before, we excluded 34 individual ratings (1.92% of the total) from our analyses in Experiment 2b in cases where we could not verify that participants had read the correct explanation materials (i.e., 26 cases in which the correct URL was not reported, 8 cases in which the participant reported they did not follow their condition's instructions by using an unauthorized source).

### 3.1.2 | Design

Both experiments had the same design as Experiment 1 with 3 between-subject conditions: control, read, search.

### 3.1.3 | Materials

Both experiments used the same seven explanatory knowledge questions and the same pre-specified explanation websites as in Experiment 1. In order to better match the content that participants saw in the read condition to the content seen in the search condition, we used full-page screenshots of the web pages that participants in the search condition were asked to find as the explanation texts in the read condition. Thus, participants in the read and search conditions were exposed to the same web page layout, images, and other potentially distracting items when reading the explanation article.

### 3.1.4 | Procedure

The procedure in both studies was similar to that of Experiment 1, with the exceptions that (1) participants were asked to report using an unauthorized source to look up answers for each question individually in both the *rating phase* and the *explanation phase* and (2) participants in the read condition were given full-length screenshots of the explanation article web pages to read (instead of the plain explanation text copied from the target websites) in order to better match the explanation content that participants in the search condition find online. Before moving to the *explanation phase*, participants in the search condition were asked to close any other browser windows they had opened to look up answers in the *rating phase*.

## 3.2 | Results

The alpha level was set at .05 for all statistical tests.

### 3.2.1 | Metacognitive judgments of explanatory ability

For each study, we conducted a one-way between subjects ANOVA on participant's ratings of explanatory ability by condition (control, read, search) collapsing across the seven individual questions. See Figure 1 for the relevant data. Confidence in explanatory ability differed significantly across conditions [Exp 2a: $F(2, 91) = 44.28$, $p < .001$; Exp 2b: $F(2, 250) = 35.24$, $p < .001$]. Participants in the search condition (Exp 2a: $M = 5.60$, SD = 1.05; Exp 2b: $M = 5.12$, SD = 1.17) rated their explanatory ability higher than did participants in the read condition (Exp 2a: $M = 4.73$, SD = 1.09; Exp 2b: $M = 4.66$, SD = 1.12) [Exp 2a: $t(60) = 3.17$, $p = .002$, $d = 0.80$; Exp 2b: $t(165) = 2.58$, $p = .011$, $d = 0.40$] and those in the control condition (Exp 2a: $M = 3.20$, SD = 0.94; Exp 2b: $M = 3.66$, SD = 1.20) [Exp 2a: $t(62) = 9.59$, $p < .001$, $d = 2.40$; Exp 2b: $t(169) = 8.06$, $p < .001$, $d = 1.23$]. Participants in the read condition rated their explanatory ability higher than did those in the control condition [Exp 2a: $t(60) = 5.92$ $p < .001$, $d = 1.51$; Exp 2b: $t(166) = 5.60$, $p < .001$, $d = 0.87$].

### 3.2.2 | Learning: Explanation accuracy

Two trained, independent coders scored participants' explanations on the final test for accuracy and completeness on a 3-point scale: 0 (no credit), 1 (half credit), 2 (full credit). Interrater reliability was good (Exp 2a: $\kappa = .75$; Exp 2b: $\kappa = .79$), and all scoring discrepancies were resolved through discussion. An accuracy score (0 to 1) was calculated for each participant based on the proportion of total possible points (14) they received across their responses to the seven questions. Homogeneity of variance was violated in Experiment 2a and Experiment 2b, as assessed by Levene's Test of Homogeneity of Variance [Exp 2a: $p = .002$; Exp 2b: $p = .026$]. A one-way between-subjects ANOVA on participants' accuracy scores revealed that explanation accuracy differed significantly by condition [Exp 2a: Welch's $F(2, 55.21) = 7.29$, $p = .002$; Exp 2b: Welch's $F(2, 162.32) = 25.89$, $p < .001$]. Participants in the control condition (Exp 2a: $M = 0.29$, SD = 0.15; Exp 2b: $M = 0.27$, SD = 0.18), who did not have access to any external explanations, produced less accurate explanations than either participants in the read condition (Exp 2a: $M = 0.49$, SD = 0.26; Exp 2b: $M = 0.49$, SD = 0.25) [Exp 2a: $t(44.79) = -3.65$, $p < .001$, $d = -0.94$; Exp 2b: $t(146.06) = -6.72$, $p < .001$, $d = -1.05$] or participants in the search condition (Exp 2a: $M = 0.40$, SD = 0.25; Exp 2b: $M = 0.41$, SD = 0.21) [Exp 2a: $t(49.97) = -2.03$, $p = .047$, $d = -0.51$; Exp 2b: $t(169) = -4.90$, $p < .001$, $d = -0.75$]. While accuracy scores did not differ significantly between the read and search conditions in Experiment 2a [$t(60) = 1.45$, $p = .151$], accuracy scores in the read condition were significantly higher than accuracy scores in the search condition in Experiment 2b [$t(165) = 2.22$, $p = .028$, $d = 0.34$].

## 4 | EXPERIMENT 3

Experiments 2a and 2b replicated the key finding that participants in the search condition rated their explanatory ability higher than did participants in the read condition, who were provided with the same explanations. This effect held even though participants in the read condition viewed a screenshot of the target web page, meaning that they, like the search condition participants, saw all the images and other extraneous content that appear on a typical web page.

One possible explanation for this result is that participants interpret their active engagement in the search process (i.e., generating a search query and selecting a result) as evidence that they have learned the explanations well. But web searching (with a popular search engine like Google) also returns results in a way that differs from simply being given an explanation to read. A search engine's algorithms identify additional information that is related to the search query, including a ranked list of relevant web pages that shows titles and short descriptions for each hit called "snippets" that are "designed to emphasize and preview the page content that best relates to a user's specific search" (Create Good Titles and Snippets in Search Results, n.d.). Many search queries also yield "featured snippets" that appear at the top of the search results page. These featured snippets (aka answer boxes) quickly highlight the answer to a user's search query by automatically pulling relevant text from web search listings to a featured box at the top of the search results page (How Google's Featured Snippets Work—Google Search Help, n.d.). As a result, searches yielding featured snippets are more likely to result in no-click searches, where the searcher does not click on any of the links, than search results that do not include a featured snippet (Fishkin, 2018; Soulo, 2017). Critically, each of the seven explanatory questions used in the present experiments yielded a featured snippet at the top of the search results page.

Effectively, a featured snippet is a preview of the to-be-viewed information—and in that sense, potentially primes the to-be-learned content. In Experiment 3, we investigated this hypothesis, adding a new preview condition where participants did not actively search the internet but were exposed to the first page of Google search results before reading the full explanation article. Unlike participants in the search condition, who were instructed to actively engage in the search process (i.e., typing the question in the search bar, initiating the search, looking through the list of search hits to locate the target article from a specific website, and clicking on the target search result), participants in the preview condition simply saw the Google search results page before reading an article that explained the answer. In other words, participants in the preview condition saw the featured snippet, as well as the snippets accompanying individual search results, prior to reading the given explanation, but they were not asked to locate a target article from a specific website to find the explanation.

### 4.1 | Method

#### 4.1.1 | Participants

Two-hundred and ninety-seven Amazon Mechanical Turk workers located in the U.S. with an approval rating above 90% participated online for compensation. The sample size for Experiment 3 was determined by a power analysis using the same criteria as Experiment 2b, which indicated a sample size of 252 was sufficient to detect a medium sized effect with 95% power. We chose to collect data from more participants than in Experiment 2b to account for potential exclusions. Thirty-six participants were excluded entirely for not following instructions on more than half of the questions (i.e., 15 participants reported looking up answers with an unauthorized outside source, 21 participants did not report the correct URL of the target website), so data were analyzed with the remaining 261 participants (111 women, 149 men, 1 nonbinary; *M* age = 36.43 years). We also excluded 49 individual ratings (2.68% of the total) from our analyses, as we could not confirm that these participants had read the correct explanation materials for those questions (i.e., 37 cases in which the correct URL was not reported, 12 cases in which the participant reported they did not follow their condition's instructions by using an unauthorized source).

#### 4.1.2 | Design

This experiment had 3 possible between-subject conditions: preview, read, search.

#### 4.1.3 | Materials

We used the same seven explanatory knowledge questions, prespecified explanation websites, and full-page screenshots as in Experiments 2a and 2b. Unlike the previous experiments, a preview condition replaced the control condition. For the new preview condition, we took full-page screenshots of the first page of Google search results for each explanatory question, which included the explanatory knowledge question (e.g., why are there leap years) typed out in the Google search bar. Thus, participants in the preview condition were exposed to similar search results as participants in the search condition, including web page titles and snippets (i.e., short descriptions relevant to the search query), for all search results on the front page of Google.

#### 4.1.4 | Procedure

The procedure was the same as the prior experiments for participants in the read and search conditions. Participants in the preview condition were shown a full-page screenshot of the first page of Google search results and given a full-page screenshot of the explanation web page to read (as in the read condition) before rating their ability to explain the answer to this question from 1 (*very poorly*) to 7 (*very well*).

## 4.2 | Results

The alpha level was set at .05 for all statistical tests.

### 4.2.1 | Metacognitive judgments of explanatory ability

We conducted a one-way between-subjects ANOVA on participant's ratings of explanatory ability by condition (preview, read, search) collapsing across individual questions. The relevant data appear in Figure 1. Confidence in explanatory ability significantly differed by condition, $F(2, 258) = 9.42$, $p < .001$. Participants in the search condition ($M = 5.38$, SD $= 0.92$) gave significantly higher ratings of explanatory ability than participants in the read condition ($M = 4.71$, SD $= 1.10$, $t(171) = 4.29$, $p < .001$, $d = 0.65$). Participants in the preview condition ($M = 5.19$, SD $= 1.07$) also gave higher ratings of explanatory ability than those in the read condition, $t(171) = 2.89$, $p = .004$, $d = 0.44$. Critically, ratings of explanatory ability did not significantly differ between the search and preview conditions, $t(174) = 1.23$, $p = .221$.

### 4.2.2 | Learning: Explanation accuracy

Two trained, independent coders scored participants' explanation responses for accuracy and completeness on a scale of 0 (no credit) to 2 (full credit). Interrater reliability was good, $\kappa = .89$, and all scoring discrepancies were resolved through discussion. An accuracy score was calculated for each participant based on the proportion of total possible points they received, averaged across their responses to the seven questions. A one-way between-subjects ANOVA on participants' accuracy scores revealed no significant differences between conditions, $F(2, 258) = 0.17$, $p = .844$ [preview condition: $M = 0.52$, SD $= 0.24$, read condition: $M = 0.54$, SD $= 0.25$, search condition: $M = 0.51$, SD $= 0.24$].

# 5 | GENERAL DISCUSSION

Across four experiments, people who searched the internet for explanations rated their explanatory ability as higher than people who read but did not search for the same explanations. This searching effect was not driven by the pictorial content that typifies web pages, even though pictures increase belief in other contexts (e.g., Cardwell et al., 2017). Additionally, this boost in confidence was not specific to active searching (e.g., Fisher et al., 2015), as participants who previewed the search results (without personally engaging in an active search process), were just as confident as participants who searched the internet themselves.

Our results are not consistent with a simple source error, whereby searching the internet inflates confidence when people confuse externally available knowledge with their own internal knowledge (e.g., Ward, 2021), as participants in both the read and search conditions were exposed to the same external explanation articles. Moreover, it is ambiguous who deserves credit for the explanatory knowledge in the search, preview, and read conditions: Participants could have attributed their knowledge of the explanations to their own memories or to the articles they read, meaning that this ambiguity cannot fully explain why participants in the search and preview conditions were consistently more confident than participants in the read condition (e.g., Fisher & Oppenheimer, 2021).

One question is whether the presence of featured snippets (seen in the search conditions and the preview condition in Exp 3) lent additional credibility to the explanations. That is, a featured snippet may give participants the impression that the explanation is well-understood and that there is consensus about the answer among the experts. This is true in some cases—Google, for example, has consulted with medical professionals to craft featured snippets of important medical information as a public service (Medical Information on Google—Google Search Help, n.d.). While we have no direct data about how our subjects perceived the credibility of the snippets, to the extent they interpreted them as evidence of available community knowledge, then the results would be consistent with the community-of-knowledge hypothesis (Sloman & Rabb, 2016).

While we acknowledge that possibility, our own preference is for a more parsimonious explanation: namely, the results of the last experiment suggest that priming contributes to the increased illusion of explanatory depth observed after searching. In the last experiment, participants showed a similar illusion when they saw a screenshot of the search results (including snippets) prior to reading the explanations (Exp 3, preview condition). By previewing the to-be-read content, search result snippets act like a modern-day equivalent of Edward B. Titchener's famous example of a person who glances both ways before crossing the street, gets distracted, and then feels as if they have already crossed the street (Titchener, 1921). That is, a brief exposure to information makes it easier to process, and such fluency is known to be interpreted in different ways. In recognition memory experiments, a new, unstudied word is more likely to be called "old" when it was briefly primed with itself (Jacoby & Whitehouse, 1989). A parallel finding occurs when people are deciding if they have experience with nonsense symbols, which are judged in the context of familiar symbols from everyday life (e.g., Nike Swoosh, BMW icon). A brief exposure to the nonsense symbol increases people's belief that it had been seen previously, outside of the experimental context, as compared to a brief glance at a different symbol or no symbol at all (Brown & Marsh, 2009). In the present work, a brief glance at upcoming web page content is enough to increase one's estimates of understanding. Increased familiarity with web page content (via a snippet) may have downstream consequences. For example, this priming may explain why in some other studies participants who searched for

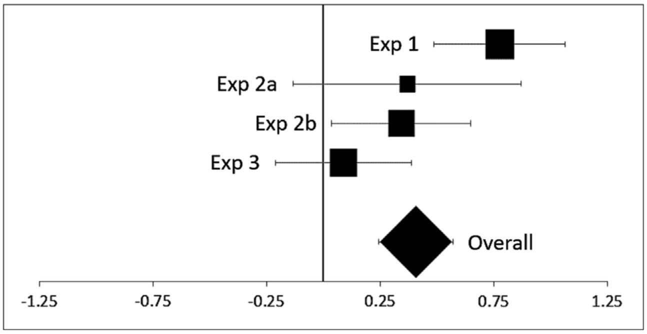


**FIGURE 2** Explanation accuracy effect sizes: Read versus search conditions. The forest plot shows the effect sizes (Cohen's *d*) for the effect of condition (read vs. search) on explanation accuracy. The size of the square for each experiment represents the weighting according to sample size, and the error bar line width represents the 95% confidence interval. The diamond represents the overall effect size from the mini meta-analysis.

information online spent less time reading articles than participants in no-search conditions (e.g., Fisher et al., 2022). Going forward, more research is needed to understand the individual (or possibly combined) contributions of fluency and beliefs about a community-of-knowledge when judging one's explanatory abilities.

Finally, we discuss the learning results across all four experiments. While the effect on confidence was clear across studies, the explanation accuracy results were less strong, with two experiments showing the effect of interest (with higher performance in the read condition than the search condition) and two experiments showing non-significant differences between the read and search conditions (albeit in the right direction). To gain traction on this issue, we performed a mini meta-analysis across the four experiments comparing explanation accuracy between the read and search conditions (see Figure 2). We used a fixed effects model in which the mean effect size (i.e., mean Cohen's *d*) was weighted by sample size (Goh et al., 2016). The mini meta-analysis confirmed that, across the four experiments, there was a significant difference in explanation accuracy between the read and search conditions, with participants in the read condition producing more accurate explanations than participants in the search condition, $Z = 4.92$, $p < .001$, $d = 0.41$, 95% CI [0.25, 0.57]. These results suggest that actively searching for explanations online led to less learning than simply reading the same explanations. As we did not find any evidence to suggest that online searching benefits explanatory learning, the observed inflation of confidence in explanatory ability reflects an illusion of explanatory depth (e.g., Rozenblit & Keil, 2002). Although online searching made people feel more confident in their explanatory knowledge, it did not translate into higher quality, more accurate explanations at test. Instead, online searching was associated with less accurate explanations overall. While our learning results varied across individual experiments, the overall finding of the mini meta-analysis aligns with similar search-related impairments that were observed on a fact-based multiple-choice test (Fisher et al., 2022).

Why might online searching be associated with less learning? The problem most likely occurs at encoding, as it is not clear why online searching would affect later retrieval of stored information. One possibility is that online searching emphasizes how easy it is to find the information online thus negating the need to memorize it (offloading hypothesis; Fisher et al., 2022). A second possibility is that people set up simplified representations based on the brief answers presented in the search result snippets, and these representations scaffold more superficial encoding of the actual web page. A third possibility, in line with our fluency hypothesis, is that people exert less effort to encode the to-be-learned information after searching (or previewing the search results), because the answer already feels familiar or fluent after seeing shorter versions of the answer in the search result snippets. This explanation is similar to that used to explain why re-reading is not an effective study strategy: Perceptual priming means that the text is easier to read the second time, leading the learner to be inappropriately confident in their learning (e.g., Bjork et al., 2013). All three of these explanations suggest that information in the search condition is not well encoded, leading to worse learning outcomes. However, we argue that impaired encoding (as opposed to voluntarily offloading) offers a more probable explanation for an intriguing finding from Fisher et al. (2022): namely, that clicking on a direct link (without searching) does not impair learning in the same way that searching the internet does (Exp 4; Fisher et al., 2022). While the authors argue that clicking on a link does not emphasize the availability of the information online, we believe instead the key difference is that clicking on a direct link skips exposure to the snippets on a search results page. Our data cannot separate whether snippets lead to simplified representations (that guide later encoding) or prime web page content (thus leading to an illusion of knowledge that reduces effort spent encoding), but the confidence data are in line with many fluency-based illusions in the literature. More research is needed to thoroughly tease these possibilities apart.

In sum, our study replicates past work showing that internet searching inflates confidence, eliminates three possible explanations for this finding (i.e., pictures, active searching, source errors) and highlights two possible mechanisms that may contribute to this effect (i.e., that featured snippets may prime and/or instill credibility in the to-be-read information). These results have implications for the design of online environments; they emphasize how different aspects of the internet may combine to affect metacognitive monitoring. Specifically, a common feature of the online search environment (e.g., featured snippets) may increase users' confidence in their learning. Internet searching changes more than just the speed with which information is retrieved; it also changes the content people see and the type of processing and encoding they engage in.

## ACKNOWLEDGMENTS

This research was supported by a Google Faculty Research Award (EJM) and the Charles Lafitte Foundation Program for Research in Psychology & Neuroscience at Duke University (EDE). The authors would like to thank Abigail C. Flyer, Stephanie Ng, and Camila Vargas Restrepo for their assistance with data collection and coding.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data from this study are openly available on the Open Science Framework at: https://osf.io/j2fm5/?view_only=875df83bcf764857868f032cb2654de6.

## ORCID

*Emmaline Drew Eliseev* https://orcid.org/0000-0002-1043-6655
*Elizabeth J. Marsh* https://orcid.org/0000-0003-0456-9547

## ENDNOTES

[1] In subsequent experiments, we chose to exclude participants in entirety if they did not follow instructions (i.e., they reported the incorrect URL, reported looking up answers using unauthorized outside sources, gave non-compliant responses) for more than half (>3) of the seven questions, as this indicates an overall lack of attention to the task instructions.

[2] One question out of the original 8 questions from Fisher et al. (2015) was not used, as a reliable answer was difficult to find online.

[3] In cases where participants reported the incorrect URL, we excluded their confidence rating and explanatory response for that specific question from our analysis.

[4] These non-compliant responses most commonly included: restating the question or instructions as the answer, responding to all questions with generic one-word responses (e.g., "yes", "n/a", "good", or "ok"), and responding to all questions with letters (e.g., "j" or "nn"). Note that variations of "I do not know" responses were included in the accuracy score, as this is a valid response to the question.

[5] We changed the question wording from "How do zippers work" to "How zippers work" to facilitate finding consistent search results on the designated website (science.howstuffworks.com) for this question.

## REFERENCES

Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science*, *26*(8), 1295–1303. https://doi.org/10.1177/0956797615588195

Benjamin, A. S., & Bjork, R. A. (1996). *Retrieval fluency as a metacognitive index. Implicit memory and metacognition*. Psychology Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Brown, A. S., & Marsh, E. J. (2009). Creating illusions of past encounter through brief exposure. *Psychological Science*, *20*(5), 534–538. https://doi.org/10.1111/j.1467-9280.2009.02337.x

Brown, A. S., & Murphy, D. R. (1989). Cryptomnesia: Delineating inadvertent plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 432–442. https://doi.org/10.1037/0278-7393.15.3.432

Cardwell, B. A., Lindsay, D. S., Förster, K., & Garry, M. (2017). Uninformative photos can increase People's perceived knowledge of complicated processes. *Journal of Applied Research in Memory and Cognition*, *6*(3), 244–252. https://doi.org/10.1016/j.jarmac.2017.05.002

Create good titles and snippets in Search Results. (n.d.). Google Developers. Retrieved March 9, 2021, from https://developers.google.com/search/docs/advanced/appearance/good-titles-snippets

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Fisher, M., Goddu, M. K., & Keil, F. C. (2015). Searching for explanations: How the internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General*, *144*(3), 674–687. https://doi.org/10.1037/xge0000070

Fisher, M., & Oppenheimer, D. M. (2021). Who knows what? Knowledge misattribution in the division of cognitive labor. *Journal of Experimental Psychology: Applied*, *27*(2), 292–306. https://doi.org/10.1037/xap0000310

Fisher, M., Smiley, A. H., & Grillo, T. L. H. (2022). Information without knowledge: The effects of internet search on learning. *Memory*, *30*(4), 375–387. https://doi.org/10.1080/09658211.2021.1882501

Fishkin, R. (2018, October 30). *Google CTR in 2018: Paid, organic, & no-click searches*. SparkToro https://sparktoro.com/blog/google-ctr-in-2018-paid-organic-no-click-searches/

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Google Search Statistics—Internet Live Stats. (2021). Retrieved March 20, 2021, from https://www.internetlivestats.com/google-search-statistics/

How Google's featured snippets work—Google Search Help. (n.d.). Retrieved March 9, 2021, from https://support.google.com/websearch/answer/9351707

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*(3), 326–338. https://doi.org/10.1037/0022-3514.56.3.326

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118*(2), 126–135. https://doi.org/10.1037/0096-3445.118.2.126

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*(1), 1–24. https://doi.org/10.1006/jmla.1993.1001

Marsh, E. J., & Rajaram, S. (2019). The digital expansion of the mind: Implications of internet usage for memory and cognition. *Journal of Applied Research in Memory and Cognition*, *8*(1), 1–14. https://doi.org/10.1016/j.jarmac.2018.11.001

Medical information on Google—Google Search Help. (n.d.). Retrieved December 28, 2022, from https://support.google.com/websearch/answer/2364942

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, *114*(1), 10–28. https://doi.org/10.1037/pspa0000102

Sloman, S. A., & Rabb, N. (2016). Your understanding is my understanding: Evidence for a community of knowledge. *Psychological Science*, *27*(11), 1451–1460. https://doi.org/10.1177/0956797616662271

Soulo, T. (2017, May 30). Ahrefs' study of 2 million featured snippets: 10 important takeaways. SEO Blog by Ahrefs. https://ahrefs.com/blog/featured-snippets-study/

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Stone, S. M., & Storm, B. C. (2021). Search fluency as a misleading measure of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 53–64. https://doi.org/10.1037/xlm0000806

Titchener, E. B. (1921). *A text-book of psychology*. Macmillan.

Ward, A. F. (2013). Supernormal: How the internet is changing our memories and our minds. *Psychological Inquiry*, *24*(4), 341–348. https://doi.org/10.1080/1047840X.2013.850148

Ward, A. F. (2021). People mistake the internet's knowledge for their own. *Proceedings of the National Academy of Sciences*, *118*(43), 1–10. https://doi.org/10.1073/pnas.2105061118

**How to cite this article:** Eliseev, E. D., & Marsh, E. J. (2023). Understanding why searching the internet inflates confidence in explanatory ability. *Applied Cognitive Psychology*, *37*(4), 711–720. https://doi.org/10.1002/acp.4058

## APPENDIX A

1. Why are there leap years?
2. How is glass made?
3. Why are there jokers in a deck of cards?
4. How (do) zippers work?[5]
5. Why are there dimples on a golf ball?
6. Why are there phases of the moon?
7. Why are there time zones?

*Note*: These explanatory questions were selected from the materials used by Fisher et al., 2015.