Searching for Information in the Digital Age: Implications for Metacognition and Learning

by

Emmaline Drew Eliseev

Department of Psychology & Neuroscience
Duke University

Date:_______________________
Approved:

____
Elizabeth Marsh, Supervisor

____
Tobias Egner

____
David Rubin

____
Gavan Fitzsimons

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Psychology & Neuroscience in the Graduate School
of Duke University

2022

ABSTRACT

Searching for Information in the Digital Age: Implications for Metacognition and Learning

by

Emmaline Drew Eliseev

Department of Psychology & Neuroscience
Duke University

Date:_______________________
Approved:

____
Elizabeth J. Marsh, Supervisor

____
Tobias Egner

____
David C. Rubin

____
Gavan Fitzsimons

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Psychology & Neuroscience in the Graduate School of
Duke University

2022

Abstract

In the current digital age, people are increasingly relying on the internet as their primary source for looking up and learning new information. In 9 experiments, this dissertation seeks to understand how searching for information affects people's metacognitive judgments and learning outcomes. First, we investigate how searching the internet for explanations impacts people's confidence in their explanatory ability and the accuracy of their subsequent explanations. Second, we examine how looking up translations online affects people's judgments of learning and their performance on a learning test. Third, we test how solving word searches influences people's estimates of their knowledge of definitions. As people use cues to infer what they know, the effect of searching on metacognitive judgments depends on the cues that are available during searching. Specifically, searching inflates confidence in one's knowledge when features in the search environment increase feelings of fluency (Ch. 2), but reduces confidence in one's knowledge when searching produces feelings of disfluency (Ch. 3 & 4). Although searching involves active engagement, our results suggest that searching does not benefit learning (Ch. 2) and can even impair learning when it disrupts the encoding of to-be-learned information (Ch. 3). Overall, our findings suggest that it is not the act of searching itself but rather the cues that are available during searching that influence how people assess their own knowledge and how well they learn new information.

Dedication

In memory of Tyomah.

Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank my advisor, Elizabeth Marsh, for her helpful guidance and flexibility as I navigated graduate school. I am grateful for the time and effort she has invested over the years in my growth as a thinker, writer, researcher, and educator. I would also like to express my gratitude to my committee members, David Rubin, Tobias Egner, and Gavan Fitzsimons, for their thoughtful feedback and enthusiastic support. I am also eternally grateful to Carissa Zimmerman for first sparking my curiosity in cognitive psychology and inspiring me to pursue my interests in graduate school. I am thankful for the support of all my family and friends, especially my mom, Jaye, for always encouraging me through challenges and celebrating my successes, Bill and Sue for their endless kindness and generosity and for giving me the best home away from home, and Nadia for always believing in me and being my biggest cheerleader. Finally, I would like to thank my husband and greatest supporter, Nikita, for his constant encouragement and unconditional love.

1. Introduction

What are the symptoms of coronavirus? How to sew a mask? How to make hand sanitizer? How do vaccines work? What are the COVID vaccine side effects? How to pronounce omicron? These questions were among the most popular Google searches in 2020 and 2021 at the height of the COVID-19 pandemic, illustrating how people rely on the internet to find answers and learn new information. While humans have a long history of depending on external stores to remember information for them (cognitive offloading; Risko & Gilbert, 2016), the internet has become the ultimate source for storing and accessing massive amounts of information in the current digital age. As people are increasingly relying on search engines to look up and learn information, it is important to understand the downstream cognitive consequences of searching for information, particularly in online search environments. How does searching for information, especially online, affect how people assess their own knowledge and learning? Do people learn information better if they find it themselves? This dissertation aims to investigate how searching for information influences people's metacognitive judgments and their learning outcomes. Specifically, we explore how different forms of searching impact people's perceptions of their own knowledge and their performance on learning assessments.

1.1 How Do People Know What They Know

> *When you know, to know that you know; and when you do not know, to know that you do not know; that is knowledge.* – Confucius

Whether a student is deciding if they should keep studying for an exam, a game show contestant is deciding if they need to phone a friend for help, a busy shopper is deciding if they need to write down a grocery list, or a doctor is deciding whether they should review a surgical technique by looking it up online, it is important for people to have an accurate assessment of their knowledge and skills. In order to make optimal decisions in these types of scenarios, people need to have awareness of the limits of their own knowledge and abilities. So how do people know what they know versus what they do not know?

Metacognition refers to awareness of one's own cognition (Flavell, 1979), and includes both monitoring and control processes (Nelson & Narens, 1990). Monitoring involves evaluating one's own cognitive processes and is often reported with judgments of learning (JOLs), feeling-of-knowing judgments (FOKs), and confidence ratings, while control refers to the ability to guide behavior based on these metacognitive judgments. In general, accurate monitoring facilitates effective control decisions (R. A. Bjork et al., 2013). For example, participants who accurately monitored their comprehension of study materials also regulated their study time more effectively (e.g., by choosing to restudy the texts they had learned less well), and these students ultimately performed better on the final learning assessment (Thiede et al., 2003).

However, metacognitive monitoring is not always accurate. Without direct access to their own cognition, people make inferences about what they do and do not know based on available cues (Koriat, 1997). The following section describes specific cues that have been shown to influence people's metacognitive judgments.

1.1.1 Cues Influencing Metacognitive Assessments

Many factors influence a person's estimates of how much they know and how much they do not know, including their opinion about the efficacy of their own memory (Hertzog et al., 1990; Beaudoin & Desrichard, 2011), their level of expertise (Atir et al., 2015; Fisher & Keil, 2016; Kruger & Dunning, 1999), their past performance on similar tasks (Hertzog et al., 2013; King et al., 1980), the amount of time they spend on a task (Koriat, 2008; Koriat & Ma'ayan, 2005), how much information they have available (Hall et al., 2007), and their preexisting beliefs about cognition (Mueller et al., 2014; Frank & Kuhlmann, 2017).

People also use their subjective feelings to inform their judgments (feelings-as-information theory; Schwarz, 2012). One of the most robust metacognitive cues that people use to make inferences about their knowledge and abilities is fluency, which refers to the subjective ease of cognitive processing (Koriat, 1997; Oppenheimer, 2008; Begg, Duft, et al., 1989; Whittlesea, 1993; Kelley & Jacoby, 1996; Benjamin & Bjork, 1996; Koriat, 2008; Koriat & Ma'ayan, 2005; Rhodes & Castel, 2008, 2009). Tasks that are easy and effortless feel fluent, while more difficult and demanding tasks feel relatively disfluent.

Fluency can be interpreted in many different ways, such that information that feels easy to process is judged to be more familiar (Schwarz, 2004), better liked (Reber et al., 1998; Lee & Labroo, 2004), more frequent (Tversky & Kahneman, 1973; Reber & Zupanek, 2002), more famous (Jacoby et al., 1989), more true (Dechêne et al., 2010; Brashier et al., 2020), and, critically for our purposes, better learned and more memorable (Begg, Duft, et al., 1989; Hertzog et al., 2003; Koriat, 2008). However, inferences based on feelings of fluency can easily be led astray when the ease of processing results from extraneous factors (e.g., easy-to-read font, high contrast colors, priming) (Schwarz, 2004, 2012). Feelings of fluency can arise from a variety of sources: Information feels fluent when it is easy to perceive (perceptual fluency) or semantically primed (conceptual fluency) and these feelings of fluency can occur when information is easy to commit to memory (encoding fluency) or brought to mind quickly (retrieval fluency) (see Alter & Oppenheimer, 2009). On the other hand, information that is relatively difficult to perceive, process, encode, or retrieve generates feelings of disfluency (Alter, 2013). Despite these differences in how and when fluency is generated, fluency has been shown to affect metacognitive judgments similarly (Alter & Oppenheimer, 2009). In general, people are more confident in information that is fluently processed information relative to information that is less fluently processed (Kelley & Lindsay, 1993; Koriat, 1993).

1.1.1.1 Perceptual Fluency

Perceptual fluency refers to how easily an item is perceived and processed. People assume that information that is easier to perceive has been well learned and will be more memorable relative to information that is more difficult to perceive (Rhodes & Castel, 2008, 2009); this assumption is also known as the ease-of-processing heuristic (Kornell et al., 2011). For example, individuals predicted they would recall more of the study words that were presented in a large, easy-to-perceive font compared to the words presented in a small, more difficult-to-perceive font, even though font size did not affect recall (Kornell et al., 2011). Similarly, font size influenced judgments of learning such that large font words received higher judgments of learning than small font words. Again, font size was not related to memory performance, and this font size bias disappeared when large font words were made harder to read (i.e., less fluent) by alternating uppercase and lowercase letters (e.g., *PiAnO*) (Rhodes & Castel, 2008). Like font size, the clarity of a word also affects metacognitive assessments. When studying both blurry words and clear words, people gave higher judgments of learning to the clear, easy-to-read words than the blurry, hard-to-read words (Yue et al., 2013). In addition to visual information, predictions of future memory are also impacted by the perceptual fluency of auditory information. For instance, loud words received higher judgments of learning than quiet words, even though volume, like font size, had no influence on recall performance (Rhodes & Castel, 2009).

1.1.1.2 Conceptual Fluency

Just as perceptual characteristics influence feelings of fluency, semantically priming concepts can also induce feelings of fluency (Alter & Oppenheimer, 2009). For instance, words (e.g., *boat*) that are primed with semantically related words (e.g., *ship*) feel more familiar than words that are not primed at all (Whittlesea, 1993). Exposure to semantically related concepts through priming makes it easier to process that specific concept and think about related ideas (Reder, 1987; Schwartz & Metcalfe, 1992; Lee & Labroo, 2004; Topolinski & Strack, 2009). Priming participants with words like *golf* and *par* makes it easier to process questions about golf like "*What is the term in golf for scoring one under par?*" (Reder, 1987). As concept priming increases processing fluency, it also influences people's metacognitive judgments about their own knowledge. For instance, priming general knowledge questions (e.g., "*What is the capital of Jamaica?*") with key words from the question (e.g., *Jamaica*) increased participants' feeling-of-knowing judgments (Schwartz & Metcalfe, 1992). Likewise, semantic priming boosts people's confidence in the accuracy of their responses. Participants in one study were more confident in their answers to trivia questions (e.g., "*What was Buffalo Bill's last name?*") if they had been previously exposed to the correct answer (e.g., *Cody*) or even a related but incorrect answer (e.g., *Hickock*) (Kelley & Lindsay, 1993).

1.1.1.3 Encoding Fluency

Encoding fluency, or the ease of committing items to memory, is another way in which fluency influences people's metacognitive assessments. In general, people expect information that feels easy to process to be more memorable. This inference is accurate in some situations; for example, easy-to-learn items, which require less study time and fewer study trials to master, also tend to be remembered better (Koriat, 2008). Similarly, related items are predicted to be more memorable and are typically remembered better than unrelated items (Castel et al., 2007).

However, encoding fluency is not always a valid predictor of future memory. For example, in one study, words that were easier to encode, like concrete and common words, received higher judgments of learning than words that were less easy to encode, like abstract and rare words. However, these predictions of memory were only correctly aligned with memory for concrete words – uncommon words were better remembered than common words, illustrating that encoding fluency is not always an accurate indicator of future memory performance (Begg, Duft, et al., 1989). Building on this early work, Hertzog and colleagues (2003) included a measure for relative encoding fluency by tracking the time it took participants to form interactive, mental images of to-be-learned paired-associates. Participants' judgments of learning were positively correlated with encoding fluency (i.e., paired-associates that took less time to visualize), yet encoding fluency did not accurately predict recall performance (Hertzog et al., 2003).

1.1.1.4 Retrieval Fluency

When assessing their knowledge, people also rely on retrieval fluency, or the ease and speed of accessing information from memory. In general, people are more confident in information that is quickly and easily retrieved from memory (Benjamin & Bjork, 1996; Benjamin et al., 1998). In one study, participants were asked to answer general knowledge trivia questions and were instructed to press a key as soon as they retrieved an answer. Faster response times for retrieving an answer to a trivia question indicated fluent retrieval and were associated with higher predictions of successful recall in the future (Benjamin et al., 1998).

In general, people use retrieval fluency as a cue for assessing their own learning (e.g., when making judgments of learning), as well as for assessing another person's learning. When participants watched other participants generate target words (e.g., *wing*) in response to paired-associate clues (e.g., *rhyme: ring - w* _ _ _) during a study task, the observers predicted better memory performance on a recognition test for those who generated answers more quickly (Matvey et al., 2001).

Nevertheless, retrieval fluency is not always an accurate cue and can mislead people into making inaccurate judgments. For instance, people were more confident in answers that came to mind quickly – even when this facilitation in retrieval occurred, not because of more accurate knowledge, but because they had seen those answers earlier on a list of answers that were explicitly labeled as incorrect (Kelley & Lindsay, 1993). Despite people's beliefs about retrieval fluency, swiftly retrieved answers are not always more memorable. Although individuals expected to remember more of the answers to trivia questions that came to mind quickly than the answers that came to mind slowly, they actually recalled more of the slowly retrieved answers than those they retrieved quickly (Benjamin et al., 1998).

1.1.1.5 Disfluency

Just as cognitive processing can feel easy and effortless, it can also feel difficult and demanding. Disfluency falls at the opposite end of the fluency spectrum and refers to subjective difficulties in cognitive processing (Diemand-Yauman et al., 2011). Like fluency, feelings of disfluency arise when information is relatively difficult to perceive and process conceptually, and this can also occur during the encoding process and at retrieval. While feelings of fluency from easy processing can lead to illusions of knowledge (Finn & Tauber, 2015), disfluency is thought to encourage people to slow down and process information more carefully and deeply (Alter, 2013; Pieger et al., 2017). For example, participants made fewer Moses illusion errors (i.e., responding "*two*" when asked "*How many animals of each kind did Moses take on the Ark?*", Erickson & Mattson, 1981) when questions were presented in a hard-to-read font compared to an easy-to-read font (Song & Schwarz, 2008).

As feelings of fluency typically increase people's metacognitive judgments, it is not surprising that disfluency appears to do the opposite. For instance, students rate disfluent texts as more difficult to understand than fluent texts (Maki et al., 1990) and expect disfluent texts to be harder to learn compared to more fluent texts (Pieger et al., 2016). When participants studied word pairs in both hard-to-read (i.e., gray, italicized) and easy-to-read (i.e., black, bolded) fonts, they gave lower judgments of learning for the word pairs presented in disfluent fonts relative to word pairs presented in fluent fonts, even though perceptual fluency did not affect recall (Magreehan et al., 2016). In another study, participants read a text, half of which was intact, while the other half was missing letters. Participants reported lower predictions of test performance for the disfluent paragraphs with missing letters than for the more fluent, intact paragraphs (Rawson & Dunlosky, 2002).

However, the evidence on how disfluency affects learning and memory outcomes is inconsistent. In some cases, disfluency leads to improved memory. For instance, students who learned from disfluent materials (i.e., hard-to-read worksheets) performed better on the exam than students who learned from fluent materials (i.e., standard worksheets) (perceptual-interference effect; Diemand-Yauman et al., 2011). However, this “disfluency effect” remains controversial as multiple attempts to replicate it (both directly and conceptually) have failed to find an effect of disfluency on memory performance (Kühl & Eitel, 2016; Eitel et al., 2014; Rummer et al., 2016; Taylor et al., 2020; Dunlosky & Mueller, 2016; Strukelj et al., 2016; Pieger et al., 2016).

Nevertheless, disfluency can benefit memory in situations where it facilitates processes that are known to enhance memory, like increased semantic processing. For example, participants read paragraphs that were complete or were missing multiple letters that they had to fill in. When tested, participants recalled more content from paragraphs that had missing letters than from paragraphs without missing letters, suggesting that the additional semantic processing required to fill in the missing letters while reading these disfluent paragraphs improved memory (Maki et al., 1990).

1.1.2 Metacognitive Illusions

Although cues can provide valid indications of one's cognition, they can also mislead people into making inaccurate judgments of their own knowledge and abilities when the cues (e.g., feelings of fluency) arise from an unrelated influence. For instance, while better learned information does tend to come to mind more quickly than information that is not well learned (Benjamin & Bjork, 1996), the experience of retrieval fluency can also reflect irrelevant factors, like previous exposure (Kelley & Lindsay, 1993), and lead one to overestimate their knowledge or memory.

While people can sometimes accurately gauge their knowledge through cue-based inferences, people often overestimate how much they know. From answering trivia questions incorrectly with high ratings of certainty (Fischhoff et al., 1977) to claiming to have impossible knowledge of made-up terms (e.g., "*pre-rated stocks*") (Atir et al., 2015), people make these types of errors with remarkably high levels of confidence. These illusions of knowledge (Glenberg et al., 1982) often occur because people make inferences from invalid cues (e.g., font size, volume) (Rhodes & Castel, 2008, 2009), they lack enough knowledge to correctly estimate how much they do not know (Kruger & Dunning, 1999), or they confuse externally stored information with their own internal knowledge by sometimes treating other people's knowledge as their own (Sloman & Rabb, 2016; Fisher & Oppenheimer, 2021).

Illusions of knowledge prevent people from identifying gaps in their knowledge. Students often experience similar illusions of comprehension when using ineffective study strategies, like rereading or highlighting their notes (R. A. Bjork et al., 2013). For instance, a textbook chapter that has been repeatedly read feels increasingly easier to process, and students are often misled by this ease into thinking they know more material than they do (Karpicke et al., 2009). People also experience illusions of explanatory depth, overestimating how well they could explain natural phenomena, like how rainbows form (Cardwell et al., 2017), and mechanical processes, like how a sewing machines work (Rozenblit & Keil, 2002). While these concepts (e.g., *rainbows*) and objects (e.g., *sewing machines*) seem familiar, people often struggle to provide in-depth explanations when their knowledge is put to the test (Rozenblit & Keil, 2002).

Unfortunately, illusions of knowledge and competence can result in suboptimal decisions, as inaccurate metacognitive monitoring leads to ineffective metacognitive control decisions (R. A. Bjork et al., 2013). For instance, a student who believes they know the material better than they actually do will likely terminate studying too early and ultimately perform poorly on the exam, a person who overestimates their understanding of the inner workings of their car's engine may also overestimate their ability to fix it themselves, and a relator who is highly confident in their knowledge of the real estate market may fail to look up the latest housing market trends.

1.2 Information at Our Fingertips in the Digital Age

> *Getting information from the internet* is *like getting a glass of water from the Niagara Falls*. – Arthur C. Clarke

In the current digital age, information is never more than a Google search away. People are constantly reaching for their smartphones, opening a browser on their laptops, or even asking Alexa to search the internet for a quick answer. Google processes more than 3.5 billion search queries a day with the average person conducting 3 to 4 searches every day (Internet Live Stats, 2021). Search engines, like Google, supply convenient access to the largest repository of human knowledge in existence. From looking up definitions and weather forecasts, to checking the news and the latest CDC guidelines, it is no surprise that the internet has become the primary source for finding information about the world – with only 7% of Americans saying they do not use the internet (Perrin & Atske, 2021). As people increasingly rely on the internet as their main source of information about the world, what are the cognitive implications of looking for information online? Does having instant access to an incredible amount of information online change how people estimate the boundaries of their own knowledge? Does finding information online change how well people learn that information? While this dissertation centers around the overarching question of how searching influences people's metacognitive judgments and learning outcomes, we examine these questions with a particular focus on the effects of online searching.

While relying on external sources of information is not new to the digital age, the internet has revolutionized people's access to information. By providing quick and constant access to information well beyond what could be stored in any individual's memory, the internet has become the ultimate transactive memory partner (Marsh & Rajaram, 2019; Sparrow et al., 2011; Ward, 2013). But the internet differs from other external forms of information storage (e.g., books, partners) in several important ways, including the unparalleled amount of information available, the constant accessibility of this information, and the rapid speed of retrieving relevant information. Unlike asking a partner or looking up information in a book, searching the internet returns millions (and even billions) of results that are related to the search query in just a fraction of a second.

1.2.1 Internet Searching

Internet search engines, like Google, have become the dominant tool for looking up and learning new information. A growing body of research has begun investigating the cognitive consequences of online searching. Searching the internet has been shown to influence several metacognitive judgments, including confidence in general knowledge (Fisher et al., 2015), feelings of knowing (Ferguson et al., 2015), and feelings of findability (Risko et al., 2016). For example, looking up the answers to questions like *How do zippers work?* online increases people's estimations of their own general knowledge of unrelated topics, like weather and American history (Fisher et al., 2015). Although online searching might lead people to overestimate their own general knowledge, they are quite accurate at gauging how easily they could find information that they do not know online. That is, information that was rated higher on a feeling of findability scale was also found faster online than information that was judged to be less easily found (Risko et al., 2016).

Although less research has examined the effects of internet searching on learning, one recent study suggests that online searching can be detrimental to learning. Individuals who searched the internet to learn about a specific topic, like autism or photosynthesis, performed worse on a fact-based, multiple-choice test. Despite this impairment to learning, participants who searched online were as confident in their learning as, and sometimes more confident than, participants who received the same topic information without searching (Fisher et al., 2021).

Having access to the internet also changes people's behavior, making them less likely to rely on their own internal memory. For instance, having access to the internet reduces people's willingness to answer general knowledge questions (e.g., *What is the largest planet in the solar system?*) using their own internally stored knowledge (Ferguson et al., 2015). Moreover, searching the internet quickly becomes habitual, as previous internet searching increases the likelihood of future internet searching – even when searching the internet is inconvenient or unnecessary (Storm et al., 2017).

Why would online searching influence people's metacognitive assessments and behaviors? One possibility is that people conflate their internally stored knowledge with the knowledge that is stored externally online (Sparrow et al., 2011), which in turn inflates their estimations of how much they know individually (Fisher et al., 2015). For instance, people have been known to treat other people's knowledge as their own (Fisher & Oppenheimer, 2021; Sloman & Rabb, 2016). However, if people believe they know more in general after searching the internet because they mistake the internet's knowledge as their own, why are they also more likely to continue using to the internet to look up answers, especially to easy questions that they could likely answer without googling (Storm et al., 2017)?

Another possibility is that features of the online search environment may inadvertently mimic metacognitive cues for knowledge, and these cues mislead people's estimates of their own knowledge and learning. A single Google search can generate millions (if not billions) of search result hits in less than a second that are related to the user's query. In the same way that people expect to recall more of the information that they retrieved quickly from memory (Benjamin et al., 1998), people also predict they will recall more information that they found quickly online (Stone & Storm, 2021). Although retrieval fluency can be a valid indicator of the accessibility of information stored in a person's memory, the speed at which information can be retrieved from an external source, like the internet, is a misleading predictor of memory, as it is unrelated to a person's ability to retrieve this information from their own memory in the future. In addition to the staggering speed of delivery, the sheer amount of relevant information displayed on a Google search results page may activate semantically related concepts, increasing the ease of processing the target answer and making it easier to bring related concepts to mind.

1.3 Overview of the Current Studies

The goal of this dissertation is to understand how searching for information in different settings influences people's metacognitive judgments and learning outcomes. Specifically, this dissertation examines these questions within the context of online searching, supplying evidence that some forms of searching boost confidence and induce illusions of knowledge (see Experiments 1-3), while other types of searching reduce confidence and promote distraction (Experiments 4-7). Overall, our findings suggest that how information is acquired has implications for both metacognition and learning, and we discuss potential explanations for these results in the following chapters.

To preview, not all search is the same. The different kinds of searches described in Chapters 2, 3, and 4 vary in the degree of information available during the searching process – ranging from a very informative search that yields a huge amount of relevant information (i.e., Googling answers to explanatory questions in Experiments 1-3) to an unembellished search that provides only the essential information (i.e., using Google Translate to look up English translations of Swahili words in Experiments 4-5) to a completely non-informative search (i.e., looking for the target word in a word search puzzle in Experiments 6-7). The differences between these forms of searching, and the diverse environments in which they take place, result in differing effects on people's judgments of their own knowledge and the success of their learning.

Experiments 1, 2a, 2b, and 3 aim to investigate how searching for explanations online influences people's perception of how much they know about the explanations they looked up and their subsequent recall of those explanations. Across all four experiments, searching for answers online increased people's confidence in their ability to explain the answers they found online. Experiment 1 compares participants who used Google to find answers to explanatory questions (e.g., *Why are there leap years?*), participants who read the same answers but did not search for them, and participants who did not have any access to the answers. Experiments 2a and 2b ruled out a potential explanation, demonstrating that this confidence boost in explanatory ability is not explained by exposure to the additional web page features (e.g., photos, links, ads) that accompany the explanation articles online. Experiment 3 rules out another possible explanation and offers evidence in support of a potential explanation: Participants in the preview condition, who were exposed to the first page of Google search results before reading the explanation articles, were just as confident in their explanatory ability as participants who searched Google for the explanation articles themselves. The results of Experiment 3 suggest that the confidence boost in explanatory ability is not solely due to the act of searching the internet itself but is likely driven by the pre-exposure of the to-be-learned answers in the search results themselves (Eliseev & Marsh, in preparation-a).

Experiments 4a, 4b, and 5 transition to a different online search environment, Google Translate, to investigate similar questions about people's learning perceptions and outcomes. Experiments 4a and 4b asked whether looking up translations online affected people's estimations of how well they learned the translations and their actual learning outcomes on a test by comparing participants who looked up the English translations of Swahili words on Google Translate to participants who were given the same translations. Experiment 5 ruled out the potential explanation that participants in the search condition are trying to guess the translation by adding a reveal condition where participants experience a brief delay between seeing the Swahili word and receiving the English translation. We also asked participants in Experiment 5 about their preferences and intuitions about whether receiving or searching for translations benefits learning the most. In contrast to the results of Experiments 1-3, participants in Experiments 4-5 who looked up the translations online were less confident in their learning and tended to recall fewer translations than participants who were given the translations. These results challenge the assumption that internet searching always increases confidence in one's knowledge, and we discuss distraction as a possible explanation for these findings (Eliseev & Marsh, in preparation-b).

Finally, we explored a non-informative type of search: Searching for a target word in a word search puzzle. In Experiments 6 and 7, we investigated whether finding a word in a word search makes people more or less confident in their knowledge of that word's definition – a judgment that is unrelated to their ability to locate the word in the word search. In Experiment 6, we asked some participants to actively solve word searches and others to passively view pre-solved word searches before rating their ability to define the target vocabulary words. Experiment 7 compared active solving and passive viewing within-subjects. Critically, in both Experiments 6 and 7, actively solving word searches made people less confident in their knowledge of vocabulary word definitions relative to passively viewing pre-solved word searches – despite demonstrating equal knowledge of the definitions on a knowledge check test (Eliseev & Marsh, in preparation-c).

Overall, our findings across these nine experiments suggest that different forms of searching affect how fluently information is processed, which in turn influences how people assess their own knowledge and how well they learn new information.

2. Searching for Explanations on Google

Whether looking up how to change a tire, why cats purr, what causes hiccups, or how bitcoin works, people are increasingly relying on the internet for finding and learning new information. As previous internet searching increases the likelihood of future internet searching (Storm et al., 2017), it is not surprising that Google processes more than 1.2 trillion search queries per year, a number that is continually growing (*Internet Live Stats*, 2022).

Past work suggests that searching for information online inflates estimations of knowledge (Fisher et al., 2015). However, this work has focused on how online searching inflates people's self-assessments of their general knowledge in broad domains, like weather, food, and American history, that are unrelated to their internet search queries. Experiments 1, 2a, 2b, and 3 expand this line of research by investigating how online searching affects people's confidence in their ability to explain the specific explanations they found online (as opposed to confidence in their general knowledge in domains unrelated to their online searching) and how accurately they can subsequently explain the answers they found online.

We focus on people's estimates of their explanatory ability, as people are particularly susceptible to illusions of explanatory depth (Rozenblit & Keil, 2002). That is, relative to their knowledge of facts, narratives, or procedures, people are especially likely to overestimate their ability to explain natural and mechanical processes

(Rozenblit & Keil, 2002). For example, participants believe they have a deep understanding of how an everyday object, like a ballpoint pen, works, but they are only able to offer shallow explanations when asked to explain its workings in detail (Rozenblit & Keil, 2002).

With the vast amount of information available online, the internet is a powerful tool for learning. When individuals have access to the internet, they tend to produce more accurate answers to questions (Ferguson et al., 2015; Pieschl, 2019), but they are less willing to answer questions independently (Ferguson et al., 2015). A recent study showed that online searching was associated with negative effects on learning: Participants who searched for information about broad topics (e.g., photosynthesis, autism) online performed worse on a multiple-choice question test than participants who were given the same information to read (Fisher et al., 2021). Experiments 1, 2a, 2b, and 3 seek to extend this work by focusing on how searching the internet for explanations to questions (e.g., *Why are there leap years?*) affects how well people learn those specific answers by testing their recall of these explanations, as opposed to their recognition of facts about broad topics.

The following four experiments (Experiments 1-3) examine how searching for explanations online impacts people's confidence in their ability to explain the specific answers they found online and how accurately they recall these explanations. To preview, we find evidence that online searching leads to illusions of knowledge, as participants who searched the internet for explanations were more confident in their explanatory ability but did not consistently produce more accurate answers when compared with participants who did not search the internet.

2.1 Experiment 1

2.1.1 Method

2.1.1.1 Participants

Three-hundred and twelve Amazon Mechanical Turk workers located in the U.S. with an approval rating above 90% participated online for compensation. Seventeen participants were excluded for failing to follow instructions, so data were analyzed with the remaining 295 participants (144 female, 150 male, 1 non-binary; *M* age = 34.94 years).

2.1.1.2 Design

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

2.1.1.3 Materials

We selected 7 explanatory knowledge questions (e.g., *Why are there leap years?*) from Fisher, Goddu, and Keil (2015).[1] 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 in which the questions appeared was randomized in all conditions.

---

[1] We chose not to use one question from the original materials from Fisher, Goddu, and Keil (2015), as a reliable answer was difficult to find online.

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

2.1.1.4 Procedure

After giving informed consent, participants completed the rating phase where they were asked to rate how well they could explain the answers to all 7 explanatory questions (e.g., *Why are there leap years?*). Before making their explanatory ability rating 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 specific 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 verify that participants found the correct target article, they were asked to report the URL of the specific web page they found. Participants in the read condition were asked to read a text to confirm the details of the explanation to the question. These explanation texts were copied and pasted from the same target web pages that participants in the search condition were asked to find online. Participants in the control condition were simply asked to evaluate how well they could explain the answer to each question based on their own knowledge without using any outside sources.

After making explanatory ability ratings for all 7 questions, participants completed the explanation phase in which participants were asked to explain the answers to the questions they saw in the previous rating phase. They were instructed to provide as much detail as possible in their explanations without consulting any outside sources. After completing the explanation phase, participants were asked to report if they had used the internet to look up the answers to any questions in either the rating phase or the explanation phase; any participants who reported using unauthorized sources to look up answers were excluded. After completing the study, participants filled out their demographic information and were debriefed.

2.1.2 Results

The alpha level for all statistical tests was set at .05. Planned comparisons tested whether explanatory ability ratings and explanation accuracy varied by condition.

2.1.2.1 Confidence in Explanatory Ability

We conducted a one-way between-subjects ANOVA on participants' ratings of explanatory ability by condition (control, read, search) collapsing across all 7 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 differed significantly by condition, $F(2, 191.83) = 65.72$, $p < .001$, $\eta^2 = .41$. Participants in the search condition ($M = 5.45$, $SD = 0.99$) gave significantly higher ratings of explanatory 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 ($M = 3.55$, $SD = 1.30$, $t(179.11) = 11.48$, $p < .001$, $d = 1.65$). Participants in the read condition also gave higher ratings of explanatory ability than those in the control condition, $t(185.39) = 6.86$, $p < .001$, $d = 0.98$.

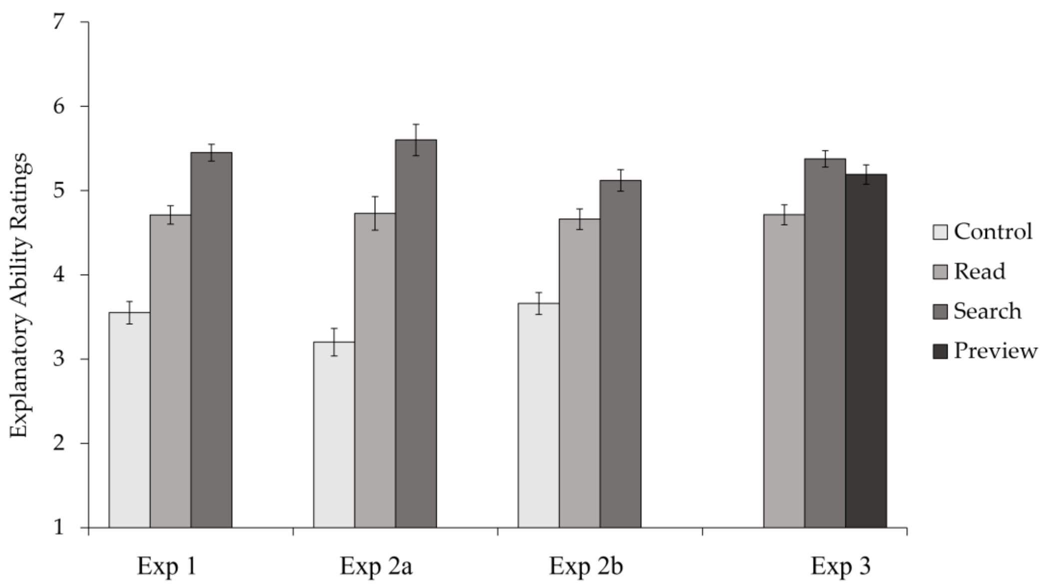

Figure 1: Mean ratings of explanatory ability by condition for Experiments 1, 2a, 2b, and 3. Error bars reflect standard error of the mean. Note that Experiment 3 included a preview condition instead of a control condition.

2.1.2.2 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 ($\kappa = .86$), and all scoring discrepancies were resolved through discussion. An accuracy score was calculated for each participant based on the proportion of total possible points (14) they received across their responses to the 7 questions ranging from 0 to 1. The relevant data appear in Figure 2. A one-way between-subjects ANOVA on participants' accuracy scores revealed accuracy differed significantly by condition, $F(2, 292) = 3.79$, $p = .024$, $\eta^2 = .03$. Participants in the search condition ($M = 0.48$, $SD = 0.23$) produced more accurate responses than participants in the read condition ($M = 0.39$, $SD = 0.22$, $t(196) = 2.78$, $p = .006$, $d = 0.40$). However, accuracy scores in the control condition ($M = 0.44$, $SD = 0.24$) did not differ significantly from the read condition, $t(194) = 1.58$, $p = .116$, $d = 0.22$, or the search condition, $t(194) = -1.14$, $p = .258$, $d = 0.17$.

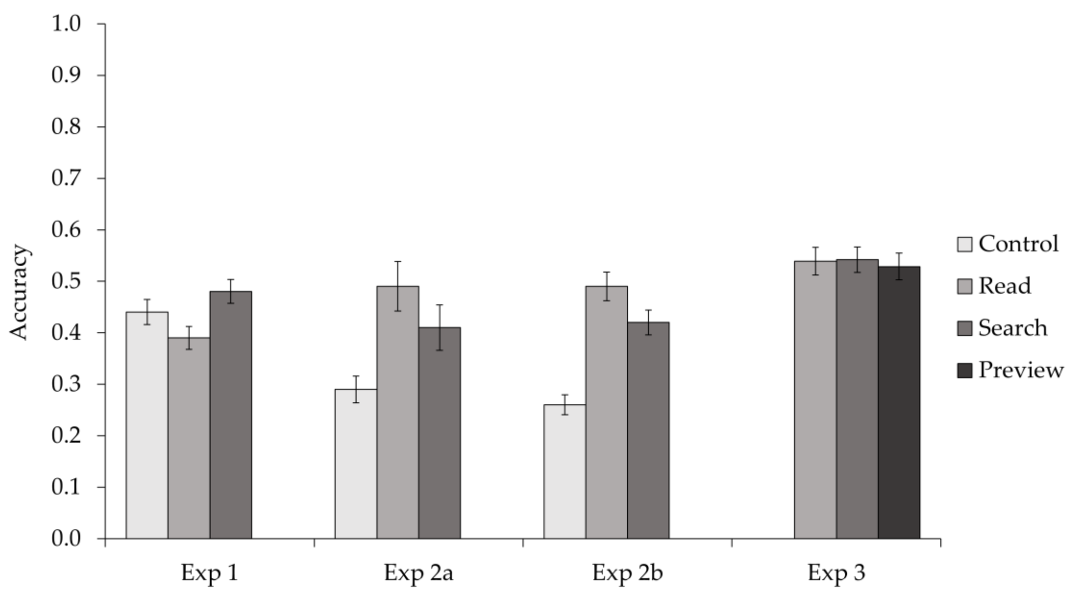

Figure 2: Mean accuracy of explanations by condition for Experiments 1, 2a, 2b, and 3. Error bars reflect standard error of the mean. Note that Experiment 3 included a preview condition instead of a control condition.

2.1.3 Discussion

As expected, participants in the search and read conditions were more confident in their ability to explain the answers to the explanatory questions than participants in the control condition, who did not have access to any of the explanations. Critically, participants who searched the internet for the target explanation articles were more confident in their explanatory ability than participants who read, but did not search for, the same explanation articles.

While participants in the read condition were given plain text that was copied and pasted from the original online article, participants in the search condition were exposed to target web pages that contained additional content beyond just the text; these web pages, 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. These superfluous elements could potentially distract participants and impair learning; for example, irrelevant images (e.g., a picture of a lightning bolt) and other “seductive details” have been shown to draw attention away from the to-be-learned text and ultimately lead to poor learning outcomes (Garner et al., 1992; Sanchez & Wiley, 2006). Additionally, photos, even uninformative ones (e.g., a picture of a rainbow), can increase people’s self-perceived knowledge and their estimations of understanding complex concepts (e.g., how rainbows form) (Cardwell et al., 2017), so it is possible that the search condition’s confidence ratings were inflated by exposure to these additional web page elements. In order to rule out this possibility, Experiments 2a and 2b asked participants in the read condition to read full-page screenshots of the target web pages, so they would be exposed to the same explanation article and approximately the same web page elements[2] as participants in the search condition.

Intriguingly, participants in the search condition produced more accurate explanations than participants in the read condition suggesting a small learning benefit from online searching. However, participants in the control condition, who did not have access to the explanatory question answers, unexpectedly scored as well on the explanation test as participants in the read and search conditions, who did have access to the full explanations. It is possible that cheating was underreported in this experiment leading to inflated accuracy scores in the control condition. In the attempt to capture a more precise measure of cheating, Experiments 2a and 2b asked participants if they used an unauthorized source to look up the answer after each question individually, rather than asking for an aggregate judgment at the end.

2.2 Experiments 2a and 2b

These experiments share the same design and methods, so we report them together. Experiment 2b was designed to directly replicate Experiment 2a with a larger sample.

[2] Some of the web page elements (i.e., advertisements) that participants see do vary individually depending on the user's personal search history and browser settings. We created the full-page screenshots of the target web pages in an internet browser after clearing history and cookies while using private browsing mode (aka incognito mode) in the attempt to avoid capturing targeted ads that were overly specific.

2.2.1 Method

2.2.1.1 Participants

In Experiment 2a, 119 Amazon Mechanical Turk workers located in the U.S. with an approval rating above 90% participated online for compensation. Twenty-five participants were excluded for failing to follow instructions[3], so data were analyzed with the remaining 94 participants (39 female, 53 male, 1 non-binary, 1 unreported; *M* age = 35.95 years).

In Experiment 2b, 288 Amazon Mechanical Turk workers located in the U.S. with an approval rating above 90% participated online for compensation. Thirty-five participants were excluded for failing to follow instructions[3], so data were analyzed with the remaining 253 participants (113 female, 136 male, 2 non-binary, 2 unreported; *M* age = 35.11 years).

2.2.1.2 Design

These experiments had the same design as Experiment 1 with three between-subjects conditions: control, read, and search.

2.2.1.3 Materials

We used the same 7 explanatory knowledge questions and the same pre-specified explanation websites as in Experiment 1. In order to better match the content

---

[3] Specific reasons for exclusion included reporting looking up answers on 3 or more of the 7 questions, not taking the task seriously, and reporting incorrect URLs.

that participants saw in the read condition to the search condition, we used full-page screenshots of the target 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 website layout, images, and other potentially distracting items when reading the explanation articles.

2.2.1.4 Procedure

The procedure was similar to that of Experiment 1, with the exceptions that 1) participants were asked to report cheating 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 articles) 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.

2.2.2 Results

The alpha level for all statistical tests was set at .05. Planned comparisons tested whether explanatory ability ratings and explanation accuracy varied by condition.

2.2.2.1 Confidence in Explanatory Ability

We conducted a one-way between-subjects ANOVA on participant's ratings of explanatory ability by condition (control, read, search) collapsing across all 7 individual questions. The relevant data appear in Figure 1. Confidence in explanatory ability differed significantly by condition [Experiment 2a: $F(2, 91) = 44.27$, $p < .001$, $\eta^2 = .49$; Experiment 2b: $F(2, 250) = 35.25$, $p < .001$, $\eta^2 = .22$]. Participants in the search condition (Experiment 2a: $M = 5.60$, $SD = 1.06$; Experiment 2b: $M = 5.12$, $SD = 1.17$) gave significantly higher ratings of explanatory ability than participants in the read condition (Experiment 2a: $M = 4.73$, $SD = 1.09$; Experiment 2b: $M = 4.66$, $SD = 1.12$) [Experiment 2a: $t(60) = 3.16$, $p = .002$, $d = 0.80$; Experiment 2b: $t(165) = 2.58$, $p = .011$, $d = 0.40$] and those in the control condition (Experiment 2a: $M = 3.20$, $SD = 0.93$; Experiment 2b: $M = 3.66$, $SD = 1.20$) [Experiment 2a: $t(62) = 9.59$, $p < .001$, $d = 2.40$; Experiment 2b: $t(169) = 8.06$, $p < .001$, $d = 1.23$]. Participants in the read condition also gave higher ratings of explanatory ability than those in the control condition [Experiment 2a: $t(60) = 5.93$ $p < .001$, $d = 1.51$; Experiment 2b: $t(166) = 5.60$, $p < .001$, $d = 0.87$].

2.2.2.2 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 (Experiment 2a: $\kappa = .75$; Experiment 2b: $\kappa = .80$), 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 7 questions. The relevant data appear in Figure 2. Homogeneity of variance was violated in Experiment 2a, as assessed by Levene's Test of

Homogeneity of Variance (*p* = .002). A one-way between-subjects ANOVA on participants' accuracy scores revealed that explanation accuracy differed significantly by condition [Experiment 2a: $F(2, 55.23) = 7.43$, $p = .001$, $\eta^2 = .21$; Experiment 2b: $F(2, 250) = 25.42$, $p < .001$, $\eta^2 = .17$]. Participants in the control condition (Experiment 2a: *M* = 0.29, *SD* = 0.15; Experiment 2b: *M* = 0.26, *SD* = 0.18), who did not have access to any external explanations, produced less accurate explanations than either participants in the read condition (Experiment 2a: *M* = 0.49, *SD* = 0.26; Experiment 2b: *M* = 0.49, *SD* = 0.25) [Experiment 2a: $t(44.77) = -3.54$, $p = .001$, $d = -0.92$; Experiment 2b: $t(146.59) = -6.89$, $p < .001$, $d = -1.07$] or participants in the search condition (Experiment 2a: *M* = 0.41, *SD* = 0.25; Experiment 2b: *M* = 0.42, *SD* = 0.22) [Experiment 2a: $t(50.06) = -2.38$, $p = .021$, $d = -0.60$; Experiment 2b: $t(169) = -5.31$, $p < .001$, $d = -0.81$]. While accuracy scores did not differ significantly between the read and search conditions in Experiment 2a [$t(60) = 1.10$, $p = .277$, $d = 0.32$], accuracy scores in the read condition were marginally higher than accuracy scores in the search condition in Experiment 2b [$t(165) = 1.93$, $p = .055$, $d = 0.30$].

2.2.3 Discussion

Experiments 2a and 2b replicated the key finding from Experiment 1 that participants in the search condition were more confident in their ability to explain the answers to questions than participants in the read condition, who read but did not search for the same explanations – even when controlling for additional web page content (e.g., links, photos, advertisements) that might influence confidence ratings. Thus, the results of Experiments 2a and 2b suggest that the search condition's boost in confidence is not explained by exposure to the additional web page content.

Experiment 3 was designed to probe the underlying mechanism of this increased confidence in explanatory ability in the search condition. On one hand, this illusion of explanatory depth could be driven by the act of searching itself. That is, participants may interpret their active engagement in the searching process, including generating a search query and finding a target search result, as evidence for actively learning the explanations. On the other hand, this inflated confidence in explanatory ability could be explained by increased familiarity with the explanations after previewing the to-be-learned information on the Google search results page. To investigate whether this boost in confidence is facilitated by the active process of searching or by previewing the explanations on the Google search results page, Experiment 3 added a new preview condition where participants did not actively search the internet themselves but were exposed to the first page of Google search results before reading the full explanation article.

Contrary to Experiment 1, participants in the search condition in Experiments 2a and 2b did not generate more accurate explanations than participants in the read condition suggesting no learning benefit from searching for answers online. Additionally, unlike Experiment 1, participants in Experiments 2a and 2b in the read and search conditions, who had previously read the explanations, generated more accurate explanations than participants in the control condition, who did not have access to the explanations. Experiment 3 aims to further investigate the effects of online searching on how well individuals learn the target explanations.

2.3 Experiment 3

2.3.1 Method

2.3.1.1 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. Thirty-six participants were excluded for failing to follow instructions (33 did not follow instructions, 3 reported cheating), so data were analyzed with the remaining 261 participants (111 female, 149 male, 1 non-binary; *M* age = 36.43 years).

2.3.1.2 Design

This experiment had three between-subjects conditions: preview, read, and search.

2.3.1.3 Materials

We used the same 7 explanatory knowledge questions, pre-specified explanation websites, and full-page screenshots as in Experiments 2a and 2b. Unlike the previous studies, a preview condition replaced the control condition. For the new preview condition, we captured full-page screenshots of the first page of Google search results for each explanatory question query. Thus, participants in the preview condition were exposed to similar search results as participants in the search condition, including web page titles, a "featured snippet" (i.e., a brief answer to the search query automatically featured in an answer box at the top of the search results page), and "snippets" (i.e., short descriptions of each search result that are relevant to the search query) on the front page of Google (see Figure 3 for examples) for each explanatory question.

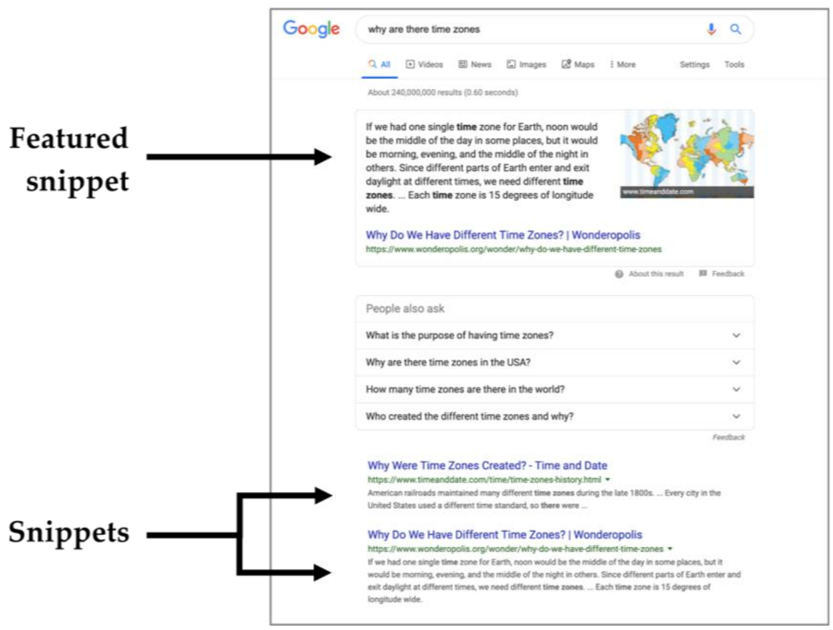

Figure 3: Examples of snippets and a featured snippet from a Google search results page.

2.3.1.4 Procedure

The procedure was the same as in Experiments 2a and 2b 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 website to read (as in the read condition) for each question before rating their ability to explain the answer to this question. After completing the rating phase, all participants completed the explanation phase as in Experiments 2a and 2b.

2.3.2 Results

The alpha level for all statistical tests was set at .05. Planned comparisons tested whether explanatory ability ratings and explanation accuracy varied by condition.

2.3.2.1 Confidence in 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 differed significantly by condition, $F(2, 258) = 9.44$, $p < .001$, $\eta^2 = .07$. Participants in the search condition ($M = 5.38$, $SD = 0.92$) gave significantly higher ratings of explanatory ability than participants in the read condition ($M = 4.71$, $SD = 1.10$, $t(171) = 4.30$, $p < .001$, $d = 0.65$). 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 differ significantly between the search and preview conditions, $t(174) = 1.23$, $p = .220$, $d = 0.19$.

2.3.2.2 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 ($\kappa = .90$), 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 across their responses to the 7 questions ranging from 0 to 1. A one-way between-subjects ANOVA on participants' accuracy scores revealed no significant differences between the preview condition ($M = 0.53$, $SD = 0.24$), read condition ($M = 0.54$, $SD = 0.25$), or search condition ($M = 0.54$, $SD = 0.23$), $F(2, 258) = 0.07$, $p = .929$, $\eta^2 = .00$. The relevant data appear in Figure 2.

2.4 Discussion of Experiments 1, 2a, 2b, & 3

Our findings suggest that how information is acquired influences both metacognitive judgments and learning. Across four experiments, people who searched the internet for answers to explanatory questions (or were exposed to the Google search results) rated their ability to explain these answers higher than people who read but did not search for the same explanations.

Given the mixed results on learning outcomes, we performed a mini meta-analysis for the difference in explanation accuracy between the read condition and the search condition in Experiments 1-3 using 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 no significant difference in explanation accuracy between the read and search conditions, $Z = -0.19$, $p = .849$, $d = -0.02$, 95% CI [-0.18, 0.15]. These results suggest that actively searching for explanations online does not benefit learning relative to directly receiving the same explanations. Thus, the inflated confidence in explanatory ability reflects an illusion of explanatory depth (e.g., Rozenblit & Keil, 2002), rather than superior learning of explanations. Although online searching makes people feel more confident in their explanatory knowledge, this increased confidence does not translate into generating higher quality, more accurate explanations at test.

Despite no difference in demonstrated explanation ability, people who searched for answers online (or were exposed to the Google search results) felt more confident in their explanatory knowledge. In Experiment 1, we document this effect directly, showing that searching for information online inflates people's confidence in their ability to explain that information. The results of Experiments 2a and 2b rule out a possible explanation, showing that this confidence boost in explanatory ability is not simply explained by exposure to the additional web page features (e.g., photos, links, advertisements) that accompany the explanation articles online.

Critically, Experiment 3 shows that when participants in the preview condition were exposed to the first page of Google results before reading the explanation articles, they were just as confident in their explanatory ability as participants who searched for the explanations online. This suggests that the confidence boost in explanatory ability is not solely due to the act of searching the internet, but it may be driven by prior exposure to the to-be-learned information presented in Google search results, like snippets (i.e., short descriptions of each search result that are relevant to the search query) and featured snippets (i.e., answers boxes that feature a brief answer automatically pulled from a top search result) (see Figure 3 for examples). Seeing the Google search results page is the modern-day equivalent of Titchener's famous example of a person who glances both ways before crossing the street, gets distracted, crosses the street but feels as if they have already crossed the street (Titchener, 1921). Brief exposure to information (whether verbatim or just semantically related) makes it easier to process (Reder, 1987), and increases people's confidence in their knowledge (Kelley & Lindsay, 1993; Schwartz & Metcalfe, 1992). This priming may explain why participants in the search and reveal conditions, who were pre-exposed to the target explanations via the Google search results page, were more confident in their ability to explain those answers than participants in the control or read conditions who were not primed with search results. These findings challenge the assumption that internet searching increases people's confidence in their knowledge, because they conflate their internally stored knowledge with the knowledge that is stored externally online (Fisher et al., 2015; Sparrow et al., 2011).

3. Looking up Translations on Google Translate

If the internet is now people's primary source for information, what are the cognitive implications of looking up and learning new information online? In the previous four experiments, searching the internet for explanations led to a clear increase in people's confidence in their ability to explain the answers they had found online; however, the impact of searching the internet on learning was less clear. The next three experiments aim to shed light on how looking up to-be-learned information online affects people's metacognitive judgments of learning and their actual learning as measured by performance on a cued-recall test. Specifically, Experiments 4a, 4b, and 5 investigate another form of internet searching using a popular translation tool, Google Translate. Unlike the interface of the main Google search engine, which includes an abundance of relevant information (i.e., featured snippet boxes, snippets for each search result), Google Translate offers a pared-down search environment and simply translates the input (e.g., a given word or phrase) from one language into another.

Prior research has shown that actively generating information benefits memory compared to passively receiving the same information. Specifically, memory is improved for information that is self-generated (either partially or completely) than for the same information that is simply read, a finding known as the generation effect (Slamecka & Graf, 1978; Bertsch et al., 2007). While the generation effect refers to a memory benefit for answers that are internally generated (i.e., retrieved from a person's own memory) during the learning phase, Experiments 4a, 4b, and 5 investigate how using an online tool to externally generate answers (i.e., retrieving information online) during the learning phase affects people's perceptions of learning and their subsequent memory performance.

In the classic generation effect paradigm, participants are presented with a list of paired-associates. For half of the list, participants are asked to simply read the intact pair (e.g., *COLD, HOT*); for the other half of the list, participants are given half of the pair (e.g., *LONG*) and are instructed to generate the second word in the pair (i.e., *SHORT*) using a specific rule (e.g., opposites, synonyms, rhymes). In some experiments, participants are given a word fragment (e.g., *CAVE, S -* or *Animal, H_R_E*) when asked to generate the target word (Slamecka & Graf, 1978; Gardiner et al., 1989). Then participants complete a memory test (e.g., recognition, cued-recall, free recall). Consistent effects have been demonstrated across different study materials and variations of the original paradigm (Anderson et al., 1971; Jacoby, 1978; Bertsch et al., 2007; McCurdy et al., 2020). For example, a memory advantage is observed when filling in the last word in a sentence (e.g., "*According to Skinner, a reward is positive RE*_ _ _ _ _ _ _ _ _ _ _") (Peynircioğlu & Mungan, 1993), when solving anagrams (e.g., *NOUBOBR*) in response to clues (e.g., *American whisky*) (Gardiner et al., 1989), and even when solving multiplication problems (e.g., *30 x 9*) (McNamara & Healy, 2000). Generation can also benefit learning new information (Dunlosky et al., 2013; Metcalfe & Kornell, 2007). Even generating errors can improve memory for to-be-learned information relative to reading, if feedback is provided (Potts & Shanks, 2014; Metcalfe, 2017).

From a desirable difficulties perspective, items that are harder to generate should result in better long-term learning outcomes (E. L. Bjork & Bjork, 2011). Generation involves processes that are known to benefit memory, including increased item-specific and relational processing (Hirshman & Bjork, 1988; McDaniel et al., 1988), relative distinctiveness (Begg, Snider, et al., 1989; Kinoshita, 1989), semantic activation (McElroy & Slamecka, 1982), and deeper semantic processing (Craik & Lockhart, 1972). The magnitude of this memory advantage of generation is almost half of a standard deviation (.40) over reading (Bertsch et al., 2007). For an updated meta-analytic review, see (McCurdy et al., 2020).

It is possible that actively looking up translations may result in a similar benefit for learning translations. In this case, instead of self-generating the answer, participants are asked to use an external aid (i.e., Google Translate) to generate the answer. On the other hand, looking up translations on Google Translate may only involve shallow levels of processing, encouraging the learner to focus on superficial characteristics of the to-be-remembered translations (e.g., its spelling) rather than deeper levels of semantic processing that are known to benefit memory (Craik & Lockhart, 1972).

In the following experiments (Experiments 4-5), participants learned English translations of Swahili vocabulary words (e.g., *wingu - cloud*). We investigate how looking up the English translations online affects people's perceptions of their learning and their performance on a learning assessment compared to people who directly receive the same translations.

3.1 Experiments 4a and 4b

Experiments 4a and 4b share a design and procedure, so we report them together.

3.1.1 Method

3.1.1.1 Participants

In Experiment 4a, 218 Duke undergraduates participated for course credit. Sixteen participants were excluded (14 reported looking up answers during the test, 2 did not complete the study due to technical difficulties). Data from the remaining 202 Duke undergraduates (129 female, 72 male, 1 unreported; *M* age = 19.24 years) were analyzed. English was the reported as the native language by 83.7% of participants, 14.4% reported a native language other than English, 1.0% reported being natively bilingual with English and another language, and 1.0% did not report a native language. Crucially, no participants reported speaking Swahili.

In Experiment 4b, 316 Duke undergraduates participated for course credit. Sixteen participants were excluded (11 participants reported looking up answers during the test, 4 participants did not finish the experiment, and 1 participant did not respond to the question about cheating). Data from the remaining 300 Duke undergraduates (193 female, 101 male, 2 non-binary, 4 unreported; *M* age = 19.31 years) were analyzed. English was the reported as the native language by 82.3% of participants, 15.7% reported a native language other than English, 1.0% reported being natively bilingual with English and another language, and 1.0% did not report a native language. Crucially, no participants reported speaking Swahili.

3.1.1.2 Design

Both Experiments 4a and 4b had two between-subjects conditions: control and search.

3.1.1.3 Materials

We selected 20 Swahili-English translation pairs (e.g., *wingu - cloud*) from previously published norms (Nelson & Dunlosky, 1994). We specifically chose Swahili nouns that had only one English translation, according to Google Translate. In the time between running Experiment 4a (Spring 2020) and Experiment 4b (Spring 2021), Google Translate had changed the English translations for seven of the Swahili words previously used in Experiment 4a (e.g., *baharia - sailor* changed to *baharia - sailors*, *lulu - pearl* changed to *lulu - lulu*), so we updated the list of Swahili-English word pairs for Experiment 4b to ensure words were translated consistently. See Appendix B for a complete list of the Swahili-English translation pairs used in Experiment 4a and Appendix C for the word pairs used in Experiment 4b.

3.1.1.4 Procedure

After giving informed consent, participants completed the learning phase. Participants were randomly assigned to either the control condition or search condition. All participants were asked to learn the English translations of the Swahili words. In the control condition, participants were shown a Swahili-English word pair (e.g., *wingu - cloud*) and were asked to type the Swahili word (i.e., *wingu*) and the English translation (i.e., *cloud*) in the designated textboxes before continuing to the next pair. Participants in the search condition were instructed to use Google Translate to look up the English translations of the Swahili words. Specifically, participants were asked to open Google Translate (https://translate.google.com/) in a separate browser tab and received instructions on how to set it up to translate words from Swahili to English. After verifying that their Google Translate setup matched an image of the correct setup, search condition participants saw a Swahili word (i.e., *wingu*) and were asked to look up the English translation using Google Translate and then to type the English translation (i.e., *cloud*) they found in the designated textbox before continuing to the next Swahili word. All participants continued their learning task (either typing out both words or looking up and typing out the translations) for all 20 Swahili-English translation pairs. After seeing all 20 Swahili-English translations, participants were asked to estimate how many of the 20 English translations that they saw in the first part of the study did they expect to remember in the second part of the study by entering a number between 0 and 20. To ensure that participants followed instructions in the learning phase, participants in the control condition were asked if they looked up any translations using an outside source, and participants in the search condition were asked if they looked up any translations using an outside source other than Google Translate. We excluded any participants who reported failing to follow instructions by looking up translations with an unauthorized source during the learning phase.

Immediately after the learning phase, participants completed the test phase. All participants took a cued-recall test where they were given the same Swahili words from the learning phase (e.g., *wingu - ???*) and asked to recall their respective English translations (e.g., *cloud*). If they did not know the correct translation, participants were encouraged to give their best guess. They were also told not to use any outside sources to look up the answers during the test. For each translation, participants were asked to rate their confidence in the accuracy of their answer from 1 (*not confident at all*) to 5 (*very confident*). After being tested on all 20 translations, participants were asked if they looked up any answers using an outside source during the test. If they answered yes, they were asked to indicate how many answers they looked up. We excluded participants who reporting looking up any answers during the test. After completing the study, participants filled out their demographic information and were debriefed.

3.1.2 Results

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

3.1.2.1 Judgments of Learning

We conducted a one-way between-subjects ANOVA on participants' aggregate judgments of learning (i.e., the number of English translations they expected to recall at test out of 20). The relevant data appear in Figure 4. Critically, participants who searched for translations online (Experiment 4a: $M = 4.90$, $SD = 3.02$; Experiment 4b: $M = 4.35$, $SD = 2.82$) expected to recall fewer translations than participants who were given the same translations to study (Experiment 4a: $M = 6.04$, $SD = 3.81$; Experiment 4b: $M = 5.70$, $SD = 3.29$) [Experiment 4a: $F(1, 191.70) = 5.56$, $p = .019$, $\eta^2 = .03$; Experiment 4b: $F(1, 298) = 14.64$, $p < .001$, $\eta^2 = .05$].

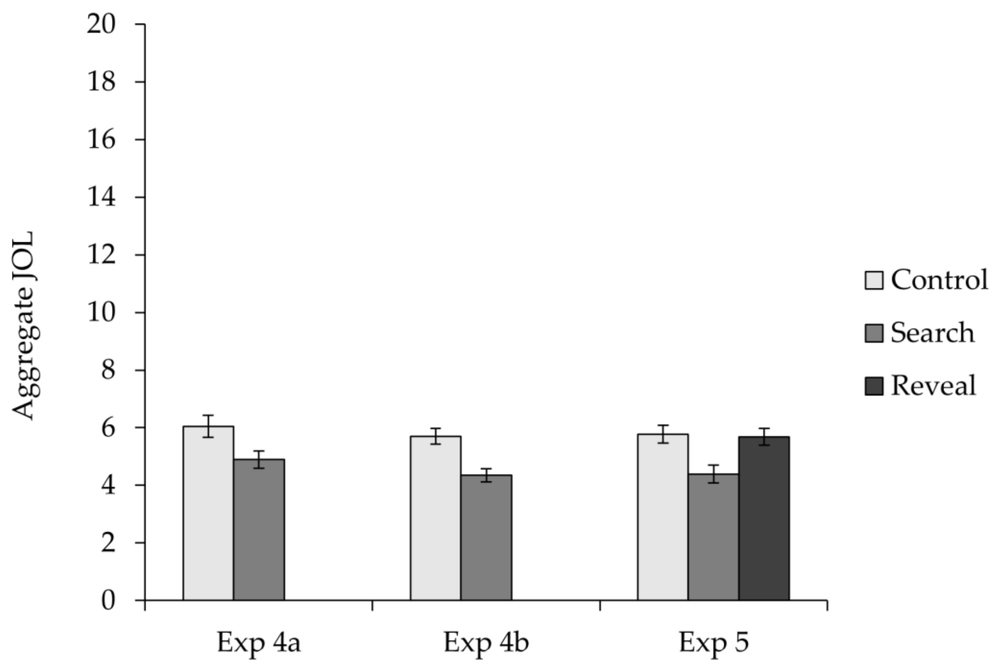

Figure 4: Mean aggregate judgments of learning by condition in Experiments 4a, 4b, and 5. Aggregate JOL refers to the number of English translations participants expected to recall out of 20. Error bars reflect standard error of the mean. Note that the reveal condition was only included in Experiment 5.

3.1.2.2 Performance on Cued-Recall Test

We conducted a one-way ANOVA on participants' test accuracy, defined as the number of English translations (out of 20) that participants correctly recalled on the cued-recall test. The relevant data appear in Figure 5. In Experiment 4a, there was no significant difference between the search condition (M = 5.49, SD = 3.46) and the control condition (M = 6.28, SD = 4.10) in the number of English translations correctly recalled, $F(1, 200) = 2.21$, $p = .139$, $\eta^2 = .01$. However, in Experiment 4b, participants in the search condition (M = 4.66, SD = 3.18) recalled significantly fewer translations than participants in the control condition (M = 6.58, SD = 4.29) who did not use Google Translate in the learning phase, $F(1, 272.60) = 19.24$, $p < .001$, $\eta^2 = .06$. Although the difference in recall between the control and search conditions was not statistically significant in Experiment 4a, the trend was consistent with Experiment 4b, as search condition participants recalled numerically fewer translations than control condition participants.

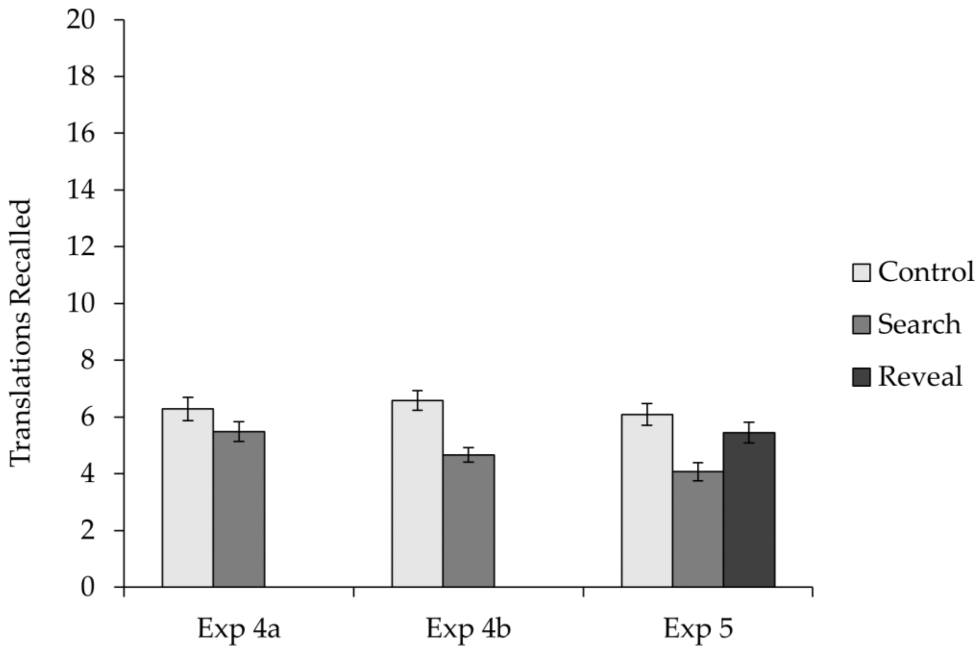

Figure 5: Mean number of translations correctly recalled by condition for Experiments 4a, 4b, and 5. Error bars reflect standard error of the mean. Note that the reveal condition was only included in Experiment 5.

3.1.2.3 Confidence Ratings at Test

We conducted a one-way ANOVA on participants' confidence ratings for the translations they correctly recalled at test.[4] For ease of interpretation, the original scale values that ranged from 1 (*not at all confident*) to 5 (*very confident*) were recoded to range from 0 (*not at all confident*) to 4 (*very confident*). Participants who searched for translations online (Experiment 4a: $M = 2.69$, $SD = 1.00$; Experiment 4b: $M = 2.46$, $SD = 1.11$) were less confident in the accuracy of their correct answers on the cued-recall test than participants who were given the same translations to study (Experiment 4a: $M = 2.98$, SD

[4] This analysis only included the confidence ratings from participants who correctly recalled at least one English translation at test.

= 0.92; Experiment 4b: *M* = 2.93, *SD* = 0.91) [Experiment 4a: $F(1, 193) = 4.41$, $p = .037$, $\eta^2 =$ .02; Experiment 4b: $F(1, 268.08) = 15.28$, $p < .001$, $\eta^2 = .05$].

3.1.3 Discussion

Unlike searching for answers to explanatory questions on Google (i.e., Experiments 1-3), looking up translations on Google Translate in Experiments 4a and 4b decreased people's confidence in their learning. This pattern of reduced confidence continued at test, as participants in the search condition remained less confident in the accuracy of their correctly recalled translations on the cued-recall test compared to participants in the control condition.

Despite the consistency in the judgments of learning and confidence ratings at test results across Experiments 4a and 4b, the test performance data are less clear. In Experiment 4a, there was no significant difference between conditions in the number of English translations correctly recalled at test; however, in Experiment 4b, participants in the search condition recalled significantly fewer translations than participants in the control condition who did not use Google Translate in the learning phase. In spite of the differing recall results, the pattern of data was consistent across Experiments 4a and 4b, with search condition participants recalling numerically fewer correct translations than control condition participants.

In addition to further investigating any potential effects of searching on learning, Experiment 5 was conducted to test a potential explanation for the reduced confidence in the search condition relative to the control condition. By not simultaneously receiving the English translation with the Swahili word (e.g., *wingu - cloud*), as in the control condition, it is possible that participants in the search condition spontaneously tried to guess the answer while looking up the translation. When information is novel and unstudied, attempting to guess the correct answer can emphasize what people do not know. For example, participants perceived learning to be more difficult and gave lower judgments of learning when they had to guess the translation of an unknown Euskara word (e.g., *igel* - ?) before receiving the correct translation (e.g., *frog*) compared to when they received an intact Euskara-English word pair (e.g., *igel - frog*) to learn (Potts & Shanks, 2014).

Attempting to retrieve answers, even wrong answers, has been shown to improve learning when accompanied by feedback (Metcalfe, 2017). For example, incorrectly guessing an unstudied target word (e.g., guessing "track") when given a weakly related cue word (e.g., *train - ?*) before studying the intact word pair (e.g., *train - caboose*) improved later recall of the correct target word (e.g., *caboose*) (Kornell et al., 2009). Unsuccessful retrieval attempts are thought to benefit learning by activating semantically related concepts (e.g., track, engine, station) to the cue word (e.g., *train*), thereby enhancing encoding of the correct target word (e.g., *caboose*) (Grimaldi & Karpicke, 2012). These errors can act as mediators linking the cue and target together and providing more cues for retrieval (Pyc & Rawson, 2010). Although we did not observe a learning benefit in the search condition in Experiments 4a and 4b, this does not necessarily rule out the possibility that participants in the search condition tried to guess the English translation before finding it online. When errors are not related to the target answer, generating incorrect guesses does not benefit recall memory (Metcalfe, 2017), like when participants are forced to guess the target word in an unrelated word pair (e.g., *pillow - leaf*) (Huelser & Metcalfe, 2012; Grimaldi & Karpicke, 2012) or the answers to completely unfamiliar questions (e.g., *Where is Disko Island?*) (Kang et al., 2011).[5] In this case, the link between the Swahili and English word pairs (e.g., *pamba - cotton*) is arbitrary and unknown to non-Swahili speakers (i.e., seeing *pamba* does not automatically activate any concepts related to cotton). If participants were attempting to guess the answers while looking up the translations, their guesses would likely be based on physical characteristics of the word (e.g., spelling), rather than any semantic relationships that would enhance subsequent encoding of the correct translation. For instance, a participant in the search condition might automatically guess that "panda" is the English translation of the Swahili word "*pamba*" solely based on its spelling before finding the correct translation "*cotton*" online.

Experiment 5 aims to test this potential explanation for the reduction in confidence by adding a new reveal condition that imposes a brief delay between seeing

---

[5] There is evidence that incorrectly guessing the translations of foreign vocabulary words can improve learning; however this learning benefit was only observed on recognition based assessments (i.e., multiple-choice tests) not recall or cued-recall tests (Potts & Shanks, 2014; Potts et al., 2019).

the Swahili word and receiving the English translation. Similar to participants in the search condition, participants in the reveal condition also have the opportunity to covertly guess the English translation before receiving it. If the judgments of learning and cued-recall test performance of participants in the reveal condition resemble the results of participants in the search condition, this would support the idea that attempting to guess unknown answers while searching reduces confidence in learning and does not lead to enhanced learning.

3.2 Experiment 5

3.2.1 Method

3.2.1.1 Participants

Three-hundred and sixty-seven Duke undergraduates participated for course credit in Experiment 5. Thirty-four participants were excluded (15 participants reported looking up answers during the test, 19 failed to follow task instructions). Data from the remaining 333 Duke undergraduates (215 female, 107 male, 3 non-binary, 8 unreported; *M* age = 19.02 years) were analyzed. English was the reported as the native language by 84.7% of participants, 12.9% reported a native language other than English, 0.6% reported being natively bilingual with English and another language, and 1.8% did not report a native language. Importantly, no participants reported speaking Swahili.

3.2.1.2 Design

This experiment had three between-subjects conditions: control, search, and reveal.

3.2.1.3 Materials

Experiment 5 included the same 20 Swahili-English word pairs as Experiment 4b (see Appendix C for the complete list).

3.2.1.4 Procedure

The procedure for Experiment 5 is similar to that of Experiments 4a and 4b with the exceptions that (a) Experiment 5 includes a new reveal condition and (b) participants were asked questions about their learning preferences and predictions after the test. Participants in the reveal condition were given a Swahili word (e.g., *wingu - ???*) and then experienced a 2-second pause before revealing the English translation (e.g., *wingu - cloud*). A delay duration of 2 seconds was selected based on the difference in the median time spent per word-pair in the learning phase between the control and search conditions in the previous experiment (Experiment 4b: control *Mdn* = 7.06 s, search *Mdn* = 9.17 s) and on an established procedure that explicitly instructed participants to covertly recall a previously learned English translation of a given Swahili word before presenting the intact Swahili-English word pair (Carpenter et al., 2008).

After completing the cued-recall test, participants were asked if they had looked up any of the answers using an outside source during the test. Any participants who reported looking up answers during the test were excluded. After the test, participants were then asked to imagine the following scenario:

> In one foreign language classroom, students are studying the translations of foreign vocabulary words. **Some students are given the translations, while other students are asked to look up the translations themselves.** All students take the same test.

Based on this fictional classroom scenario, participants were asked 1) to predict which group of students would learn the translations the best (those given the translations, those who looked them up, or no difference), 2) which way they would prefer to learn (being given the translations, looking them up themselves, or no preference), and 3) which way they think would help them remember better (being given the translations, looking them up themselves, or no difference). See Appendix D for the full list of questions and answer choices. At the end of the study, participants filled out their demographic information and were debriefed.

3.2.2 Results

The alpha level for all statistical tests was set at .05, and planned contrasts tested whether judgments of learning, test performance, and confidence ratings at test differed by condition.

3.2.2.1 Judgments of Learning

A one-way between-subjects ANOVA on participants' aggregate judgments of learning revealed significant differences between conditions, $F(2, 330) = 6.65$, $p = .001$, $\eta^2 = 0.04$. The relevant data appear in Figure 4. Participants in the search condition (M =

4.39, $SD = 3.24$) expected to recall fewer English translations at test (out of 20) than either participants in the control condition ($M = 5.77$, $SD = 3.22$, $t(220) = 3.20$, $p = .002$, $d = 0.43$) or the reveal condition ($M = 5.68$, $SD = 3.03$, $t(220) = 3.06$, $p = .002$, $d = 0.41$). There was no significant difference in aggregate judgments of learning between the control and reveal conditions, $t(220) = 0.24$, $p = .813$, $d = 0.03$.

3.2.2.2 Performance on Cued-Recall Test

We conducted a one-way between-subjects ANOVA on the number of English translations participants correctly recalled on the cued-recall test. The relevant data appear in Figure 5. Homogeneity of variance was violated according to Levene's Test ($p = .030$), so we report the adjusted Welch's F-statistic. The one-way ANOVA revealed significant differences in the number of translations correctly recalled between conditions, $F(2, 218.79) = 9.10$, $p < .001$, $\eta^2 = 0.05$. Participants in the search condition ($M = 4.07$, $SD = 3.35$) recalled fewer English translations than either participants in the control condition ($M = 6.09$, $SD = 3.97$, $t(214.00) = 4.09$, $p < .001$, $d = 0.55$) or the reveal condition ($M = 5.45$, $SD = 3.81$, $t(220) = 2.86$, $p = .005$, $d = 0.38$). There was no significant difference in the number of English translations correctly recalled between the control and reveal conditions, $t(220) = 1.22$, $p = .222$, $d = 0.16$.

3.2.2.3 Confidence Ratings at Test

We also conducted a one-way between-subjects ANOVA on participants' confidence ratings for the translations they correctly recalled at test.[6] For ease of interpretation, the original scale values that ranged 1 (*not at all confident*) to 5 (*very confident*) were recoded to range from 0 (*not at all confident*) to 4 (*very confident*). The one-way ANOVA revealed significant differences in confidence ratings at test, $F(2, 317) = 4.06$, $p = .018$, $\eta^2 = 0.03$. Participants in the search condition ($M = 2.42$, $SD = 1.19$) were less confident in the accuracy of their correct answers on the cued-recall test than either participants in the control condition ($M = 2.83$, $SD = 1.03$, $t(211) = 2.68$, $p = .008$, $d = 0.37$) or the reveal condition ($M = 2.74$, $SD = 1.08$, $t(214) = 2.04$, $p = .043$, $d = 0.28$). There was no significant difference between the control condition's and search condition's confidence in the accuracy of their correct answers on the cued-recall test, $t(209) = 0.65$, $p = .516$, $d = 0.09$.

3.2.2.4 Predictions & Preferences

We conducted chi-square tests of independence to determine if participants' learning preferences and predictions were influenced by their condition (control, reveal, search). The relevant frequency data for participants' choices by condition appear in Table 1. Participants' predictions for how other students would learn the translations

[6] This analysis only included the confidence ratings from participants who correctly recalled at least one English translation at test.

best did not differ by condition, $\chi^2(4, N = 333) = 7.67$, $p = .104$. Participants' own preferences for how they would like to learn the translations also did not differ by condition, $\chi^2(4, N = 333) = 3.58$, $p = .466$. However, the method participants thought would help them remember more translations did significantly differ by condition, $\chi^2(4, N = 333) = 12.61$, $p = .013$. A residual analysis with a Bonferroni correction (adjusted alpha level = .006)[7] revealed that the participants in the search condition were disproportionately more likely to predict that they would remember more translations if they were given the translations ($p = .002$) and disproportionately less likely to predict better memory if they looked up the translations themselves ($p = .003$) relative to participants in the control and reveal conditions. This result likely reflects the reduced confidence reported by participants in the search condition who expected to recall fewer translations after looking them up themselves.

---

[7] Based on the published method in (Beasley & Schumacker, 1995) and (MacDonald & Gardner, 2000)

Table 1: The response frequency of participants' learning preferences and predictions by condition in Experiment 5. See Appendix D for the full list of questions and answer choices.

	Question 1: Prediction for best student learning		
	Receiving translations	Looking up translations	No difference in learning
Control	15	89	7
Search	25	74	12
Reveal	13	88	10
	Question 2: Preference for own learning		
	Receiving translations	Looking up translations	No preference
Control	49	50	12
Search	62	40	9
Reveal	58	45	8
	Question 3: Prediction for best memory for self		
	Receiving translations	Looking up translations	No difference in memory
Control	21	78	12
Search	35	59	17
Reveal	16	77	18

3.3 Discussion of Experiments 4a, 4b, & 5

Experiments 4a, 4b, and 5 investigated how using an online search tool (i.e., Google Translate) to learn English translations of Swahili vocabulary words affects people's metacognitive judgments of learning and their learning outcomes as measured by performance on a cued-recall test.

Across these three experiments, people who looked up the translations online were less confident in their learning, expecting to recall fewer words than people who were given (either immediately or after a brief delay) the same translations to learn.

Much of the research investigating the relationship between internet searching and metacognitive judgments has observed boosts in confidence (Fisher et al., 2015) including our findings when participants searched Google for explanations in Experiments 1-3. However, the results of Experiments 4-5 suggest that online searching does not always boost people's confidence. Although some search environments include features that increase feelings of fluency (e.g., snippets that prime the target answer), other search environments, like Google Translate, do not. Taken together, these findings support the idea that metacognitive judgments are not influenced by the act of searching itself, but by the cues available in different types of search environments.

While searching for translations online reduced people's confidence in their learning, low confidence in learning is not always associated with negative learning outcomes. When learners engage in strategies that create desirable difficulties, they initially report lower levels of confidence; however, they ultimately perform better on learning assessments, because desirable difficulties promote processes that enhance encoding and promote better long-term memory (R. A. Bjork, 1994; E. L. Bjork & Bjork, 2011). For instance, when engaging in an effective study strategy, like retrieval practice, learners tend to make more mistakes and perform worse on an immediate test; however, learning is substantially better in the long run (Roediger & Karpicke, 2006).

Given the mixed results around performance on the cued-recall test, we meta-analyzed the difference in the number of English translations correctly recalled between the control and search conditions in Experiments 4-5 using 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). Our mini meta-analysis confirmed that, across the three experiments, participants in the control condition, who were given the translations during the learning phase, recalled more of the English translations on the cued-recall test than participants in the search condition, who looked up the translations online during the learning phase, $Z = 5.79$, $p < .001$, $d = 0.44$, 95% CI [0.29, 0.58].

Despite participants in the search condition presumably exerting more effort in the learning phase by looking up the translations themselves and reporting lower confidence in their learning, they performed worse on the learning assessment. It is possible that a learning benefit from searching would only emerge after a delay, in the same way that the benefits of retrieval practice become evident in the long-term (Roediger & Karpicke, 2006; E. L. Bjork & Bjork, 2011). Nonetheless, looking up information using an online translator does not appear to promote any desirable encoding and retrieval processes that benefit learning in the same way desirable difficulties do (E. L. Bjork & Bjork, 2011; R. A. Bjork & Bjork, 2020), making it unlikely for a long-term learning benefit from searching to emerge.

Although learners do not always know what is best for their learning, Experiment 5 sheds light on participants' intuitions about what they believe is most helpful for learning (see Table 1 for response data and Appendix D for the full list of questions and answer choices). When asked which method of learning would lead to the best learning for students in a hypothetical classroom, 75% of participants predicted better learning outcomes for students who looked up the translations themselves, 16% of participants predicted better learning outcomes for students who were given translations, and 9% of participants expected equal learning outcomes regardless of how students acquired the to-be-learned translations. Despite 3 out of 4 participants expecting a learning benefit of looking up translations online, our results do not support their intuition. When asked which method would help themselves remember the most translations, the majority (64%) still predicted looking up translations would benefit memory the most; however, participants in the search condition were disproportionately less likely to predict a memory benefit from looking up translations and were disproportionately more likely to expect that receiving the translations would aid memory. This change of heart is likely due to participants in the search condition reflecting on their own experience looking up translations online in the learning phase and how that reduced their overall confidence in learning. When participants were asked how they would prefer to learn translations themselves, 51% of participants preferred to be given translations, 40% preferred to look up translations, and 9% reported no preference between the two options. Interestingly, only 40% reported preferring to look up translations themselves to learn, despite the majority of participants (75%) predicting that looking up translations would benefit learning more.

This apparent mismatch in participants' preferences may suggest an inclination for less effortful learning strategies, which aligns with our previous work showing that students preferred writing open-book essays to closed-book essays (Arnold et al., 2021).

Critically, the reveal condition's results in Experiment 5 do not seem to support the hypothesis that unsuccessfully guessing the translations of the unfamiliar Swahili words before finding the answers reduces confidence when participants look up translations themselves. It is possible that the delay between seeing the Swahili word and receiving the English translation in the reveal condition was not long enough to spontaneously prompt participants to attempt to guess the translation. To definitively rule out this explanation, a follow-up experiment, modeled after similar designs in the literature (Grimaldi & Karpicke, 2012; Smith et al., 2013; Kornell et al., 2009), should compare the control and search conditions to a new, forced-guess condition in which participants are required to explicitly guess the translation before seeing the correct translation.

An alternative explanation is that the act of looking up the translation itself is distracting enough to disrupt the encoding of the to-be-learned translation. In order to look up translations, participants in the search condition had to navigate to a different tab in their internet browser to enter the Swahili word into Google Translate and then return to their original browser tab and enter the English translation in the designated textbox. Previous research has shown how multitasking with technology (e.g., laptops, smartphones) leads to distraction and worse performance outcomes (Sana et al., 2013; Wood et al., 2012; Chen & Yan, 2016; Wiradhany & Koerts, 2021). Indeed, individuals appear to be aware of the costs of dividing attention across tasks, predicting worse performance when engaging in two tasks at once (Finley et al., 2014). Accordingly, the continual switching between tasks in the search condition potentially distracted the learner's attention away from encoding the to-be-learned translations, resulting in both lower judgments of learning and worse performance on the cued-recall test. To test this distraction hypothesis, a follow-up study could add a new distraction condition that requires participants to complete a distraction task (e.g., solving arithmetic problems) in the learning phase to see if adding distraction also reduces learning predictions and performance as in the search condition.

4. Solving Word Searches

The purpose of Experiments 6 and 7 is to isolate any effects on metacognitive judgments from the act of searching in a constrained and non-informative environment. Specifically, we investigate how searching for a target word in a word search puzzle affects how well people think they can define the target vocabulary word. Finding a word in a word search does not provide any additional information about the word (e.g., definition, part of speech, origin, affect), and an individual's ability to locate the target word has no bearing on how well they could define the word. Unlike Experiments 1-5 where participants gain information through searching (i.e., either answers to explanatory questions or English translations of foreign words), participants in Experiments 6 and 7 do not receive any information about the target word's definition.

How does the action of searching affect people's assessments of their own knowledge when there is no knowledge to be gained through searching? There is evidence that active involvement in a task can inflate people's judgments and lead to overestimations of their abilities. For example, the IKEA effect shows that individuals place a higher value on objects that they create themselves (e.g., assembling IKEA furniture, building LEGO sets, making origami animals) (Norton et al., 2012). Not only are consumers willing to pay more for the product they assembled themselves, they also report higher feelings of competence after building an object themselves compared to people who were given but did not build their own object (Mochon et al., 2012). Once a task is successfully completed, individuals also report increased feelings of self-efficacy (Bandura, 1977). If actively and successfully solving a word search puzzle also increases people's feelings of competence, it is possible that these feelings could mislead people to inflate their metacognitive estimates of the depth of their vocabulary knowledge – even though their ability to solve a word search is unrelated to their knowledge of vocabulary definitions.

In two experiments, we tested how actively engaging in a non-informative search, by solving a word search puzzle, would affect people's metacognitive judgments of how well they could define the target word they found. While the ability to successfully locate a target word in a word search has no bearing on one's knowledge of the word's meaning, Experiments 6 and 7 investigate how this extraneous searching task influences people's estimates of their knowledge of vocabulary word definitions.

4.1 Experiment 6

4.1.1 Method

4.1.1.1 Participants

Four-hundred and thirty-two Duke undergraduates participated in Experiment 6 for course credit. Twenty-two participants were excluded (19 reported looking up answers on the test, 3 failed to complete the study). Data from the remaining 410 Duke undergraduates (257 female, 147 male, 2 non-binary, 4 unreported; *M* age = 19.33 years) were analyzed.

4.1.1.2 Design

This experiment had two between-subjects conditions: active and passive.

4.1.1.3 Materials

We selected 12 GRE vocabulary words that were rated as difficult by The Economist GRE Tutor (*Most Common GRE Vocabulary*, 2017) and included 6 nouns and 6 adjectives. See Appendix E for the complete list of words. For the multiple-choice definition knowledge test, we selected another 12 GRE filler vocabulary words from The Economist GRE Tutor that also included 6 nouns and 6 adjectives. All 12 target words and 12 filler words were matched with a plausible lure GRE word of the same part of speech (e.g., nouns with nouns).

For each of the 12 target words, we generated an individual word search (10 letters by 10 letters) using the online Puzzlemaker from Discovery Education (*Puzzlemaker Create Your Own Word Search Puzzle*, n.d.). We created two versions of each word search: one for the active condition to solve and one pre-solved version for the passive condition to view. We then embedded these word searches into the Qualtrics study using Javascript code (Andre, 2018) adapted from Robert Calin-Jageman's word search task (Cusack et al., 2015). See Figure 6 for an example of a word search in each condition.

Active word search

Passive word search

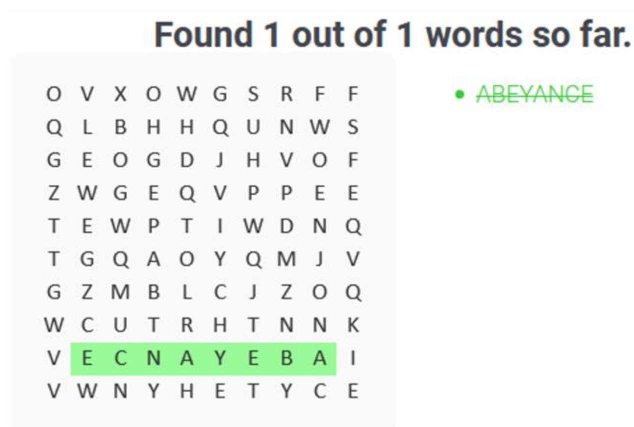

Figure 6: Examples of the word searches used in Experiments 6 and 7. Unsolved word searches (see upper panel) were solved by participants in the active condition (Experiment 6) or on active trials (Experiment 7), while pre-solved word searches (see lower panel) were viewed by participants in the passive condition (Experiment 6) or on passive trials (Experiment 7).

4.1.1.4 Procedure

After giving informed consent, participants completed the first phase of the experiment where they were asked to rate their ability to define each of the target vocabulary words (e.g., *abeyance*) on a scale of 1 (*very poorly*) to 5 (*very well*). Before beginning the first phase, participants in the active condition received specific instructions on how to use their cursor to highlight the target word in the online word search and successfully completed a practice trial. For each trial, participants in the active condition were asked to locate the target word (e.g., *abeyance*) in a word search (see upper panel of Figure 6). After solving the word search, participants rated the difficulty of solving the word search from 1 (*very difficult*) to 5 (*very easy*) and rated their ability to the define the target word (e.g., *abeyance*). For each trial in the passive condition, participants were shown an image of a word search that had been allegedly solved by another person (see lower panel of Figure 6) and were asked to rate their ability to define the target vocabulary word (e.g., *abeyance*) that was already highlighted in the word search before continuing to the next word search.

After completing all 12 trials in the first phase of the experiment, all participants advanced to the test phase. Participants answered 24 multiple-choice questions, including 12 critical questions that tested definition knowledge of the target words presented in the previous phase and 12 filler questions that tested definition knowledge of new GRE words. For each question, participants were given a definition (e.g., *A state of temporary disuse or suspension*) and asked to select the word that best matched the definition. Each question had three answer choices, including the correct answer (e.g., a*beyance*), a plausible lure (e.g., *antipathy*), and a *Don't know* option. Participants were instructed not to look up any definitions during the test and were encouraged to select the *Don't know* option instead of guessing. After completing the study, participants filled out their demographic information and were debriefed.

4.1.2 Results

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

4.1.2.1 Knowledge Ratings

We conducted an independent samples t-test to compare ratings of definition knowledge between the active and passive conditions. The relevant data appear in Figure 7. The Levene's Test for Equality of Variances (p = .017) indicated unequal variances, so we report the adjusted t-statistic. Participants in the active condition, who solved word searches, rated their ability to define the target GRE vocabulary word lower (M = 1.54, SD = 0.54) than participants in the passive condition, who viewed pre-solved word searches (M = 1.67, SD = 0.68), $t(390.14)$ = 2.19, p = .029, d = 0.22. Overall, participants did find the GRE vocabulary words to be difficult with the average rating of definition ability (M = 1.61) being near the low end of the 1 (*very poorly*) to 5 (*very well*) scale.

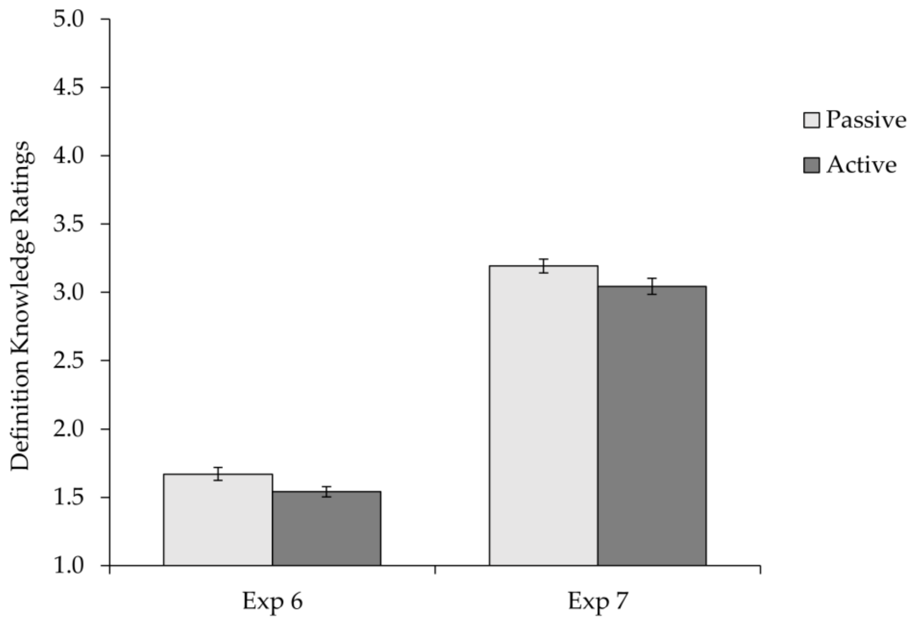

Figure 7: Mean ratings of ability to define target GRE words by condition in Experiment 6 and by trial type in Experiment 7. Error bars reflect standard error of the mean. Note that condition was manipulated between subjects in Experiment 6 and trial type was manipulated within subjects in Experiment 7.

4.1.2.2 Definition Knowledge Check

We conducted an independent samples *t*-test to compare performance on the definition matching test between conditions. The relevant data appear in Figure 8. Specifically, we compared the number of target questions that were answered correctly in the active condition to the number of correct answers in the passive condition. Performance on the definition matching test did not differ between the active condition (M = 3.83, SD = 2.75) and the passive condition (M = 4.04, SD = 2.96), $t(408) = 0.76$, p = .446, d = 0.07), confirming that the difference in definition knowledge ratings between the active and passive conditions did not reflect an underlying difference in knowledge.

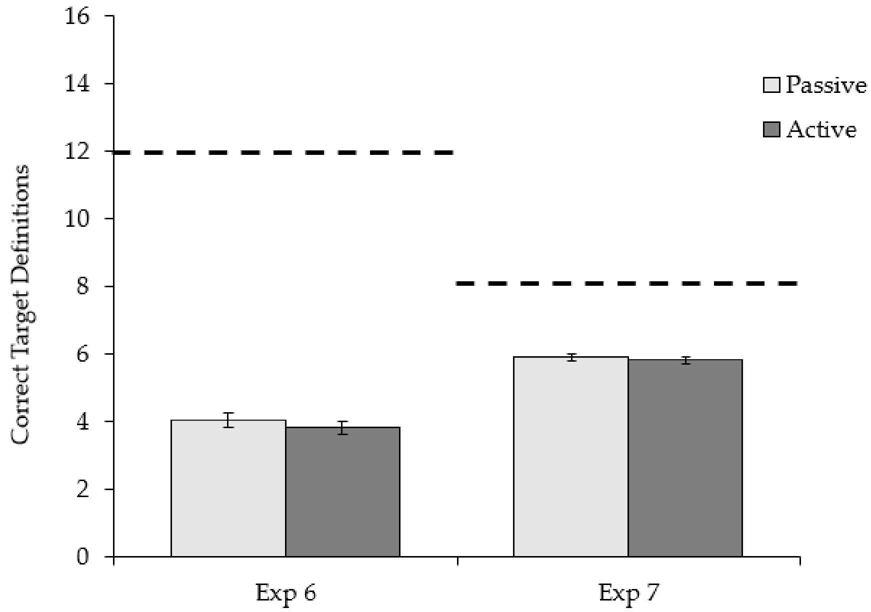

Figure 8: Mean number of correct answers on the definition knowledge test by condition in Experiment 6 and by trial type in Experiment 7. The dashed lines represent the maximum number of correct answers: 12 total in Experiment 6 and 8 passive plus 8 active in Experiment 7. Error bars reflect standard error of the mean. Note that condition was manipulated between subjects in Experiment 6 and trial type was manipulated within subjects in Experiment 7.

4.1.3 Discussion

In Experiment 6, participants who actively solved word searches were less confident in their ability to define the target GRE words they found compared to participants who passively viewed the same pre-solved word searches. Unlike in the previous experiments, participants in Experiment 6 did not search for or receive any information regarding the target word definitions. Both conditions performed equally well on the definition knowledge test, suggesting that the observed difference in self-assessed definition ability is not driven by an actual difference in definition knowledge.

However, performance across both conditions suggests that participants struggled with the difficulty level of the target GRE words: Participants rated their ability to define the target words very low (*M* = 1.61 on a 1 (*very poorly*) to 5 (*very well*) scale) and answered few of the definition matching questions correctly (*M* = 3.94 out of 12) on the test. To avoid potential floor effects, Experiment 7 replaces the difficult GRE vocabulary words with GRE vocabulary words that are rated as medium difficulty.

It is possible that the active condition's reduced confidence in their ability to define vocabulary words is due to their subjective experience during the searching task. While seeing a word search that has already been solved is effortless and requires no additional processing, searching for a target word in a word search requires extra time and effort, and could introduce spurious feelings of disfluency. Participants could interpret these feelings of disfluency as indicative of a lack of definition knowledge. Effects of fluency are relative, so they are typically larger when manipulated within subjects than between subjects (e.g., Wänke & Hansen, 2015; Dechêne et al., 2010). Since any differences in fluency are more noticeable when compared to a baseline, Experiment 7 employs a within-subjects design to amplify any relative differences in participants' subjective experience between active and passive trials.

4.2 Experiment 7

4.2.1 Method

4.2.1.1 Participants

Two-hundred and sixty Duke undergraduates participated in Experiment 7 for course credit. Four participants were excluded (2 reported looking up answers on the test, 2 failed to complete the study). Data from the remaining 256 Duke undergraduates (175 female, 78 male, 2 non-binary, 1 unreported; *M* age = 19.01 years) were analyzed.

4.2.1.2 Design

This experiment manipulated task (active search, passive viewing) within subjects.

4.2.1.3 Materials

We selected 16 new, target GRE vocabulary words that were rated as medium difficulty by The Economist GRE Tutor (*Most Common GRE Vocabulary*, 2017) and included 8 nouns and 8 adjectives. We randomly divided the 16 target words into 2 sets (Set A and Set B) of 8 words each for counterbalancing (see Appendix F for word sets and counterbalances). For the multiple-choice definition knowledge test, we selected another 16 GRE filler vocabulary words (8 nouns and 8 adjectives) that were also rated as medium difficulty by The Economist GRE Tutor and 32 plausible lure GRE words that were matched to the target and filler words on both difficulty and part of speech.

As in Experiment 6, we generated an individual word search (10 letters by 10 letters) for each of the new 16 target words using the online Puzzlemaker from Discovery Education (*Puzzlemaker Create Your Own Word Search Puzzle*, n.d.). We created two versions of each word search: one for solving in the active trials and one pre-solved version for viewing in the passive trials. We then embedded these word searches into the Qualtrics study using Javascript code (Andre, 2018) adapted from Robert Calin-Jageman's word search task (Cusack et al., 2015). See Figure 6 for examples of an active trial word search versus a passive trial word search.

4.2.1.4 Procedure

After giving informed consent, participants were randomly assigned to one of four counterbalances (see Appendix F). All participants completed a block of 8 active trials and a block of 8 passive trials, and the order of blocks was counterbalanced to avoid order effects. For each trial (active or passive), participants were asked to rate their ability to define the target vocabulary word (e.g., *capricious*) on a scale of 1 (*very poorly*) to 5 (*very well*). Half of the participants completed the block of active trials first in which they were asked to solve 8 individual word searches. Participants were given instructions on how to use the cursor to highlight the target word in the word search and successfully completed a practice trial before beginning the block of active trials. For each active trial, participants were asked to rate the difficulty of solving the word search from 1 (*very difficult*) to 5 (*very easy*), in addition to rating their ability to define the target word. The other half of participants completed the block of passive trials first in which they were asked to look at 8 individual word searches that had already been allegedly solved by another person and rate their ability to define the target vocabulary word (e.g., *capricious*) that was already highlighted in the word search. After completing the initial block, participants then completed the other block of either passive or active trials depending on their assigned counterbalance.

After completing 8 active trials and 8 passive trials, participants then advanced to the test phase of the experiment. Participants answered 32 multiple-choice questions, including 16 critical questions that tested definition knowledge of the target words presented in the previous active and passive trials and 16 filler questions about new GRE words. For each question, participants were given a definition (e.g., *Given to sudden and unaccountable changes of mood or behavior*) and asked to select the word that best matched the definition. Each question had 3 answer choices, including the correct answer (e.g., *capricious*), a plausible lure (e.g., a*cerbic*), and a *Don't know* option. Participants were instructed not to look up any definitions during the test and were encouraged to select the *Don't know* option instead of guessing. After completing the study, participants filled out their demographic information and were debriefed.

4.2.2 Results

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

4.2.2.1 Knowledge Ratings

We conducted a paired samples *t*-test to compare ratings of definition knowledge on active trials to those on passive trials. The relevant data appear in Figure 7. Participants rated their ability to define the target GRE vocabulary words lower on active trials where they solved word searches ($M = 3.04$, $SD = 0.92$) than on passive trials where they viewed pre-solved word searches ($M = 3.19$, $SD = 0.83$), $t(255) = 2.61$, $p = .010$, $d = 0.16$. Overall, participants rated the GRE vocabulary words as moderately difficult to define with the average rating of definition ability ($M = 3.12$) landing just above the midpoint of the 1 (*very poorly*) to 5 (*very well*) scale.

4.2.2.2 Definition Knowledge Check

We conducted a paired samples *t*-test to test performance on the critical test questions. The relevant data appear in Figure 8. As the target words were counterbalanced across active and passive trials, we compared test accuracy (defined as the number of definition questions answered correctly) for the target words that were shown on active trials to those target words that were shown on passive trials. Performance on the definition matching test did not differ between words that were shown on active trials ($M = 5.83$, $SD = 1.63$) and those that were shown on passive trials ($M = 5.91$, $SD = 1.68$), $t(255) = 0.87$, $p = .385$, $d = 0.05$, confirming that the difference in definition knowledge ratings between active and passive trials was not due to an underlying difference in knowledge.

4.3 Discussion of Experiments 6 & 7

Experiments 6 and 7 tested how actively engaging in the act of searching affects confidence in one's knowledge. Across two experiments, participants reported less confidence in their ability to define vocabulary words after actively solving a word search puzzle than after passively viewing a pre-solved puzzle.

A possible explanation for this post-search reduction in confidence is that the act of searching involves additional cognitive effort, which may introduce feelings of disfluency. While fluency is associated with processing that is easy and effortless, processing that is more difficult and effortful can feel disfluent (Diemand-Yauman et al., 2011; Dunlosky & Mueller, 2016). People interpret these feelings of disfluency as a metacognitive cue indicating a lack of skill or knowledge (Magreehan et al., 2016; Rawson & Dunlosky, 2002). Even though the task of solving a word search is irrelevant to a participant's knowledge of a word's definition, any feelings of friction and disfluency created by solving the puzzle could bleed into participants' metacognitive judgments, resulting in lower knowledge ratings. Although participants who solved word searches rated the puzzles as fairly easy to solve [Experiment 6: $M = 4.29$; Experiment 7: $M = 4.25$ on a scale of 1 (*very difficult*) to 5 (*very easy*)], solving puzzles, even easy ones, could still induce enough feelings of relative disfluency to reduce confidence in one's knowledge. For instance, participants who furrowed their brow (a relatively simple task that is associated with exerting cognitive effort) while answering trivia questions were less confident in the accuracy of their answers than participants who puffed out their cheeks (Alter et al., 2007).

One indicator of disfluency is time spent on task. Prior work has shown a positive correlation with feelings of disfluency and the amount of time spent on a task, such that more disfluent tasks take longer to complete. For instance, as the difficulty of reading a text increased so too did the time participants spent reading (Seufert et al., 2017). Unsurprisingly, participants spent more time solving word searches than viewing solved ones. In Experiment 6, participants in the active condition spent 3.8x as long solving all 12 word searches (*Mdn* = 209.32 s) as participants in the passive condition spent viewing all 12 solved word searches (*Mdn* = 55.26 s). Likewise, participants in Experiment 7 spent 4.1x as long solving all 8 word searches on active trials (*Mdn* = 139.66 s) as they spent viewing all 8 solved word searches on passive trials (*Mdn* = 34.01 s). While not conclusive in itself, the additional time spent solving word searches lends support to the disfluency hypothesis.

More research is needed to test the hypothesis that solving word searches decreases confidence in one's knowledge by introducing feelings of disfluency. For example, a follow-up study could manipulate disfluency by increasing or decreasing the difficulty of solving the word search. If increasing the difficulty of the searching task (e.g., by increasing the puzzle size, using more distractor letters that are shared with the target word, including more difficult solution orientations like backwards and diagonal, using a hard to read font) increases the size of this effect, or if decreasing the difficulty of the searching task (e.g., by reducing the puzzle size, using a different color font to increase the distinctiveness of the target word) decreases or even eliminates the effect, these results would support the hypothesis that the search-induced reduction in confidence is driven by extraneous feelings of disfluency.

5. Conclusions

Overall, our findings suggest that different forms of searching impact how people assess their own knowledge and how they learn new information in differing ways. When searching increases feelings of fluency, it leads to higher confidence in one's knowledge (Experiments 1-3), but when searching decreases feelings of fluency, it reduces confidence in one's knowledge (Experiments 4-7). As for learning outcomes, searching for information online does not appear to improve how well that information is learned (Experiments 1-3). But when the searching task disrupts the encoding of the to-be-learned material, searching impairs how well that information is learned (Experiments 4-5).

5.1 Metacognitive Conclusions

From the nine experiments previously described, the metacognitive findings initially appear inconsistent: In some experiments, searching increased confidence in one's own knowledge (Experiments 1-3), whereas in other experiments, searching led to reduced confidence (Experiments 4-7). Overall, these findings indicate that it is not the act of searching itself but how the process of searching generates feelings of fluency (or disfluency) that influences people's metacognitive assessments of their knowledge and abilities. Specifically, searching appears to inflate confidence in one's knowledge when features in the search environment (e.g., featured snippets) prime the answer and increase feelings of fluency (Experiments 1-3). Conversely, searching seems to reduce confidence in one's knowledge when searching yields feelings of disfluency by either increasing distraction (i.e., switching tabs to look up translations in Experiments 4-5) or adding difficulty through an unrelated but effortful task (i.e., solving a word search puzzle in Experiments 6-7).

These differing effects of searching seem to depend on the cues that are available during searching, which varied across the three studies. For instance, exposure to relevant information increases ease of processing, and these feelings of fluency inform people's metacognitive judgments (Cardwell et al., 2017; Kelley & Lindsay, 1993; Schwartz & Metcalfe, 1992). Accordingly, the level of exposure to related information varied across the different types of searching investigated in Chapters 2, 3, and 4. Searching Google for explanations in Experiments 1-3 provided exposure to additional, informative content, while looking up translations via Google Translate in Experiments 4-5 offered the essential result with no surplus information and solving word searches in Experiments 6-7 did not contribute any relevant information about the word's meaning.

5.2 Learning Conclusions

On the surface, actively searching for information may appear similar to other active learning techniques and effortful strategies (i.e., desirable difficulties) that require additional effort but involve processes known to benefit memory (E. L. Bjork & Bjork, 2011; Freeman et al., 2014; Karpicke, 2012). If searching facilitated deeper, semantic processing of to-be-learned information like other desirable difficulties, we would expect to see a learning benefit. However, the results of Experiments 1-5 do not offer strong evidence that searching for information improved how well that information was learned – if anything searching decreased learning. The mini meta-analysis on Experiments 1-3 confirmed that there was no learning benefit of online searching: Participants who searched for explanations online produced just as accurate explanations as participants who received the same explanations to read. Moreover, the mini meta-analysis on Experiments 4-5 revealed an overall learning detriment of online searching: Participants who looked up translations online recalled fewer translations on the cued-recall test than participants who were given the translations during the learning phase.

In the same way that ineffective study strategies (e.g., rereading, highlighting) feel helpful but are not actually effective for learning, searching the internet for information may feel beneficial for learning – especially when there are cues in the search environment that facilitate feelings of fluency and lead to inflated estimates of knowledge (Experiments 1-3). Indeed, the majority of participants in Experiment 5 predicted that looking up translations oneself would lead to better learning outcomes than being given the same translations. While searching for information feels more active compared to passively receiving to-be-learned information, the act of searching itself does not encourage elaboration, rehearsal, generation, deep semantic processing, or any other desirable difficulties known to benefit learning (Dunlosky et al., 2013; E. L.

Bjork & Bjork, 2011; R. A. Bjork, 1994). Likewise, despite searching taking relatively more time and effort, this additional time and effort is not necessarily spent engaging in effective encoding strategies that are known to improve learning.

5.3 Limitations

Presumably, seeing additional information that is relevant to the answer on a search results page makes it easier to process the answer, while spending extra time and effort to look up foreign words or solve a word search puzzle introduces unnecessary difficulty. However, the influence of fluency in these studies is assumed, rather than directly measured (e.g., by asking participants to rate the difficulty of processing material or using an objective measure of processing like study time allocation or eye tracking reading behaviors) (see Dunlosky & Mueller, 2016). In Experiments 6 and 7, participants rated the difficulty of the word searches they solved; however, participants were not asked to rate the difficulty of the control task (i.e., passively viewing the solved word searches), so we are not able to directly compare participants' subjective experiences of task difficulty.

As all nine studies were conducted online, we cannot be certain that participants truthfully reported whether they looked up test answers using unauthorized sources. It is possible that underreported cheating introduced additional noise into our data on learning outcomes in Experiments 1-5, potentially explaining the inconsistencies in the learning results. As past internet searching increases the likelihood of future internet searching (Storm et al., 2017), participants in the search conditions in Experiments 1-5, who were instructed to look up answers online in the first phase of the experiment, may have also been more likely to use the internet to look up answers during the test compared to participants in other conditions. Nevertheless, even if participants in the search condition are more likely to look up answers online during the test, we have no reason to expect they would also be less likely to report cheating than participants in any of the other conditions.

Additionally, the online nature of the studies prevented us from comparing online searching to physical searching (e.g., looking through a filing cabinet, flipping through an encyclopedia) to determine if digital searching impacts metacognition and learning differently than physical forms of searching. These limitations can be addressed in future follow-up studies.

5.4 Future Directions

This dissertation investigated how different forms of searching affect people's metacognitive judgments and learning outcomes. Building off this work, future research should continue to investigate the cues that exert the most influence during searching by directly manipulating fluency (e.g., by varying the difficulty of the search) to further elucidate the role fluency plays in how searching impacts people's estimates of their own knowledge and their subsequent learning performance.

Moreover, many open questions remain about the effects of searching in other judgment domains. If searching impacts people's subjective experiences of fluency, what other judgments might also be affected by searching for information, particularly in online environments? Prior research has demonstrated that information that feels fluent and easy to process is more likely to be judged as more true (Dechêne et al., 2010; Brashier et al., 2020), more likeable (Reber et al., 1998; Lee & Labroo, 2004), more familiar (Schwarz, 2004), and more frequent (Tversky & Kahneman, 1973; Reber & Zupanek, 2002) relative to information that is processed disfluently (Alter & Oppenheimer, 2009).

While this dissertation focused on the effects of searching on metacognitive monitoring processes (i.e., assessing the state of one's cognition), more research is needed to investigate how searching, especially in online settings, affects metacognitive control processes (i.e., using those assessments to guide behavior). For example, if looking up translations online reduces judgments of learning, how would this affect self-regulated learning? Specifically, would learners choose to change how they allocate their search time, alter their chosen search strategy, or adjust when they decide to terminate searching?

While these questions are similar to the questions that metacognitive research has been exploring for the past fifty years, these familiar questions are worth investigating in new environments. For instance, does searching for information online impact metacognition and learning differently than physically searching for information? The internet has revolutionized how people find and learn information by increasing how much information is available, how fast it can be retrieved, and how easy it is to access at any time, from anywhere (see Marsh & Rajaram, 2019).

What the internet has not changed, however, are the core cognitive processes of the mind. As the internet only became mainstream in the 1990s, the timescale is far too short for these types of widescale changes to human cognition to occur. Instead, technology has changed the situations in which these cognitive processes are deployed (see Eliseev & Marsh, 2021). While internet search engines have not changed how people assess their own knowledge and abilities, these online environments may inadvertently provide cues (e.g., how easily related information is processed, how quickly information is retrieved) that influence people's metacognitive judgments. Similarly, online searching has not altered how people learn new information, but aspects of some online environments may distract the learner's attention, disrupting the encoding process and resulting in worse learning outcomes. As technology advances and the available tools to access information continue to evolve, understanding the factors that influence cognition (and under what circumstances) is crucial to predicting human behavior, building more efficient online environments, and designing better tools.

Appendix A

List of the 7 explanatory questions used in Experiments 1-3

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?[8]
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 Fisher, Goddu, and Keil (2015).

[8] We changed the question wording from "*How do zippers work*" to "*How zippers work*" to facilitate finding reliable search results on science.howstuffworks.com.

Appendix B

List of the 20 Swahili-English word pairs used in Experiment 4a

Swahili Word	English Translation
baharia	sailor
fagio	broom
fahali	bull
farasi	horse
godoro	mattress
goti	knee
jibini	cheese
kasuku	parrot
lozi	almond
lulu	pearl
mbwa	dog
mfupa	bone
ndoo	bucket
punda	donkey
theluji	snow
tumbili	monkey
usingizi	sleep
wingu	cloud
yai	egg
ziwa	lake

Note. These Swahili-English translation pairs were selected from previously published norms (Nelson & Dunlosky, 1994).

Appendix C

List of the 20 Swahili-English word pairs used in Experiments 4b and 5

Swahili Word	English Translation
bahasha	envelope
duara	circle
fahali	bull
farasi	horse
gharika	flood
godoro	mattress
jibini	cheese
malkia	queen
mfupa	bone
ndoo	bucket
pamba	cotton
punda	donkey
sahani	plate
samaki	fish
theluji	snow
tumbili	monkey
usingizi	sleep
wingu	cloud
yai	egg
ziwa	lake

Note. These Swahili-English translation pairs were selected from previously published norms (Nelson & Dunlosky, 1994).

Appendix D

Questions about learning predictions and preferences used in Experiment 5

In one foreign language classroom, students are studying the translations of foreign vocabulary words. **Some students are given the translations, while other students are asked to look up the translations themselves.** All students take the same test.

1. Which students in this classroom do you think will learn the translations the best?

    - The students who were given the translations
    - The students who looked up the translations themselves
    - No difference in learning

2. If you were a student in this classroom, how would you prefer to learn the translations?

    - I would prefer to be given the translations
    - I would prefer to look up the translations myself
    - No preference

3. If you were a student in this classroom, which way would help you remember the most translations?

    - I would remember more translations if I were given them
    - I would remember more translations if I looked them up myself
    - No difference in memory

Appendix E

List of the 12 GRE Vocabulary words used in Experiment 6

Word List
Abeyance
Anodyne
Canard
Coterie
Dross
Ersatz
Halcyon
Hermetic
Lacuna
Nadir
Puerile
Venal

Note. All 12 GRE vocabulary words were rated as difficult by The Economist GRE Tutor (*Most Common GRE Vocabulary*, 2017).

Appendix F

List of the GRE Vocabulary word sets used in Experiment 7

Set A	Set B
Barrage	Acumen
Discordant	Capricious
Droll	Credulous
Esoteric	Exemplar
Fetid	Lucid
Impasse	Qualm
Proxy	Syntax
Tenet	Verbose

Note. All 16 GRE vocabulary words were rated as medium difficulty by The Economist GRE Tutor (*Most Common GRE Vocabulary*, 2017) and were randomly assigned to set A or set B.

Counterbalances used in Experiment 7

Counterbalance	First Task	Second Task
Counterbalance 1	Active A	Passive B
Counterbalance 2	Passive B	Active A
Counterbalance 3	Active B	Passive A
Counterbalance 4	Passive A	Active B

References

Alter, A. L. (2013). The Benefits of Cognitive Disfluency. *Current Directions in Psychological Science*, *22*(6), 437–442. https://doi.org/10.1177/0963721413498894

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Anderson, R. C., Goldberg, S. R., & Hidde, J. L. (1971). Meaningful processing of sentences. *Journal of Educational Psychology*, *62*(5), 395–399. https://doi.org/10.1037/h0031625

Andre, Q. (2018). *WordSearchJS* [JavaScript]. https://github.com/QuentinAndre/WordSearchJS

Arnold, K. M., Eliseev, E. D., Stone, A. R., McDaniel, M. A., & Marsh, E. J. (2021). Two routes to the same place: Learning from quick closed-book essays versus open-book essays. *Journal of Cognitive Psychology*, *33*(3), 229–246. https://doi.org/10.1080/20445911.2021.1903011

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

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Beasley, T. M., & Schumacker, R. E. (1995). Multiple Regression Approach to Analyzing Contingency Tables: Post Hoc and Planned Comparison Procedures. *The Journal of Experimental Education*, *64*(1), 79–93. https://doi.org/10.1080/00220973.1995.9943797

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*(2), 211–241. https://doi.org/10.1037/a0022106

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory Predictions Are Based on Ease of Processing. *Journal of Memory and Language*, *28*(5), 610–632.

Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 977–989. https://doi.org/10.1037/0278-7393.15.5.977

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval Fluency as a Metacognitive Index. In *Implicit Memory and Metacognition*. Psychology Press.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68. https://doi.org/10.1037/0096-3445.127.1.55

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In *Metacognition: Knowing about knowing* (pp. 185–205). The MIT Press.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

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

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, *194*, 104054. https://doi.org/10.1016/j.cognition.2019.104054

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

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*(2), 438–448. https://doi.org/10.3758/MC.36.2.438

Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*(1), 107–111. https://doi.org/10.3758/BF03194036

Chen, Q., & Yan, Z. (2016). Does multitasking with mobile phones affect learning? A review. *Computers in Human Behavior*, *54*, 34–42. https://doi.org/10.1016/j.chb.2015.07.047

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Cusack, M., Vezenkova, N., Gottschalk, C., & Calin-Jageman, R. J. (2015). Direct and Conceptual Replications of Burgmer & Englich (2012): Power May Have Little to No Effect on Motor Performance. *PLOS ONE*, *10*(11), e0140806. https://doi.org/10.1371/journal.pone.0140806

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the Italicized): Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Dunlosky, J., & Mueller, M. L. (2016). Recommendations for exploring the disfluency hypothesis for establishing whether perceptually degrading materials impacts performance. *Metacognition and Learning*, *11*(1), 123–131. https://doi.org/10.1007/s11409-016-9155-9

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency Meets Cognitive Load in Multimedia Learning: Does Harder-to-Read Mean Better-to-Understand? *Applied Cognitive Psychology*, *28*(4), 488–501. https://doi.org/10.1002/acp.3004

Eliseev, E. D., & Marsh, E. J. (2021). Externalizing autobiographical memories in the digital age. *Trends in Cognitive Sciences*, *25*(12), 1072–1081. https://doi.org/10.1016/j.tics.2021.08.005

Eliseev, E. D., & Marsh, E. J. (in preparation-b). *Looking up and learning translations online*.

Eliseev, E. D., & Marsh, E. J. (in preparation-c). *Puzzling over a puzzle: Non-informative searching decreases confidence in knowledge*.

Eliseev, E. D., & Marsh, E. J. (in preparation-a). *Understanding why searching the Internet inflates confidence in explanatory ability*.

Ferguson, A. M., McLean, D., & Risko, E. F. (2015). Answers at your fingertips: Access to the Internet influences willingness to answer questions. *Consciousness and Cognition*, *37*, 91–102. https://doi.org/10.1016/j.concog.2015.08.008

Finley, J. R., Benjamin, A. S., & McCarley, J. S. (2014). Metacognition of Multi-Tasking: How Well Do We Predict the Costs of Divided Attention? *Journal of Experimental Psychology. Applied*, *20*(2), 158–165. https://doi.org/10.1037/xap0000010

Finn, B., & Tauber, S. K. (2015). When Confidence Is Not a Signal of Knowing: How Students' Experiences and Beliefs About Processing Fluency Can Lead to Miscalibrated Confidence. *Educational Psychology Review*, *27*(4), 567–586. https://doi.org/10.1007/s10648-015-9313-7

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., & Keil, F. C. (2016). The Curse of Expertise: When More Knowledge Leads to Miscalibrated Explanatory Insight. *Cognitive Science*, *40*(5), 1251–1269. https://doi.org/10.1111/cogs.12280

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. (2021). Information without knowledge: The effects of Internet search on learning. *Memory*, *0*(0), 1–13. https://doi.org/10.1080/09658211.2021.1882501

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gardiner, J. M., Dawson, A. J., & Sutton, E. A. (1989). Specificity and Generality of Enhanced Priming Effects for Self-Generated Study Items. *The American Journal of Psychology*, *102*(3), 295–305. https://doi.org/10.2307/1423051

Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). “Seductive details” and learning from text. In *The role of interest in learning and development* (pp. 239–254). Lawrence Erlbaum Associates, Inc.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*(6), 597–602. https://doi.org/10.3758/BF03202442

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. (n.d.). Retrieved March 20, 2021, from https://www.internetlivestats.com/google-search-statistics/

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*(4), 505–513. https://doi.org/10.3758/s13421-011-0174-0

Hall, C. C., Ariss, L., & Todorov, A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Human Decision Processes*, *103*(2), 277–290. https://doi.org/10.1016/j.obhdp.2007.01.003

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, *5*(2), 215–227. https://doi.org/10.1037/0882-7974.5.2.215

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 22–34. https://doi.org/10.1037/0278-7393.29.1.22

Hertzog, C., Hines, J. C., & Touron, D. R. (2013). Judgments of Learning are Influenced by Multiple Cues In Addition to Memory for Past Test Accuracy. *Archives of Scientific Psychology*, *1*(1), 23–32. https://doi.org/10.1037/arc0000003

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 484–494. https://doi.org/10.1037/0278-7393.14.3.484

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*(4), 514–527. https://doi.org/10.3758/s13421-011-0167-z

Jacoby, L. L. (1978). On Interpreting the Effects of Repetition: Solving a Problem Versus Remembering a Solution. *Journal of Verbal Learning and Verbal Behavior*, *17*(6), 649–667.

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

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, *103*(1), 48–59. https://doi.org/10.1037/a0021977

Karpicke, J. D. (2012). Retrieval-Based Learning: Active Retrieval Promotes Meaningful Learning. *Current Directions in Psychological Science*, *21*(3), 157–163. https://doi.org/10.1177/0963721412443552

Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479. https://doi.org/10.1080/09658210802647009

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

Kelley, C. M., & Jacoby, L. L. (1996). Adult Egocentrism: Subjective Experience versus Analytic Bases for Judgment. *Journal of Memory and Language*, *35*(2), 157–175. https://doi.org/10.1006/jmla.1996.0009

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of Knowing: The Influence of Retrieval Practice. *The American Journal of Psychology*, *93*(2), 329–343. https://doi.org/10.2307/1422236

Kinoshita, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, *17*(5), 563–571. https://doi.org/10.3758/BF03197079

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*(2), 416–428. https://doi.org/10.3758/MC.36.2.416

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*(4), 478–492. https://doi.org/10.1016/j.jml.2005.01.001

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998. https://doi.org/10.1037/a0015729

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The Ease-of-Processing Heuristic and the Stability Bias: Dissociating Memory, Memory Beliefs, and Memory Judgments. *Psychological Science*, *22*(6), 787–794. https://doi.org/10.1177/0956797611407929

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kühl, T., & Eitel, A. (2016). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning*, *11*(1), 1–13. https://doi.org/10.1007/s11409-016-9154-x

Lee, A. Y., & Labroo, A. A. (2004). The Effect of Conceptual and Perceptual Fluency on Brand Evaluation. *Journal of Marketing Research*, *41*(2), 151–165. https://doi.org/10.1509/jmkr.41.2.151.28665

MacDonald, P. L., & Gardner, R. C. (2000). Type I Error Rate Comparisons of Post Hoc Procedures for I j Chi-Square Tables. *Educational and Psychological Measurement*, *60*(5), 735–754. https://doi.org/10.1177/00131640021970871

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning*, *11*(1), 35–56. https://doi.org/10.1007/s11409-015-9147-1

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 609–616. https://doi.org/10.1037/0278-7393.16.4.609

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

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory & Cognition*, *29*(2), 222–233. https://doi.org/10.3758/BF03194916

McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review*, *27*(6), 1139–1165. https://doi.org/10.3758/s13423-020-01762-3

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A Contextual Account of the Generation Effect: A Three-Factor Theory. *Journal of Memory and Language*, *27*(5), 521–536.

McElroy, L. A., & Slamecka, N. J. (1982). Memorial Consequences of Generating Nonwords: Implications for Semantic-Memory Interpretations of the Generation Effect. *Journal of Verbal Learning and Verbal Behavior*, *21*(3), 249–259.

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language*, *43*(4), 652–679. https://doi.org/10.1006/jmla.2000.2720

Metcalfe, J. (2017). Learning from Errors. *Annual Review of Psychology*, *68*, 465–489. https://doi.org/10.1146/annurev-psych-010416-044022

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*(2), 225–229. https://doi.org/10.3758/BF03194056

Mochon, D., Norton, M. I., & Ariely, D. (2012). Bolstering and restoring feelings of competence via the IKEA effect. *International Journal of Research in Marketing*, *29*(4), 363–369. https://doi.org/10.1016/j.ijresmar.2012.05.001

Most common GRE vocabulary: A list organized by difficulty. (2017, September 4). Economist GRE Tutor. https://gre.economist.com/gre-advice/gre-vocabulary/which-words-study/most-common-gre-vocabulary-list-organized-difficulty

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Nelson, T. O., & Dunlosky, J. (1994). Norms of Paired-Associate Recall During Multitrial Learning of Swahili-English Translation Equivalents. *Memory*, *2*(3), 325–335. https://doi.org/10.1080/09658219408258951

Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, *22*(3), 453–460. https://doi.org/10.1016/j.jcps.2011.08.002

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. https://doi.org/10.1016/j.tics.2008.02.014

Perrin, A., & Atske, S. (2021). *7% of Americans don't use the internet. Who are they?* https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/

Peynircioğlu, Z. F., & Mungan, E. (1993). Familiarity, relative distinctiveness, and the generation effect. *Memory & Cognition*, *21*(3), 367–374. https://doi.org/10.3758/BF03208269

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency – Does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*, *44*, 31–40. https://doi.org/10.1016/j.learninstruc.2016.01.012

Pieger, E., Mengelkamp, C., & Bannert, M. (2017). Fostering Analytic Metacognitive Processes and Reducing Overconfidence by Disfluency: The Role of Contrast Effects. *Applied Cognitive Psychology*, *31*(3), 291–301. https://doi.org/10.1002/acp.3326

Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during learning: What is the locus of the effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1023–1041. https://doi.org/10.1037/xlm0000637

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644–667. https://doi.org/10.1037/a0033194

Puzzlemaker Create Your Own Word Search Puzzle. (n.d.). Discovery Education. Retrieved January 4, 2022, from https://puzzlemaker.discoveryeducation.com/word-search

Pyc, M. A., & Rawson, K. A. (2010). Why Testing Improves Memory: Mediator Effectiveness Hypothesis. *Science*, *330*(6002), 335–335. https://doi.org/10.1126/science.1191465

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 69–80. https://doi.org/10.1037/0278-7393.28.1.69

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9*(1), 45–48. https://doi.org/10.1111/1467-9280.00008

Reber, R., & Zupanek, N. (2002). Effects of processing fluency on estimates of probability and frequency. In *ETC. Frequency processing and cognition* (pp. 175–188). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198508632.003.0011

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, *19*(1), 90–138. https://doi.org/10.1016/0010-0285(87)90005-3

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology. General*, *137*(4), 615–625. https://doi.org/10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550–554. https://doi.org/10.3758/PBR.16.3.550

Risko, E. F., Ferguson, A. M., & McLean, D. (2016). On retrieving information from external knowledge stores: Feeling-of-findability, feeling-of-knowing and Internet search. *Computers in Human Behavior*, *65*, 534–543. https://doi.org/10.1016/j.chb.2016.08.046

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

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

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacognition and Learning*, *11*(1), 57–70. https://doi.org/10.1007/s11409-015-9151-5

Sana, F., Weston, T., & Cepeda, N. J. (2013). Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education*, *62*, 24–31. https://doi.org/10.1016/j.compedu.2012.10.003

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, *34*(2), 344–355. https://doi.org/10.3758/BF03193412

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1074–1083. https://doi.org/10.1037/0278-7393.18.5.1074

Schwarz, N. (2004). Metacognitive Experiences in Consumer Judgment and Decision Making. *Journal of Consumer Psychology*, *14*(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2

Schwarz, N. (2012). Feelings-as-information theory. In *Handbook of theories of social psychology, Vol. 1* (pp. 289–308). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n15

Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science*, *45*(2), 221–238. https://doi.org/10.1007/s11251-016-9387-8

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

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

Smith, M. A., Roediger, H. L. I., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1712–1725. https://doi.org/10.1037/a0033569

Song, H., & Schwarz, N. (2008). Fluency and the Detection of Misleading Questions: Low Processing Fluency Attenuates the Moses Illusion. *Social Cognition*, *26*(6), 791–799. https://doi.org/10.1521/soco.2008.26.6.791

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

Storm, B. C., Stone, S. M., & Benjamin, A. S. (2017). Using the Internet to access information inflates future use of the Internet to access other information. *Memory*, *25*(6), 717–723. https://doi.org/10.1080/09658211.2016.1210171

Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning*, *11*(1), 71–88. https://doi.org/10.1007/s11409-015-9146-2

Taylor, A., Sanson, M., Burnell, R., Wade, K. A., & Garry, M. (2020). Disfluent difficulties are not desirable difficulties: The (lack of) effect of Sans Forgetica on memory. *Memory*, *28*(7), 850–857. https://doi.org/10.1080/09658211.2020.1758726

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

Titchener, E. B. (1921). *A Text-book of Psychology*. Macmillan.

Topolinski, S., & Strack, F. (2009). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General*, *138*(1), 39–63. https://doi.org/10.1037/a0014678

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Wänke, M., & Hansen, J. (2015). Relative Processing Fluency. *Current Directions in Psychological Science*, *24*(3), 195–199. https://doi.org/10.1177/0963721414561766

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

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253. https://doi.org/10.1037/0278-7393.19.6.1235

Wiradhany, W., & Koerts, J. (2021). Everyday functioning-related cognitive correlates of media multitasking: A mini meta-analysis. *Media Psychology*, *24*(2), 276–303. https://doi.org/10.1080/15213269.2019.1685393

Wood, E., Zivcakova, L., Gentile, P., Archer, K., De Pasquale, D., & Nosko, A. (2012). Examining the impact of off-task multi-tasking with technology on real-time classroom learning. *Computers & Education*, *58*(1), 365–374. https://doi.org/10.1016/j.compedu.2011.08.029

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*(2), 229–241. https://doi.org/10.3758/s13421-012-0255-8

Biography

Emmaline Drew Eliseev attended Rice University in Houston, Texas and earned a B.A. in psychology and French studies. As an undergraduate, she spent a semester abroad in Nice, France where she studied French, and spent another semester abroad in Copenhagen, Denmark where she studied cross-cultural psychology. At Rice, Emmaline conducted a senior research project with Dr. Carissa Zimmerman examining how online textbooks affect learning. In 2016, Emmaline graduated *summa cum laude*, Phi Beta Kappa, and with Distinction in Research and Creative Work.

Emmaline began her graduate studies in cognitive psychology at Duke University in 2016. She conducted research on human learning and memory with Dr. Elizabeth J. Marsh. She published three journal articles and one book chapter. While at Duke, Emmaline was a Bass Digital Education Fellow (2021-2022), a Preparing Future Faculty Fellow (2019-2020), and a James B. Duke Fellow (2016-2020). She served as a Teaching Assistant for 4 courses, taught her own summer course as the instructor of record, and received the Dean’s Award for Excellence in Teaching. Emmaline earned her M.A. in psychology in 2019 and will graduate with her Ph.D. in psychology and a Certificate in College Teaching in May 2022.