NORTHWESTERN UNIVERSITY

You Could Look it up: Exposures to Inaccurate Information and Online Search

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Learning Sciences

By

Amalia Donovan

EVANSTON, ILLINOIS

June 2020

Abstract

Reading informs our understandings of the world. Information we encounter in both academic and everyday reading situations can be integrated into our general knowledge, influencing our perceptions and decisions. When the information we read is valid, these influences are desirable. However, we are also routinely exposed to inaccurate information which, if influential, would be problematic. Unfortunately, readers have been shown to rely on *both* accurate and inaccurate information contained in what they read to complete subsequent tasks. For the purposes of this dissertation, reliance on inaccurate information is defined as instances in which readers' responses on post-reading tasks reflect pieces of false information with which they were previously presented. Reliance often takes the form of readers directly supplying false information from texts to answer later questions or agreeing with false statements at elevated rates. This phenomenon has proven difficult to attenuate, with few studies demonstrating reliable reductions in the influences of inaccurate information.

The overarching purpose of this dissertation is to investigate how opportunities for online search (i.e., using online resources in the service of identifying and checking valid responses when completing assigned tasks) might influence rates of reliance on inaccurate information. While online search is ubiquitous in daily life, it has rarely figured in empirical research on the influence of exposures to inaccurate information. As such, examining how engagement in online search might impact rates of reliance on inaccurate information provides a critical test of the external generalizability of this phenomenon. Allowing for search should result in more externally valid estimates of the rates at which readers might actually use false information in everyday reading scenarios.

The first chapter of this dissertation offers a literature review of projects providing evidence of people's reliance on inaccurate information, obtained across a range of methodologies that exemplify the scope of the problem. Subsequently, the first chapter moves its focus to studies that utilize the same general procedures and materials that have offered well-replicated demonstrations of the issue, and that will be applied in the four experiments comprising this dissertation. Next, the chapter provides a review of previously tested interventions designed to mitigate reliance on inaccurate information. Finally, the chapter argues for the utility of online search as a means of attenuating the deleterious effects of exposure to inaccurate information, grounded in the review of applicable literature.

The next four chapters each present an introduction, methods, results, and discussion of findings from four experiments examining different facets of inaccurate information use and online search. Experiment 1 provides evidence for the efficacy of online search opportunities as a means of reliably reducing people's use of inaccurate information on post-reading tasks within a general adult population. Experiment 2 provides a replication of these findings with a different sample and experimental context, in addition to supplementing findings with qualitative analyses of participants' search behaviors. These analyses offer insight into the reliability of participants' search reports, how participants went about searching, and the specific ways in which search impacts task performance. Experiment 3 tests a means by which the frequency of participants' search behaviors might be increased to in turn reduce reliance on inaccurate information. Experiment 4 utilized a think-aloud and semi-structured interview protocol to provide an additional, richer qualitative assessment of participants' thought processes as they completed the experimental tasks and engaged in online search. The final chapter summarizes the key findings, implications, and limitations of these four experiments, as well as directions for future work.

Table of Contents

ABSTRACT..... 2

CHAPTER I. LITERATURE REVIEW..... 9

A. Influences of Exposure to Inaccurate Information..... 10

B. Attenuating Reliance on Inaccurate Information..... 14

C. Processes Underlying Reliance on Inaccurate Information..... 19

D. Online Search as a Means of Attenuating Inaccuracy Use..... 21

CHAPTER II. EXPERIMENT 1..... 25

A. Experiment 1 Methods..... 26

1. Participants..... 26

2. Materials..... 26

3. Procedure..... 28

4. Design..... 29

B. Experiment 1 Results..... 29

1. Online search rates..... 29

2. Questionnaire coding..... 30

3. Questionnaire response rates..... 31

4. Inaccurate – lure responses..... 31

5. Accurate responses..... 34

6. Inaccurate – other responses..... 35

7. No answer responses..... 36

C. Experiment 1 Discussion..... 37

CHAPTER III. EXPERIMENT 2..... 46

A. Experiment 2 Methods..... 46

1. Participants..... 46

2. Materials..... 46

3. Procedure..... 47

4. Design..... 47

B. Experiment 2 Results..... 47

1. Online search rates..... 47

2. Questionnaire coding..... 48

3. Questionnaire response rates..... 49

4. Inaccurate – lure responses..... 49

5. Accurate responses..... 51

6. Inaccurate – other responses..... 52

7. No answer responses..... 53

8. Screen recordings..... 54

C. Experiment 2 Discussion..... 59

CHAPTER IV. EXPERIMENT 3..... 69

A. Experiment 3 Methods..... 70

1. Participants..... 70

2. Materials..... 70

3. Procedure..... 71

4. Design..... 71

B. Experiment 3 Results..... 71

1. Online search rates..... 71

2. Questionnaire coding..... 73

3. Questionnaire response rates..... 73

4. Inaccurate – lure responses..... 74

5. Accurate responses..... 75

6. Inaccurate – other responses..... 76

7. No answer responses..... 77

C. Experiment 3 Discussion..... 78

CHAPTER V. EXPERIMENT 4..... 83

A. Experiment 1 Methods..... 83

1. Participants..... 83

2. Materials..... 83

3. Procedure..... 84

4. Design..... 85

B. Experiment 4 Vignettes..... 85

1. Participant 1..... 85

2. Participant 2..... 88

3. Participant 3..... 89

4. Participant 4..... 92

5. Participant 5..... 94

6. Participant 6..... 97

7. Participant 7..... 99

8. Participant 8..... 101

9. Participant 9..... 102

10. Participant 10..... 104

C. Experiment 4 Results Summary & Discussion..... 111

CHAPTER VI. GENERAL DISCUSSION..... 114

A. Review of Experimental Outcomes..... 114

B. Implications..... 118

C. Limitations and Future Directions..... 120

D. Conclusion and Summary..... 122

TABLES..... 123

REFERENCES..... 133

VITA..... 141

List of Tables

Table 1. Sample Story Passages

Table 2. Critical Item List

Table 3. Response Rates as a Function of Search, Experiment 1

Table 4. Response Rates as a Function of Search, Experiment 2

Table 5. Search Codes Applied to Screen Recordings, Experiment 2

Table 6. Response Rates as a Function of Search, Experiment 3

Table 7. Interview Protocol, Experiment 4

Chapter I: Literature Review

The information we read can influence our understandings of the world. This influence can be beneficial when the information we read is accurate, reflecting valid claims and reliable evidence (Kintsch, 1988; Kendeou & O'Brien, 2014). However, we may also at times be exposed to inaccurate information. Unfortunately, inaccuracies also exert an influence on readers' comprehension and performance, as has been demonstrated in text processing and memory experiments utilizing a variety of text topics and methodologies (e.g., Ecker, Lewandowsky, & Tang, 2010; Fazio, Barber, Rajaram, Ornstein, & Marsh, 2013; Fazio & Marsh, 2008; Fazio, Rand, & Pennycook, 2019; Frenda, Nichols, & Loftus, 2011; Gerrig & Prentice, 1991; Johnson & Seifert, 1994; Loftus, 2005; Loftus & Hoffman, 1989; Marsh & Fazio, 2006; Marsh, Meade, & Roediger, 2003; Okado & Stark, 2005; Rapp & Braasch, 2014; Rapp, Hinze, Kohlhepp, & Ryskin, 2014).

Influences of exposure to inaccurate information

A number of studies, for example, have investigated the effects of reading about events for which readers are initially presented with inaccurate causal information that is later discredited. Findings have demonstrated that, when asked to infer the likely cause of the event after reading, participants continue to rely on the discredited cause (e.g., Johnson & Seifert, 1998; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Rich & Zaragoza, 2016). To illustrate, consider the following example from Johnson & Siefert (1994) regarding a police report about a warehouse fire. Participants in this study read one of three experimentally manipulated reports, with all three reports initially suggesting that gas cylinders were to blame for the start of the fire. As the report continued, however, some participants learned that the gas cylinders were actually empty (and therefore were no longer considered a viable cause). Other participants learned about

an alternative cause to the gas cylinders (i.e., flammable rags), while the remaining participants received no discrediting or alternative information with respect to the cylinders. After reading the report, participants were asked a variety of questions about the incident.

Of particular interest here are responses to questions regarding the cause of the fire. One might predict that only participants who received no additional discrediting information would mention the gas cylinders as a cause in their responses. However, all participants included references to gas cylinders as a cause of the fire in their responses, even if they had received discrediting information or a suggested alternative cause. This finding has been replicated with other materials, and termed the *continued influence effect* (CIE), which highlights people's mentions of earlier mentioned information even after it has been discounted. CIEs have proven challenging to attenuate, providing a prominent example of the negative implications of exposure to inaccurate information (e.g., Ecker, Hogan, & Lewandowsky, 2017; Ecker, Lewandowsky, Cheung, & Mayberry, 2015; Ecker et al., 2010; Johnson & Seifert, 1998; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Rich & Zaragoza, 2016; Thorson, 2016).

A second example of people's reliance on inaccurate information comes from the eyewitness memory literature. Studies in this area often employ the *misinformation paradigm* in which participants witness an event (such as a crime depicted in a film clip), and afterwards are presented with inaccurate information about that event. Participants then complete tasks assessing their memory of the initially presented event. Research demonstrates that inaccurate post-event information can contaminate memories of the originally witnessed event, as participants inappropriately include details from the misleading post-event information on subsequent recall tasks (e.g., Loftus, 1975; Loftus, 2005; Zaragoza & Lane, 1994).

Consider the following example from Loftus, Miller, & Burns (1978). In this study, participants viewed slides depicting an accident involving a car and a pedestrian. After viewing the slides, participants answered a series of questions about the accident. Question content was manipulated to either be consistent with details of the event (e.g., asking about the car stopping at a stop sign) or to be inconsistent with details of the event (e.g., asking about the car stopping at a yield sign). Later on, participants were asked to indicate on a recognition task whether they originally viewed the car stopping at a stop sign or at a yield sign. Having previously answered a question containing misleading information (e.g., being asked about the car stopping at a yield sign, when it actually stopped at a stop sign) led to higher error rates on the recognition task with participants exhibiting greater difficulty in remembering what had been originally presented. As with the continued influence effect literature, research on misinformation effects as obtained in the eyewitness memory literature points to undesirable consequences of exposure to inaccurate information (e.g., Blank & Launay, 2014; Chan, Thomas, & Bulevich, 2009; Eakin, Schreiber, & Sergent-Marshall, 2003; Forgas, Laham, & Vargas, 2005; Frenda et al., 2011; Higham, Black, & Luna, 2017; Loftus, 1975; Loftus, 2005; Loftus & Hoffman, 1989; Zaragoza & Lane, 1994).

While the above projects highlight various experiences with inaccurate and contradictory information that can confuse and mislead participants, a more recent set of projects have examined specific instances in which participants are presented with inaccurate information about general knowledge topics. Specifically, these recent studies (1) manipulate information pertinent to real world topics (as opposed to information created for the purposes of an experiment) and (2) manipulate that information in the contents of fictional texts. Studies of this type examine people's *reliance on inaccurate information*, with reliance referring to readers

directly supplying false information from texts to answer later questions or agreeing with false statements at elevated rates.

The methods employed in these studies often involve two major tasks. First, participants read fictional stories containing a range of accurate, inaccurate, and unspecified general knowledge information. Fictional texts are used in these studies as they can readily be modified to incorporate various kinds of accurate and inaccurate information. For example, fictional texts typically include character dialogue, descriptions of scenes and settings, and a range of background information for the story, all of which present opportunities for manipulating the accuracy of information contained therein. Indeed, the fictional texts people read in everyday scenarios often contain information about real world topics that varies in accuracy. Manipulations of information accuracy in fictional texts may therefore be less likely to be perceived by readers as atypical, relative to other modalities subject to more stringent standards. In turn, this helps reduce the likelihood of readers evaluating text contents more rigorously than they might in everyday scenarios.

As a second task, participants complete either validity judgments or short-answer questions related to information contained in the texts after reading. These tasks assess the potential influence of the previously presented inaccuracies on participants' responses, as measured by agreements with or reproductions of those inaccuracies. To illustrate, consider stories containing assertions about various human behaviors and scientific processes relevant to real world phenomena. These assertions represent cases in which the accumulated evidence from both personal experience and outside sources should suggest one particular view or claim is valid, endorsed by both experts and the lay public (Appel & Richter, 2007; Gerrig & Prentice, 1991; Prentice, Gerrig, & Bailis, 1997; Rapp, et al., 2014). After reading inaccurate assertions such as

"Tides are not controlled by the moon's gravitational pull" or "Tooth brushing increases the likelihood of gum disease," participants subsequently have more difficulty judging the validity of these statements (i.e., indicating whether they are true or false) than they do after reading accurate versions of such assertions (e.g., "Tides are controlled by the moon's gravitational pull" or "Tooth brushing reduces the likelihood of gum disease"). These effects emerge despite norming of the assertions indicating that the vast majority of participants should have the requisite prior knowledge to endorse the specific accurate claims (Rapp, et al., 2014).

Effects of exposure to inaccurate information have also been obtained with self-contained, declarative statements that are less prescriptive for behavior or indicative of well understood processes (e.g., Hinze, Slaten, Horton, Jenkins, & Rapp, 2014; Marsh et al., 2003; Marsh & Fazio, 2006). For example, after reading inaccurate declarative statements (e.g., "I was plagued with narcolepsy, I just couldn't get to sleep no matter how hard I tried"), participants subsequently produce more inaccurate responses on related questions (e.g., "What is the medical term for the inability to sleep?") than after reading accurate statements of the same information (e.g., "I was plagued with insomnia, I just couldn't get to sleep no matter how hard I tried") or after reading unspecified information (e.g., "I was plagued with an inability to sleep, I just couldn't get to sleep no matter how hard I tried"). Participants reproduce the inaccuracies they read, both when the information is unfamiliar, as well as when it should be known to be wrong (albeit at a lower rate). These findings indicate that exposures to inaccurate information can negatively impact post-reading performance, even when participants likely possess the background knowledge necessary for validating the information presented, based on knowledge norms (Nelson & Narens, 1980; Tauber, Dunlosky, Rawson, Rhodes, & Sitzman, 2013). Encouragingly, analogous effects emerge following exposures to accurate information, with

participants producing more accurate responses on related post-reading questions after reading accurate statements as compared to inaccurate or unspecified statements. Reliance on accurate, valid information that we read is clearly to our benefit. But as the results reviewed indicate, readers also use inaccurate information. Such use can result in sub-optimal outcomes for readers with respect to post-reading decisions, communications, and conceptualizations.

Attenuating reliance on inaccurate information

To date, a number of interventions designed to attenuate the above mentioned effects (i.e., reproductions of inaccurate information; judgments reflecting an influence of inaccurate information) have been tested. Some interventions attempt to alter the ways in which information from texts is encoded in the hopes that readers will be more likely to notice and discount inaccuracies. For example, Marsh et al. (2003) manipulated the number of times participants read inaccurate information. Participants were randomly assigned to read texts containing potentially inaccurate information once, twice, or not at all. The effects of exposure to inaccurate information were assessed with a general knowledge test containing items pertinent to the falsehoods presented in the stories. Participants' responses to those items in turn served as a metric of influence. The researchers found that rereading inaccuracies actually increased the frequency with which readers reproduced them. This undesirable outcome may have been attributable to repeated exposures increasing the salience of the inaccuracies in readers' memories, either ensuring encoding or enhancing retrieval on the subsequent test.

In a later project, Fazio and Marsh (2008) investigated whether slowing the presentation speeds of stories containing inaccurate information would reduce rates of reliance on those ideas post-reading. They hypothesized that slower presentation speeds would afford participants additional processing time that could in turn promote more evaluative appraisals of the

inaccurate information contained in the stories. Participants were assigned to one of two conditions. In a "fast" condition, participants listened to a set of stories containing inaccurate information at a speed 25% faster than the base rate at which the stories were recorded. In a "slow" condition, participants listened to the same stories at a speed reduced 25% below the base rate. Results from performance on a post-reading general knowledge test did not support the authors' hypothesis. Participants in the "slow" presentation condition actually exhibited increased reliance on inaccurate information relative to their counterparts in the "fast" condition. Similarly to repeated exposures, slow presentations may have increased the availability of inaccuracies in readers' memories.

Further evidence that manipulations intended to enhance detection and rejection of inaccuracies during reading do not in fact reduce later reliance was obtained in Eslick, Fazio, and Marsh (2011). This study tested the possibility that highlighting inaccuracies in stories for participants would make it easier for them to detect and discount the false information. Identifying the problematic information for participants might free up cognitive resources for validation, or make it more obvious that information in the text should be rejected. To this end, participants read a set of fictional stories containing potentially inaccurate statements, some of which were highlighted in red. The researchers then compared reproductions rates for highlighted inaccuracies to rates for non-highlighted inaccuracies. No differences obtained, with participants reproducing both highlighted and non-highlighted inaccuracies at similar rates on the post-reading task.

More recently, Donovan, Theodosis, & Rapp (2018) investigated the influence of interruptions during reading of inaccurate story content. Interruptions during reading were hypothesized to be a potential means of attenuating inaccurate responses as they could (1) afford

participants additional time to evaluate information and (2) disrupt readers' encoding of inaccuracies. Such disruption could in turn render inaccuracies potentially unavailable for subsequent retrieval. Across three experiments, participants were assigned to read an extended story containing inaccurate assertions about the world with or without intermittent interruptions. Participants' performance on a validity judgment task containing items pertinent to the story information was then used to assess potential reliance on inaccuracies. Analyses of the judgment data indicated that, in contrast to the researchers' hypotheses, participants made incorrect judgments at commensurate rates regardless of whether or not they received interruptions during reading.

While the aforementioned studies suggest that reducing reliance on inaccurate information with manipulations designed to change how readers encode text content may be ineffective, other types of reading manipulations have successfully yielded reductions to varying degrees. Marsh and Fazio (2006), for example, assessed two particular approaches. In their first experiment, the researchers assessed the efficacy of providing pre-reading warnings about potential inaccuracies. Participants in an experimental condition were told that "Authors of fiction often take liberties with certain facts or ideas in order to make the story flow better or be more entertaining. Therefore, some of the information you will read may be incorrect" (p. 1142). Participants in a control condition received no such warning. All participants were then presented with fictional texts containing inaccurate statements of general knowledge topics. After reading, participants completed a short-answer test containing items related to the statements presented in the texts. Analyses of participants' responses indicated that the pre-reading warning modestly reduced how frequently participants supplied inaccurate story contents as test answers.

In a follow-up experiment, the authors again tested the utility of a pre-reading warning, but also manipulated the reading difficulty level of the stories. Participants thus read stories written at either a sixth grade or a twelfth grade reading level. This secondary manipulation served to assess the possibility that effects of a warning might depend on the resource demands associated with reading the texts. The researchers hypothesized that it would be more difficult for participants to adequately consider or act on the pre-warning with more demanding texts given the required cognitive resources necessary to comprehend them, in contrast with less demanding texts. Reducing the difficulty level of the texts could also free up cognitive resources for monitoring the accuracy of text content. However, the findings did not support this hypothesis. Participants who read the sixth grade level texts reproduced inaccuracies from the stories at commensurate rates to participants who read the twelfth grade level texts, suggesting that reducing resource demands did not enhance benefits of the pre-reading warning.

One potential explanation for why these experiments did not more substantially reduce the effects of exposures to inaccurate information is that the warnings were presented only once, and prior to reading. The degree to which participants attended to the warning over the course of reading is thus uncertain. In line with this possibility, Marsh and Fazio conducted a third experiment designed to foreground the warning during the reading task. Participants were asked to actively check each sentence of the stories for errors, pressing a key labeled “next” if they judged the sentence to be error free, or a key labeled “error” if they detected an inaccuracy. Participants were somewhat successful as detecting inaccuracies, making more “error” key presses for sentences contained inaccuracies than for sentences that did not. However, the previously obtained patterns of reproductions on the post-reading task were again replicated, even when participants had detected errors during reading.

Rapp et al. (2014) employed a more evaluative detection task to determine whether obtaining greater reductions in inaccuracy use might be possible. Participants read an extended fictional narrative containing a variety of inaccurate statements, but rather than making key presses to indicate error detection while reading, participants were instructed to make direct corrections to any detected inaccuracies. Participants were specifically told that they should physically edit any statements in the story that they believed to be inaccurate. The researchers hypothesized that instructions prompting participants to identify *and* modify inaccuracies in the texts could make accurate understandings more readily available for participants to utilize when completing the post-reading task. Participants indeed made fewer judgment errors (e.g., indicating that inaccurate ideas presented in the story were true) than did participants in a control experiment requiring no active revision of the materials. The editing manipulation helpfully reduced the effects of exposure to inaccurate information, although it did not completely eliminate it.

In addition to interventions specifically seeking to alter how readers encode texts, a variety of other approaches have been tested. For example, Fazio et al. (2013) examined the potential effects of a pre-experiment retrieval task. Participants completed a general knowledge test two weeks prior to reading texts containing potential inaccurate information. The general knowledge test specifically assessed participants' prior knowledge on topics that would be manipulated in the stories. Participants then completed the same general knowledge test after reading to assess their potential reproductions of the inaccurate story content. The researchers predicted that retrieving accurate prior knowledge prior to exposure to inaccurate information would attenuate reproductions by predisposing participants to instead rely on what they already know to be true. However, analyses indicated that retrieving prior knowledge in advance of

exposure to inaccuracies did little to mitigate subsequent use. Participants reproduced inaccurate information from the stories to answer questionnaire items even in cases in which they had correctly answered the same questionnaire items two weeks prior. Moreover, the authors found no differences in rates of reliance on inaccurate information for items previously answered correctly as compared to items participants were unable to answer. Similar effects were also obtained in Rapp (2008), with participants' immediately prior retrieval of accurate information doing little to reduce the deleterious effects of inaccurate exposures. These findings suggest that, at least in some cases, possessing relevant prior knowledge may be insufficient to deter the influence of recently presented falsehoods.

Processes underlying reliance on inaccurate information

A variety of explanations have been offered for the above effects, based not just on findings indicating an influence of inaccurate information, but also based on the negligible reductions attained in many projects. One explanation is based on the view that comprehension necessitates that readers activate prior knowledge while simultaneously encoding episodic memory traces of text content (Rapp, Jacovina, & Andrews, 2014). As newly encountered text information is encoded, relevant concepts in a reader's prior knowledge are activated to varying degrees (Albrecht & O'Brien, 1993; Kendeou & O'Brien, 2014; Kintsch, 1988; Rapp & van den Broek, 2005). These activated associations from long-term memory are then accessed over the course of reading in the service of comprehension (Kintsch, 1988; McNamara & Kintsch, 1996). Competition between newly encoded episodic traces and previously established long-term memory traces occurs as comprehension unfolds, with each trace strengthened or weakened over the course of reading (van den Broek, Rapp, & Kendeou, 2005). The degree to which new traces

are more available than existing traces might speak to the effects obtained in many studies examining reliance on inaccurate information.

Consider how readers might process the following statement, which inaccurately cites Galileo, rather than Copernicus: “Galileo published the idea that the Earth revolves around the sun.” If a reader does not attempt or is unable to assess the veracity of the statement, the new association may become more strongly encoded and available than any pre-existing memory traces relevant to this topic (Hinze et al., 2014). Similarly, to the extent that Galileo may be more familiar to the reader than Copernicus, the new, inaccurate association could be privileged over prior associations. Previous work has demonstrated similar effects, as newly established associations that are more familiar to the reader can outcompete previously established traces for activation (e.g., Storm, 2011).

New associations established during reading may also be more accessible to readers than prior knowledge, generally speaking. Episodic memory traces, even when linked to prior knowledge, are more recently encoded and thus may be more readily available to readers in working memory as compared to memories of information learned previously (e.g., Oppenheimer, 2008). As a result, participants may be especially likely to rely upon those recent encodings when completing tasks involving the same topics. Additionally, the questions and prompts included in post-reading tasks may serve as retrieval cues for recently encoded information, including recently presented inaccuracies (Tulving & Thomson, 1973). For example, when asked to answer, “Who was the first person to publish the idea that the Earth revolves around the sun?”, readers may retrieve and use recent episodic traces pertinent to the topic, despite that those traces may contain potential falsehoods. Taken together, these various processing tendencies may help explain the effects of inaccurate exposures demonstrated in

previous work, as discussed in some recent accounts of this phenomenon (e.g., Marsh, Cantor, & Brashier, 2016; Rapp et al., 2014; Rapp & Donovan, 2017).

Prentice & Gerrig's (1999) dual-process model of reading offers further insight into why readers might reproduce inaccurate information. This model posits that readers' engagement in evaluative processing of information depends on two components. First, readers must be sufficiently *motivated* to expend the time and cognitive resources needed to evaluate the veracity of information they read. Secondly, readers need to be *able* to validate text content, meaning they must possess relevant background knowledge. Consideration of both factors is useful in attempting to explain and mitigate reliance on inaccurate information. Motivation to engage in validation of text content may be lacking in many situations, including but by no means limited to the experimental settings being studied. Ability, or rather a lack thereof, to validate information can likewise influence readers' susceptibility to inaccuracies. For example, if readers are unable to recall or lack confidence in their applicable prior knowledge, their ability to successfully validate text content may be compromised. Moreover, when readers lack relevant prior knowledge all together, or fail to consider relevant prior knowledge that could be useful, their ability to validate may be constrained.

Online search as a means of attenuating inaccuracy use

What kinds of interventions might best be suited to addressing the challenge of attenuating reliance on inaccurate information, given the outcomes of prior work and the explanations offered above? Evidence and accounts of inaccurate information use suggest that accurate information needs to be readily available and easily accessible to readers in order for it to be retrieved and used in lieu of recently seen inaccuracies. One way to promote availability and access to valid information is by granting access to outside resources. While access to

outside resources for purposes like information validation may be characteristic of some research environments, resource use is typically restricted in experimental research in general and in studies on reliance on inaccurate information specifically (e.g., Fazio & Marsh, 2008; Marsh et al., 2003; Rapp & Braasch, 2014; Rapp et al., 2014). Certainly, restrictions can be beneficial (or even essential) for some research questions and objectives. With respect to assessing rates of reliance on inaccurate information, however, the dearth of studies allowing for outside resource access could lead to an *overestimation* of how often such reliance actually occurs.

Consider that in everyday reading scenarios, people typically have access to outside resources (e.g., search engines, media, other people, books) which they could use to assess the veracity of information they read. Online resources in particular are widely and routinely accessed in the service of obtaining information (Lai, Lee, Chen, & Yu, 2017) with internet users identifying search to be the most important activity to engage in online (Ozkara, Ozmen, & Kim, 2016). As such, online search plays a fundamental role in how people learn about and understand the world (Rieh, Collins-Thompson, Hansen, & Lee, 2016; Zhou, 2015). It follows that in everyday reading scenarios, readers might at least occasionally engage in search after reading inaccurate information and differentially rely on those inaccuracies as a result. Limiting outside resource access in experimental protocols could thus unduly limit the generalizability of findings from experiments on inaccurate information use to real world reading.

There are several reasons to believe that providing people with the opportunity to engage in online search might attenuate effects of exposure to inaccurate content. First, engaging in online search can make accurate information more readily available to participants than it would be otherwise, particularly for topics about which they are uncertain or hold little prior knowledge. Moreover, item difficulty has previously been shown to be a strong predictor of information

seeking behaviors in general, and online search specifically (Hao, Wright, Barnes, & Branch, 2016). As such, participants may be especially likely to search for the items for which it would benefit them most (i.e., information that they cannot validate independently). Search could also lead participants to discredit information from previously read texts when they encounter evidence online demonstrating that it is inaccurate. Retrieving information that contradicts statements presented in experimental texts, for example, could lead participants to doubt the veracity of what was just previously read.

In addition to developing more externally valid estimates of reliance on inaccurate information, examining online search engagement could also beneficially contribute to extant understandings of why reliance on inaccurate information occurs. As previously described, accounts of this phenomenon often appeal to the role of information availability in memory. If inaccurate information has been encountered more recently than accurate information, inaccuracies may be more accessible and thus likely to be used in subsequent task completion. Inaccurate presentations need not have this recency advantage, however, if readers have and capitalize on opportunities for information seeking. The current project thus allows for interrogating the role of information availability in reliance on inaccurate information in a reading context that allows readers greater control over the information they can access.

These possibilities, however, assume that people will actually utilize the resources they have access to in the service of ascertaining veracity. That is, simply because people often have the option to utilize outside resources as desired during everyday reading does not mean that they actually capitalize on those resources in the situations of interest here. Search, after all, is often used to meet needs or address goals associated with important outcomes, such as making informed decisions and purchases (Ozkara et al., 2016). Whether and to what extent people

would perceive online search to be a useful or worthwhile activity in the context of interest here remains unclear.

Moreover, simply because search is a ubiquitous aspect of contemporary daily life need not mean that any and all searches will lead to successful outcomes (e.g., obtaining valid, relevant information). In some cases, search may be difficult for users to enact (Dhillon, 2007) or may be performed incompletely, with minimal effort or attention to detail (Rieh, Kim, & Markey, 2012). Changes in rates of reliance on inaccurate information as a function of opportunities for online search would be expected only if participants are willing to expend the time and cognitive resources necessary for successful searches. The utility of online search as a means of attenuating reliance on inaccurate text content thus remains an open question.

Chapter II: Experiment 1

The purpose of Experiment 1 was to assess the extent to which participants would utilize the opportunity to engage in online search to support their post-reading performance, and whether capitalizing on this opportunity might reduce reproductions of inaccurate information. As the use of outside resources and tools has not commonly been a component of research on people's experiences with inaccurate information, we first assessed whether people would actually engage in online search behaviors. While people may sometimes elect to engage in online search when reading, doing so is costly in terms of both time and cognitive resources, and thus may not be especially prevalent, particularly in a low-stakes task.

If, however, people opt to search, doing so could yield benefits that mitigate the effects of exposure to inaccurate information. Recall that contemporary accounts cite availability as an influential factor in those effects, as recently read inaccuracies may be more readily available than previously learned, accurate understandings. The use of outside resources can make accurate content more readily available to rely upon on post-reading tasks. Moreover, the potential benefits of using outside resources need not be limited to instances involving exposures to inaccuracies. For questions that are simply unfamiliar or otherwise difficult, online search could also enhance performance on post-reading tasks, relative to reliance solely on prior knowledge or information included in the stories.

To assess the frequency of online search behaviors and their implications for post-reading performance, we recruited a sample of 216 online participants. Participants read a set of four fictional stories each containing accurate, inaccurate, and neutral statements. All statements pertained to general knowledge topics (e.g., historical events and figures, geography, scientific inventions and discoveries, popular culture) and were previously normed by Tauber et al. (2013)

for their familiarity within the general United States adult population. Such statements were selected for use in the current project not only because of their recent norming and use in related research but also because they appear to be highly searchable. That is, if participants wished to validate responses to any or all items, they could be reasonably be expected to do so successfully.

After reading participants completed a short-answer questionnaire with critical items relating to the statements manipulated in the stories. Half of the participants were given the option to search online during questionnaire completion, and the other half were not. We predicted that participants allowed to search would reproduce inaccurate information at a lower rate, and produce accurate responses at a higher rate, than would participants who were not allowed to search. This was predicated on the idea that engaging in online search would provide participants greater access to accurate and reliable information than they would otherwise have available, and that information being recently available would make increase its utility.

Experiment 1 Methods

Participants

216 participants recruited through Amazon's Mechanical Turk (MTurk) completed the study for $6 payment. All participants identified as native English speakers. Participants ranged in age from 23 to 69 (M = 37.58, SD = 11.03). Due to a program error, gender information was collected only for 66% of the sample. Of those 144 participants, 73 identified as male and 71 as female.

Materials

The study was presented online via Qualtrics survey software. Four short stories were adapted from Marsh (2004) to include general knowledge statements from Tauber et al. (2013). The original Marsh materials used statements from Nelson & Narens, 1980, some of which have

become outdated. Stories pertained to mundane topics, including a summer job, a school science fair, a student starting medical school, and a trip to Europe. Stories spanned fifth to eighth grade reading levels and ranged in length from 77-105 sentences. Table 1 provides sample story passages.

Each story contained a set of eight general knowledge statements, for a total of 32 critical statements. Each statement was presented in one of three counterbalanced types: inaccurate (i.e., "She was from Wilmington, the capital of Delaware"), accurate (i.e., "She was from Dover, the capital of Delaware"), or neutral (i.e., "She was from the capital of Delaware" without a city name specified). Statement type was counterbalanced across three versions of each story, such that each participant saw 8-12 statements of each type. The manipulated statements were embedded in character dialogue and story descriptions that had no bearing on the plot or narrative outcomes.

In addition to the manipulation of statement type, statements were also counterbalanced for difficulty. Half (16) of the experimentally manipulated items were categorized as easy, meaning the accurate idea was likely to be known to participants based on norming from Tauber et al. (2013). The remaining half of the experimentally manipulated items were classified as hard, such that they were unlikely to be known to participants, based on the same norms.

A 74-item short-answer questionnaire assessed people's reproductions of information from the stories. Thirty-two of the items queried information presented in the 32 critical story statements, with the remaining 42 items serving as filler. Each item consisted of a single sentence question (e.g., "What is the capital of Delaware?"), appearing identically across conditions. Table 2 presents the complete list of critical test items.

Participants were instructed to type their answers to each question into blank entry fields provided for each item. The questionnaire instructions did not reference the stories in any way so as not to unduly predispose readers to using information in the stories. Across conditions, the questionnaire instructions differed only with respect to how participants were told to respond if unsure of an answer. In the search allowed condition, participants were instructed, "If you are unsure of an answer, you may either write 'no answer,' or if you wish you can search online for the correct answer. Use whatever search engine or app you like to help you." In the no search allowed condition, participants were instructed, "If you are unsure of an answer, please write 'no answer.' We are interested in what you think is the right answer. Please do not use the internet to search for answers."

Procedure

Participants completed the study in a single session lasting approximately one hour. The experiment was conducted using Amazon's Mechanical Turk, with participants completing the experiment in an environment and with a device of their choosing. Participants first read the set of four fictional stories in a self-paced manner. Stories were presented in full one at a time in a counterbalanced order with participants advancing to the next story by pressing an arrow button. After reading all four stories, participants completed a brief set of word problems as a distractor task lasting for approximately five minutes. Next, participants were presented with the questionnaire instructions for the search allowed or the no search allowed condition, depending on assignment. The 74 short-answer questionnaire items were then presented one at a time. Participants in the search condition only were asked to self-report whether or not they looked up each item by checking boxes marked "yes" or "no." All participants then completed a brief set of demographic items, after which they were debriefed and thanked for their participation.

Design

The design of the experiment was 3 (statement type: accurate, inaccurate, or neutral) x 2 (statement difficulty: easy or hard) x 2 (search condition: search allowed, or no search allowed). Statement type varied within subjects such that all participants saw accurate, inaccurate, and neutral statements in each of the four stories they read. Collapsing across stories, participants saw 8-12 statements of each type. Statement difficulty also varied within subjects, such that half of the experimental items each participant saw were easy and half were hard. Condition varied between subjects with half of the participants (n = 108) assigned to the search allowed condition and the other half (n = 108) assigned to the no search allowed condition.

Experiment 1 Results

Online search rates

As participants in the search allowed condition were permitted to search, but the decision to do so was entirely their own choice, we first examined how frequently participants indicated they had engaged in a search. Participants in the search allowed condition self-reported searching for 17.30% (SD = 17.65) of the 74 total questionnaire items (both filler and critical), and 18.08% (SD = 18.09) of the 32 critical items specifically. Individual search rates ranged from 0.00% to 65.63% of the critical items. We submitted online search rates for the critical statements to analysis using a repeated measures ANOVA, with statement type (i.e., accurate, inaccurate, or neutral) and statement difficulty (i.e., easy, hard) entered as within-subjects variables. A main effect of difficulty indicated that participants were more likely to report searching for hard (M = 32.56%, SD = 32.75) than for easy items (M = 4.67%, SD = 11.24), [F (1, 107) = 127.84, MS = 12.60, $p < .001$, $\eta_p^2 = .54$]. This suggests participants were strategic in their searches, more often seeking out unfamiliar information than double-checking what they likely already knew.

This main effect of statement difficulty was qualified by an interaction with statement type [F (2, 106) = 5.00, MS = 0.88, p = .008, η_p^2 = .04]. Participants more often conducted searches for hard items when related information had been presented in neutral statements (M = 35.24%, SD = 34.86) relative to accurate statements (M = 29.44%, SD = 30.35), (p = .02). Search rates for information in inaccurate statements (M = 33.01%, SD = 32.94) were no different than for neutral or accurate statements (all p's > .27). In contrast, search rates for easy items were similarly low across levels of statement type (all p's > .10). No main effect of statement was obtained (F = 1.7).

Questionnaire coding

Questionnaire responses to all experimental items (N = 6,912) were categorized using a four-category coding scheme. Responses reflecting inaccurate information from the stories were coded as *inaccurate - lure* (e.g., stating that the capital of Delaware is Wilmington, with Wilmington being the specific lure used in the experimentally manipulated stories). This code fit 6.23% of responses in the data set. Responses that did not contain the particular inaccurate information presented but were nonetheless incorrect were coded as *inaccurate - other* (e.g., stating that the capital of Delaware is Delaware City, which constitutes an inaccurate response never stated in any of the experimentally manipulated stories). The *inaccurate - other* code fit 11.18% of responses in the data set. Responses reflecting a correct answer were coded as *accurate* (e.g., stating that the capital of Delaware is Dover, with Dover being the correct city name used in the experimentally manipulated stories). 65.74% of responses in the data set were coded as accurate. Responses that reflected a non-answer were coded as *no answer* (e.g., stating "I don't know" or "no answer" in response to the questionnaire item regarding the capital of Delaware). This code fit 16.13% of responses in the data set. 50% of the responses were coded

by two independent raters, with the remaining half of responses coded by one rater only. Interrater reliability for dual-coded responses was reliably high ($\kappa = .98$), with all disagreements resolved through discussion. All response categories were mutually exclusive and fit the entirety of the data set (i.e., all responses fit into exactly one of the four identified coding categories). Table 3 presents response rates across all coded categories as a function of search.

Questionnaire response rates

We assessed response rates (i.e., the frequency of responses fitting each code) for each of the four coded categories with separate repeated measures ANOVAs. Statement type (i.e., accurate, inaccurate, or neutral) and statement difficulty (i.e., easy, hard) were entered as within-subjects variables. Search condition (i.e., search allowed, no search allowed) was entered as a between-subjects variable. Response rate was then entered as the dependent variable. As indicated previously, we generated hypotheses pertinent to only two of our coded categories - *inaccurate - lure* and *accurate*. Nevertheless, we completed analyses of responses fitting all coding categories and report all findings for this experiment and those that follow. We elected to analyze the data set comprehensively so as to provide a complete picture of how the search manipulations impacted questionnaire performance. Accordingly, we utilized Bonferroni corrections throughout our analyses to account for alpha inflation ($\alpha_{altered} = .05/4 = .0125$).

Inaccurate - lure responses

As the primary objective of Experiment 1 was to assess whether allowing participants to search online might reduce their reproductions of inaccurate information, we first assessed participants' response rates on the questionnaire for the *inaccurate - lure* category. In line with previous work using the same materials and manipulations, a significant main effect of statement type was obtained. Participants were more likely to reproduce inaccuracies as answers to the

questionnaire after reading related inaccurate statements (*M* = 11.10%, *SD* = 17.65) than they were to spontaneously produce those same inaccurate responses after reading related accurate (*M* = 3.12%, *SD* = 8.48) or neutral statements (*M* = 4.63%, *SD* = 10.36), [*F* (2, 214) = 60.88, *MS* = 4.01, p < .001, η_p^2 = .22]. Also replicating previous work, a main effect of statement difficulty obtained, with participants more likely to answer hard items (*M* = 11.42%, *SD* = 16.37) than easy items (*M* = 1.14%, SD = 5.50) with inaccuracies from the stories [*F* (1, 215) = 329.00, *MS* = 23.83, *p* < .001, η_p^2 = .61].

These main effects were qualified by an interaction between statement type and statement difficulty [*F* (2, 214) = 27.00, *MS* = 1.33, *p* < .001, η_p^2 = .11]. For hard items, participants were more likely to reproduce inaccuracies after reading inaccurate statements (*M* = 13.35%, *SD* = 20.46) than to spontaneously produce them after reading neutral statements (*M* = 9.23%, *SD* = 13.09), (*p* < .001). Rates of inaccurate productions following both inaccurate and neutral statements were also higher than following accurate statements (*M* = 5.38%, *SD* = 11.16), (*p*'s < .03). For easy items, participants reproduced inaccuracies more often after reading inaccurate statements (*M* = 2.74%, *SD* = 8.90) than they spontaneously generated those inaccuracies after reading accurate (*M* = 0.00%, *SD* = 0.0) or neutral statements (*M* = 0.32%, *SD* = 2.38), (*p*'s < .001). Rates of spontaneous inaccuracy use for easy items did not differ between accurate and neutral statements (*p* > .99).

Critically in this experiment, a main effect of condition was observed, with participants in the search allowed condition (*M* = 5.17%, *SD* = 10.97) less likely to reproduce inaccurate information on the questionnaire than were participants in the no search allowed condition (*M* = 7.40%, *SD* = 15.10) [*F* (1, 215) = 11.01, *MS* = 0.83, *p* = .001, η_p^2 = .05]. The earlier reported main effects of statement type and statement difficulty were also qualified by interactions with

search condition. Participants in the search allowed condition (M = 8.04%, SD = 13.74) were less likely than participants in the no search allowed condition (M = 14.15%, SD = 20.42) to reproduce inaccuracies after reading inaccurate statements [F (2, 214) = 9.10, MS = 0.60, p < .001, η_p^2 = .04]. Rates of spontaneous inaccurate productions following accurate or neutral statements did not differ across conditions (all p's > .33). Participants in the search allowed condition (M = 9.32%, SD = 13.42) were also less likely than participants in the no search allowed condition (M = 13.52%, SD = 18.64) to produce inaccurate information as answers to hard items [F (1, 215) = 9.026, MS = 0.65, p = .003, η_p^2 = .40]. In contrast, for easy items, rates of inaccurate reproductions were similar across conditions (p = .55).

A three-way interaction between statement, difficulty, and condition was also observed. Participants assigned to the search allowed condition reproduced fewer inaccuracies on hard items after reading inaccurate statements (M = 13.35%, SD = 15.91) than did participants assigned to the no search allowed condition (M = 24.75%, SD = 22.76), [F (2, 214) = 8.45, MS = 0.42, p < .001, η_p^2 = .05]. The overall pattern of results indicates that opportunities for online search can reduce rates of inaccurate responses, particularly for topics that participants were unlikely to possess sufficient background knowledge to validate independently.

Next, looking at the search allowed condition specifically, we compared rates of *inaccurate - lure* responses for questionnaire items that participants self-reported searching for versus items that they did not look up. Given that the majority of participants did not report searching for information related to easy items at all, response rates for easy, searched items could not be calculated for the large majority of participants. We therefore collapsed across levels of difficulty for all comparisons of response rates for searched versus unsearched items. Participants were less likely to use inaccurate information to answer questions for which they

reported searching (M = 0.84%, SD = 3.05) as compared to questions for which they did not report searching (M = 9.10%, SD = 8.69), [t (76) = -7.47, p < .001]. This indicates that the reductions in inaccurate reproductions were not due merely to being assigned to a search allowed condition (i.e., potentially supporting an evaluative mindset or careful attention to the questions), but likely due to engagement in search for particular items. The results overall provide evidence that opportunities for online search can reduce an influence of inaccurate information on post-reading responses, particularly for items that participants lack the requisite background knowledge to address on their own.

Accurate responses

We next completed analogous analyses for responses fitting the *accurate* coding category. A main effect of statement type was observed, with participants producing more accurate responses after reading accurate statements (M = 70.28%, SD = 31.09) than after reading inaccurate (M = 62.18%, SD = 35.52) or neutral statements (M = 65.04%, SD = 34.67), [F (2, 214) = 23.95, MS = 5.44, p < .001, η_p^2 = .17]. A main effect of statement difficulty emerged, with participants providing more accurate responses to easy (M = 87.80%, SD = 17.40) than hard items (M = 43.87%, SD = 32.23), [F (1, 214) = 659.27, MS = 287.69, p < .001, η_p^2 = .75].

These main effects were qualified by a marginally significant interaction between statement type and statement difficulty [F (2, 214) = 3.37, MS = 1.33, p = .035, η_p^2 = .02]. Participants were more likely to provide accurate responses for hard items after reading accurate statements (M = 50.30%, SD = 29.53) than after reading neutral statements (M = 43.12%, SD = 33.37), with rates of accurate responses following both accurate and neutral statements higher than following inaccurate statements (M = 38.19%, SD = 32.65), (all p's < .04). For easy items, accurate response rates were higher following accurate statements (M = 90.27%, SD = 16.19)

than following both neutral (M = 86.96%, SD = 18.19) and inaccurate statements (M = 86.16%, SD = 17.55) (p's < .002), while rates of accurate responses did not differ between neutral and inaccurate statements (p > .99).

A main effect of search condition was also observed. Participants in the search allowed condition produced more accurate responses overall (M = 72.37%, SD = 30.74) than did participants in the no search allowed condition (M = 59.30%, SD = 35.74), [F (1, 215) = 31.82, MS = 22.25, p < .001, η_p^2 = .13]. An interaction between search condition and statement difficulty also obtained, with participants in the search allowed condition (M = 55.66%, SD = 32.43) producing accurate responses for hard items more often than did participants in the no search allowed condition (M = 32.01%, SD = 27.40), [F (1, 215) = 44.41, MS = 19.38, p < .001, η_p^2 = .17]. In contrast, rates of accurate responses for easy items were equitably high across conditions (p = .68). No other significant effects or interactions were observed (all F's < 3.05).

We again compared performance specifically for searched and unsearched items for the search allowed condition. Participants produced significantly more accurate responses for searched items (M = 88.60%, SD = 15.29) than for unsearched items (M = 66.46%, SD = 17.58), [t (76) = 6.24, p < .001]. This further suggests that actual search behaviors drove benefits.

Inaccurate - other responses

We next evaluated response rates for the coded category of *inaccurate - other*. A main effect of statement type indicated that participants generated a greater number of inaccurate - other responses after reading neutral statements (M = 12.54%, SD = 17.93) than after reading inaccurate statements (M = 9.98%, SD = 15.23), while response rates after reading accurate statements (M = 11.32%, SD = 17.04) did not differ from those other statement types [F (2, 214) = 6.38, MS = 0.53, p =.002, η_p^2 = .03]. We also observed a main effect of statement difficulty,

with participants more likely to generate inaccurate - other responses for hard items (M = 16.61%, SD = 19.06) than for easy items (M = 5.95%, SD = 12.00), [F (1, 215) = 22.30, MS = 167.95, p < .001, η_p^2 = .44]. No other significant effects or interactions were observed (all F's < 2.19). Participants across the search allowed (M = 10.60%, SD =16.21) and no search allowed conditions (M = 11.96%, SD = 17.34) performed equitably with respect to their production of inaccurate - other responses.

No answer responses

Finally, for the *no answer* response category, we observed a main effect of statement type. Participants produced more no answer responses after reading neutral statements (M =17.31%, SD = 25.78) than after reading accurate statements (M = 14.59%, SD = 23.82), while rates following presentations of inaccurate statements (M = 16.26%, SD = 25.54) did not differ from either of the other statement types [F (2, 214) = 5.31, MS = 0.44, p = .005, η_p^2 = .02].

We also observed a main effect of statement difficulty, with participants generating more no answer responses for hard (M = 27.60%, SD = 29.39) than for easy items (M = 4.51%, SD = 11.28), [F (1, 215) = 284.33, MS = 91.29, p < .001, η_p^2 = .57]. This main effect of statement difficulty was qualified by two interactions. First, we observed a significant interaction between statement difficulty and search condition, such that participants in the search allowed condition (M = 19.10%, SD = 27.90) produced fewer no answer responses for hard items than did their counterparts in the no search allowed condition (M = 36.10%, SD = 28.40), [F (1, 215) = 33.82, MS = 10.86, p < .001, η_p^2 = .14]. In contrast, rates of no answer responses for easy items were equitably low across the search allowed (M = 4.27%, SD = 11.25) and no search allowed (M = 4.74%, SD = 11.32) conditions (p = .65).

A marginally significant three-way interaction between statement type, statement difficulty, and search condition also obtained. Participants in the search allowed condition (M = 18.42%, SD = 27.57) generated fewer no answer responses to hard items than did participants in the no search allowed condition (M = 39.78%, SD = 28.47) specifically when they had seen neutral statements [F (2, 214) = 4.27, MS = 0.30, p = .02, η_p^2 = .02]. Finally, we observed a main effect of search condition, with participants in the search allowed condition (M = 11.69%, SD = 22.51) generating fewer no answer responses overall as compared to participants in the no search allowed condition (M = 20.42%, SD = 26.70), [F (1, 215) = 17.49, MS = 12.33, p < .001, η_p^2 = .08].

As a follow-up, we again compared performance for searched and unsearched items for participants in the search allowed condition. Participants produced significantly fewer no answer responses for searched items (M = 2.26%, SD = 11.95) than for unsearched items (M = 25.31%, SD = 34.15), [t (76) = -3.43, p = .001]. This again offers evidence that engagement in search supported questionnaire performance.

The results across the four coded categories support the view that granting participants opportunities to search online can improve post-reading task performance. Participants in the search allowed condition were less likely to utilize previously read inaccurate information than were participants in the no search allowed condition. Participants in the search allowed condition were also more likely to provide answers to questionnaire items in general (as opposed to responding “no answer”), and to provide accurate answers to questions.

Experiment 1 Discussion

Studies have demonstrated that readers will rely on inaccurate information presented in texts to complete subsequent tasks. These effects have emerged with a range of materials and test

items (e.g., Ecker et al., 2010; Fazio et al., 2013; Fazio & Marsh, 2008; Frenda et al., 2011; Gerrig & Prentice, 1991; Johnson & Seifert, 1994; Loftus, 2005; Loftus & Hoffman, 1989; Marsh & Fazio, 2006; Marsh, Meade, & Roediger, 2003; Okado & Stark, 2005; Rapp & Braasch, 2014; Rapp et al., 2014) and have proven difficult to attenuate. A primary purpose of this experiment was to assess whether and to what extent opportunities for online search might reduce the influence on inaccurate information. To our knowledge, this is the first such experiment to assess an online search manipulation in the context of exposure to inaccuracies. There are several reasons to predict that search would be useful. Finding relevant online resources can make accurate information more readily available for use in completing post-reading tasks. Searches may also equip readers to discredit information from previously read texts. Discovering that a piece of information presented in a text was inaccurate could also lead readers to suspect the validity of other claims.

Another reason to examine online search pertains to external validity. In everyday reading scenarios, people can consult outside resources (e.g., search engines, online applications, other people) to augment their attempts to comprehend and validate text content. However, participants in lab experiments are usually not permitted to consult outside sources during task completion, calling into question the external validity of previously obtained results (e.g., Fazio & Marsh, 2008; Marsh et al., 2003; Rapp & Braasch, 2014). In the particular case of reliance on inaccurate information, restricting people's use of outside resources could result in an overestimation of how likely they are to rely on inaccurate claims. That is, people might be more likely to rely on inaccuracies under restrictive conditions disallowing consultation of other sources and repositories than they would be otherwise.

Of course, simply because outside resources are frequently *available* does not mean they would actually be *utilized* for the purpose of validating information. People might not engage in search all that frequently given that it involves expending resources (i.e., time, attention) beyond what the task at hand may require or encourage. Motivation to expend those resources on supplemental search might also be unlikely for tasks and topics that are not of the readers' choosing. In addition, outside resources need not necessarily supply accurate information as they can undoubtedly vary in reliability. Moreover, their goodness of fit with the question or task at hand critically determines their utility. The extent to which opportunities for online search might influence rates of reliance on inaccurate information was therefore an open question.

We addressed this question by recruiting a sample of 216 participants through Amazon's Mechanical Turk. All participants read a set of stories containing general knowledge statements, some of which were presented inaccurately. After reading, participants completed a short-answer questionnaire containing items pertinent to the information manipulated in the story contents. Half of the sample was asked not to look up answers to any questionnaire items while the remaining half of the sample was instructed that, if they wished, they could use any online resources to assist them in completing the questionnaire. These participants were then asked to indicate whether or not they searched for each item.

Our first analysis assessed the frequency with which people actually self-reported searching for test items. Participants in the search condition reported looking up 18.08% of critical items, collapsing across statement difficulty. Given that hard items in this experiment are unlikely to be known by participants and that difficulty has been shown to be a strong predictor of online search engagement (Hao et al., 2016), participants could reasonably be expected to look up answers more frequently for hard as compared to easy items. Indeed, participants

reported searching for 31.83% of hard items on average, as compared to only 4.51% of easy items (which participants were expected to be able to answer on the basis of prior knowledge alone given existing knowledge norms). This difference suggests that participants made strategic use of the opportunity to search at least with respect to prior knowledge. That is, participants seemed to expend the resources necessary to search more so in cases in which it would augment their questionnaire performance. Search rates for hard items also varied as a function of statement type, with participants most often searching for hard items that had been presented without any relevant information (i.e., neutrally) in the stories. This interaction suggests that participants utilized search as a means of supplementing or augmenting the information available to them, in this case being any information potentially recalled from the experimental texts or in prior knowledge.

While self-reports of online search suggested that participants were capitalizing on opportunities for outside resource use in an efficient manner, it was critical to ascertain whether and to what extent those efforts were actually effective by assessing their questionnaire performance. If participants in the search allowed condition outperformed participants in the no search allowed condition by reproducing fewer inaccuracies from the stories, providing more accurate answers, generating fewer wrong guesses, and/or forgoing responses on fewer questions, this would indicate that allowing search was beneficial. If, however, performance on the questionnaire for these dependent variables was similar for the two groups, this would suggest the search manipulation was insufficient to improve questionnaire performance.

Response analyses revealed that participants produced significantly more inaccurate information after previously reading inaccurate statements as compared to after reading accurate or neutral statements. This pattern exemplifies one kind of influence of inaccurate information

consistently obtained in the extant literature. Participants were also significantly more likely to reproduce inaccurate information for hard as compared to easy items. These results suggest that susceptibility to inaccurate information is greater in instances in which participants hold little to no relevant prior knowledge.

Recall, however, that participants in the search allowed condition self-reported looking up 18.08% of critical items overall and approximately one third (31.83%) of the hard items. Did these self-reported instances of search translate into improved questionnaire performance with respect to inaccurate responses? The obtained effect of condition provides evidence for benefits. Participants in the search allowed condition reproduced inaccurate information on only half as many trials on average as did participants in the no search allowed condition. Participants in the search allowed condition were also significantly less likely to reproduce inaccurate information on hard items relative to participants who were not allowed to search. Thus, while the standard pattern of reliance was observed here, the search manipulation led to significant reductions in the use of inaccurate information, with follow-up analyses indicating these reductions were specifically attributable to searching for specific items.

Another critical test pertained to whether or not participants in the search allowed condition produced more accurate responses on the questionnaire than did participants in the no search allowed condition. While all participants generally produced more accurate answers after having read accurate as compared to inaccurate or neutral statements in the stories, and for easy as compared to hard items, participants in the search allowed condition significantly outperformed participants in the no search allowed condition in overall items answered accurately. We see additional evidence that self-reported search for items translated into actual performance benefits when specifically looking at hard items, as participants in the search

allowed condition outperformed their counterparts in the no search allowed condition by 23.79%. These findings indicate that participants were quite often successful at utilizing online resources to answer questionnaire items.

We next examined rates of *inaccurate - other* responses, which included any inaccurate responses other than those potentially presented in the stories. Participants, regardless of condition, provided more answers of this type after reading neutral statements than after reading accurate or inaccurate statements. This finding may be attributable to neutral statements lacking any specific information to support answering questions and, in turn, predisposing participants to guessing. Participants also provided more inaccurate - other responses for hard as compared to easy items, regardless of assignment to condition. These patterns are not surprising given that elevated levels of guessing are to be expected when participants have little information to work with from prior knowledge or from recently read texts. However, we might reasonably expect that opportunities for online search, if taken up, could reduce the frequency of wrong guesses. One reason for the lack of such between-condition effects could be that participants in the search condition may have, in at least some cases, sought out outside resources but chose a different response than the one that we coded as accurate.

For example, one test item asked participants to identify the city in which Heathrow Airport is located. We observed an increase in instances of participants in the search condition only citing “Longford” rather than “London” as the answer to this item, with the latter response only coded as correct. Longford, however, is one of the immediately adjacent villages to the airport, such that we could consider this an informed, albeit technically inaccurate response. The lack of between-condition effects observed in this experiment could thus point to an important consideration for the online search manipulation. The efficacy of search relies not only on

participants' routine willingness to engage in a search, but also on which sources they choose to consult, how they consult them, and why. Readers can access and utilize a diverse array of information online, much, but certainly not all, of which should align with our response coding criteria. If searches are ended too soon, or are conducted in an unfocused or lackadaisical way, the answers people obtain might not align with correct responses

Finally, we examined rates of *no answer* responses. Participants in the search allowed condition were significantly less likely to produce no answer responses than were participants in the no search allowed condition, indicating an additional benefit of searching. While participants regardless of condition provided no answer responses more often after reading neutral statements and for hard items, participants in the search allowed condition exhibited these patterns to a lesser degree. These findings, in conjunction with the patterns for *inaccurate - lure* and *accurate* responses, highlight the benefits of using online resources.

Collectively, the findings from Experiment 1 indicate that the search manipulation improved questionnaire performance. Participants in the search allowed condition reproduced fewer inaccurate concepts than did participants in the no search allowed condition. They also answered more questions in general (rather than choosing not to answer), and provided more accurate responses overall. To what might we attribute the benefits associated with online search? Recall that one explanation for the influence of inaccurate information pertains to the availability of recently read falsehoods. In projects like the present experiment, participants are exposed to inaccuracies just shortly before completing the post-reading questionnaire. Inaccuracies that are more readily available in memory, given the short time period from which they were experienced, might increase the likelihood of their being used later, as compared to prior knowledge. Moreover, for hard items that participants are unlikely to know on the basis of prior knowledge

alone, participants may have few other answers readily retrievable from memory. Online search, however, can quickly provide access to valid information, making that information readily available to use for completing the questionnaire. Conducting a search when completing the questionnaire can make accurate information available at exactly the moment it is needed.

Given the relatively short duration of the experimental procedures utilized here, both inaccurate information *and* information gleaned from search may both be accessible to participants. Why might online information win out, even beyond its slight temporal advantage over information from the stories? One possibility pertains to access to source information. The stories in the experiment omitted author details and other information that might afford judgments of source credibility. The search manipulation, however, allowed participants to find information from sources of their own choosing. Self-selected sources may be perceived as more credible than the content of the fictional stories for a variety of reasons, including perceived expertise and trustworthiness (Wathen & Burkell, 2002). A related possibility is that readers who choose to engage in search, and through doing so discover that information from the stories was inaccurate, may be more likely to perceive other pieces of information from the stories as lacking credibility. Engaging in online search for one item could thus have beneficial downstream effects, with participants more likely to discount subsequent story information even without opting to engage in search.

While this experiment provided promising evidence that opportunities for online search can reduce readers' reproductions of inaccurate information, additional replication attempts and further interrogations of search behaviors and associated outcomes were necessary. Given that Experiment 1 was conducted online, our assessments of search behavior relied entirely on participant self-report. Assessing search behaviors directly would allow for a more stringent

examination of the impact search might have on participants' post-reading responses. Experiment 2 thus sought to replicate these findings with a different sample and in a different context. This second experiment also augmented the procedures of Experiment 1 with the collection of screen recordings as an additional form of data. Screen recordings of participants' computer use during the experiment allowed for assessing *how* participants went about engaging in online search, providing a qualitative component lacking from Experiment 1. Investigating participants' search processes themselves could usefully provide insight into the ways in which our search manipulation might influence task performance after reading.

Chapter III: Experiment 2

We next sought to extend findings from Experiment 1 by investigating *how* participants engaged in searches for information through the addition of screen recordings to our experimental procedures. Recall that in Experiment 1, the assessments of online search rested wholly on self-report data as to whether a participant chose to search for any given item. Screen recordings provide a more precise measure of how often participants might search, and allow for examining a participant's particular search activities. To this end, we conducted Experiment 2 in person in a lab setting. This sampling change not only allowed for the collection of screen recordings but also for examining whether effects identified in Experiment 1 would replicate in a different setting and with a different population. Otherwise the experiment used the same materials, manipulations, and experimental conditions as in Experiment 1. We predicted that participants allowed to search online to answer questions would be less likely to reproduce inaccurate information and more likely to respond accurately on the post-reading questionnaire relative to participants who were not allowed to search. This prediction matched that offered in Experiment 1.

Experiment 2 Methods

Participants

96 participants (40 male, 53 female, and 3 participants who did not identify with the gender binary or chose not to report gender identity) were recruited from Northwestern University's Introduction to Psychology subject pool. They each participated in a single session lasting approximately one hour in exchange for course credit. Participants ranged in age from 18-23 (M = 18.81).

Materials

The materials were identical to those used in Experiment 1.

Procedure

The procedure was identical to Experiment 1 with two changes. First, participants completed the study in person rather than remotely. Second, we screen recorded participants' activities. Screen recordings were initiated via QuickTime Player immediately following consent. All participants were informed that screen recordings would be collected.

Design

The design was identical to Experiment.

Experiment 2 Results

Online search rates

As in Experiment 1, we first assessed how frequently participants in the search allowed condition self-reported looking up answers to questionnaire items. On average, participants reported searching for 35.84% (*SD* = 19.71) of the 74 total questionnaire items (both filler and critical) and 36.22% (*SD* = 33.59) of the 32 critical items. Mean individual search rates ranged from 0.00% to 68.75% of the critical items. We submitted search rates for critical items to analysis using a repeated measures ANOVA, with statement type (i.e., accurate, inaccurate, or neutral) and statement difficulty (i.e., easy, hard) entered as within-subjects variables. A main effect of difficulty indicated that participants were more likely to report searching for hard (*M* = 56.55%, *SD* = 31.96) than easy items (*M* = 15.89%, *SD* = 20.23), [F (1, 47) = 72.40, MS = 2.89, $p < .001$, $\eta_p^2 = .61$]. Search rates in Experiment 2 were thus somewhat higher than in Experiment 1, but exhibited the same general pattern, with participants more likely to seek out information about unfamiliar than familiar topics.

We also observed a significant interaction between statement type and difficulty [F (2, 46) = 90.14, MS = 4.70, p < .001, η_p^2 = .66]. Participants were more likely search for hard items after reading neutral statements (M = 62.90%, SD = 33.00) than after reading accurate statements (M = 51.60%, SD = 30.67), (p = .002). Search rates following inaccurate statements (M = 55.16%, SD = 31.77) were no different than following neutral or accurate statements (all p's > .06). This same pattern emerged for easy items at an overall lower rate, with participants more likely to conduct searches after reading neutral statements (M = 20.65%, SD = 23.08) than after reading accurate statements (M = 12.65%, SD = 18.99), (p = .046). Search rates following inaccurate statements (M = 14.36%, SD = 17.75) were no different from the other statement types (all p's > .36). These patterns suggest that participants' search behaviors were influenced by both relevant prior knowledge and recently encountered information, as in Experiment 1.

Questionnaire coding

The same four-category coding scheme from Experiment 1 was used in Experiment 2 (N = 3,072). Responses reproducing the inaccuracies presented in the stories were coded as *inaccurate - lure.* This code fit 5.91% of responses in the data set. Responses reflecting an accurate response were coded as *accurate* and comprised 64.36% of responses. Responses that did not contain the particular inaccurate lure presented but were nonetheless incorrect were coded as *inaccurate - other*, constituting 11.34% of responses. Responses that reflected a non-answer (e.g., "I don't know") were coded as *no answer*, constituting 18.36% of responses. All response categories were mutually exclusive and fit the entirety of the data set (i.e., all responses fit into exactly one of the identified coding categories). As in Experiment 1, 50% of the responses were independently coded by two raters trained in the coding scheme, with the remaining half of responses coded by one rater only. Interrater reliability for dual-coded

responses was reliably high ($\kappa = .98$), with all disagreements resolved through discussion. Table 4 presents response rates across all coded categories as a function of search.

Questionnaire response rates

Following the same analytic procedures employed in Experiment 1, we assessed response rates (i.e., the frequency of responses fitting each code) for each of the four coded categories with separate repeated measures ANOVAs. As in Experiment 1, we generated hypotheses pertinent to only two of our coded categories: *inaccurate - lure* and *accurate*. Nevertheless, we completed analyses of responses fitting all coded categories to provide a comprehensive overview of the data. As in Experiment 1, Bonferroni corrections were applied to account for alpha inflation. Statement type (i.e., accurate, inaccurate, or neutral) and statement difficulty (i.e., easy, hard) were entered as within-subjects variables. Condition (i.e., search allowed, search not allowed) was entered as a between-subjects variable. Response rate was the dependent variable.

Inaccurate - lure responses

A main effect of statement type was obtained, with participants significantly more likely to reproduce inaccurate information after reading it in the text ($M = 13.07\%$, $SD = 17.69$) than they were to spontaneously produce those inaccuracies after reading accurate ($M = 1.92\%$, $SD = 5.60$) or neutral statements ($M = 2.75\%$, $SD = 7.14$), [$F\ (2, 94) = 63.26$, $MS = 3.76$, $p < .001$, $\eta_p^2 = .40$]. We also observed a main effect of statement difficulty, with participants more likely to provide inaccurate responses to hard ($M = 10.44\%$, $SD = 15.32$) as compared to easy questions ($M = 1.39\%$, $SD = 12.53$), [$F\ (1, 95) = 179.41$, $MS = 8.21$, $p < .001$, $\eta_p^2 = .66$]. In addition, we observed a main effect of search condition [$F\ (1,95) = 14.94$, $MS = 0.72$, $p < .001$, $\eta_p^2 = .14$], with participants in the search allowed condition ($M = 4.30\%$, $SD = 9.61$) significantly less likely to

provide inaccurate information to complete questionnaire items than were participants in the no search allowed condition (M = 7.52%, SD = 14.73).

Three interactions qualified these main effects. First, a statement by difficulty interaction was obtained [F (1, 94) = 32.91, MS = 1.29, p < .001, η_p^2 =.26]. For hard items, participants were more likely to reproduce inaccurate information after reading inaccurate statements (M = 21.98%, SD = 19.06) than they were to spontaneously produce those same inaccuracies after reading accurate (M = 3.83%, SD = 7.45) or neutral statements (M = 5.50%, SD = 9.33). At an overall lower rate, participants also produced inaccuracies as answers to easy items more often after reading inaccurate statements (M = 2.74%, SD = 8.90) than they spontaneously generated inaccuracies after reading accurate (M = 0.00%, SD = 0.0) or neutral statements (M = 0.32%, SD = 2.38), (p's < .001). Inaccurate productions did not differ following presentations of accurate and neutral statements for hard (p = .27) or easy items (p > .99). Second, a marginally significant statement by condition interaction indicated that participants in the search allowed condition (M = 9.48%, SD = 13.84) were less likely to reproduce inaccuracies after reading inaccurate information than were participants in the no search allowed condition (M = 16.66%, SD = 20.29) [F (2, 94) = 4.02, MS = 0.24, p = .033, η_p^2 = .04]. Inaccurate productions following accurate or neutral statements did not differ across conditions (all p's > .14). Third, an interaction between condition and difficulty obtained, with participants in the search allowed condition (M = 7.66%, SD = 12.12) producing inaccurate information significantly less often for hard items than did participants in the no search allowed condition (M = 13.22%, SD = 17.57), [F (1, 95) = 8.60, MS = 0.39, p = .004, η_p^2 =.08]. In contrast, inaccurate productions did not differ for easy items across the search allowed (M = 0.95%, SD = 4.00) and no search allowed conditions (M = 1.82%, SD = 7.85), (p = .25). No other significant main effects or interactions were obtained (all F's < 2.47).

As in Experiment 1, we also compared performance on searched and unsearched items for participants in the search allowed condition. Participants produced significantly fewer *inaccurate - lur*e responses for searched items (M = 1.12%, SD = 2.36) than for unsearched items (M = 12.93%, SD = 13.92), [t (42) = -5.74, $p < .001$]. This replicated the benefits of searching observed in Experiment 1.

Accurate responses

For responses coded as *accurate,* we observed a main effect of statement type, with participants producing more accurate responses after reading accurate statements in the stories (M = 71.09%, SD = 30.81) as compared to after having read neutral (M = 64.43%, SD = 32.90) or inaccurate statements (M = 57.55%, SD = 34.05), [F (2,94) = 31.47, MS = 4.54, $p < .001$, η_p^2 =.25]. We also observed a main effect of statement difficulty, with participants more likely to provide accurate responses for easy (M = 84.01%, SD = 19.13) than for hard items (M = 44.70%, SD = 32.30), [F (1,95) = 524.85, MS = 113.01, $p < .001$, η_p^2 =.85]. An interaction between statement difficulty and search condition was also obtained [F (1, 95) = 56.02, MS = 12.06, $p < .001$, η_p^2 =.37]. Participants in the search allowed condition provided more accurate responses for hard items (M = 64.74%, SD = 28.54) as compared to participants in the no search allowed condition (M = 24.66%, SD = 23.20). This also obtained for easy items at an overall lower rate, with participants in the search allowed condition (M = 90.55%, SD = 16.07) providing accurate responses more often than did participants in the no search allowed condition (M = 77.48%, SD = 19.76). This interaction in turn drove a main effect of search condition, with participants in the search allowed condition (M = 77.64%, SD = 26.49) providing more accurate responses overall than did participants in the no search allowed condition (M = 51.07%, SD =

33.60), [F (1, 95) = 81.42, MS = 44.94, p < .001, η_p^2 =.46]. These results replicated Experiment 1. No other significant main effects or interactions were obtained (all F's < 3.03).

We also compared performance on searched versus unsearched items for participants specifically in the search allowed condition. They produced significantly more accurate responses for searched items (M = 89.08%, SD = 9.23) than for unsearched items (M = 68.96%, SD = 16.44), [t (42) = -5.74, p < .001], again replicating Experiment 1.

Inaccurate - other responses

We next examined response rates for the coded category of *inaccurate - other*, which contained any incorrect responses other than the inaccurate information appearing in the stories. We observed a main effect of statement difficulty, with participants significantly more likely to provide inaccurate - other responses for hard (M = 15.93%, SD = 17.82) than for easy items (M = 6.74%, SD = 16.27), [F (1, 95) = 92.01, MS = 8.71, p < .001, η_p^2 =.49]. This main effect was qualified by a marginally significant interaction between statement difficulty and search condition [F (1, 95) = 5.57, MS = 0.53, p = .02, η_p^2 =.06]. Participants in the search allowed condition generated more inaccurate - other responses for hard items (M = 17.23%, SD = 19.14) than did participants in the no search allowed condition (M = 14.63%, SD = 16.36). The opposite pattern was obtained for easy items, with participants in the search allowed condition generating fewer inaccurate - other responses (M = 5.54%, SD = 12.67) than did participants in the no search allowed condition (M = 7.94%, SD = 13.36). These results may be attributable to the diverse information participants in the search condition may have viewed. For hard items, seeing diverse information may pose a challenge in that participants might, at least occasionally, struggle to choose the most suitable response from amongst the various possibilities presented. For easy items, diverse information would not be expected to pose such a challenge as

participants would be more likely to be able to validate response possibilities against their own background knowledge. No other significant main effects or interactions were obtained (all *F*'s < 1.23). As in Experiment 1, we again observed no significant difference in rates of *inaccurate - other* responses across the search allowed (M = 11.38%, SD = 17.23) and no search allowed conditions (M = 11.28%, SD = 15.28).

No answer responses

For the final coded category of *no answer*, a main effect of difficulty obtained [F (1, 95) = 168.92, MS = 29.53, $p < .001$, $\eta_p^2 = .64$], with participants more likely to provide a non-answer for hard (M = 28.93%, SD = 29.44) than for easy items (M = 7.79%, SD = 14.51). This main effect was qualified by two interactions. First, there was a marginally significant statement type by statement difficulty interaction [F (2, 94) = 3.38, MS = 0.25, p = .04, $\eta_p^2 = .08$]. For hard items, participants were significantly more likely to provide non-answers after reading neutral statements (M = 33.00%, SD = 31.61) than after reading inaccurate (M = 26.89%, SD = 27.43) or accurate statements (M = 26.88%, SD = 28.99), (p's < .026). In contrast, no answer responses for easy items did not vary by statement type (all p's > .11). Additionally, we observed an interaction between search condition and statement difficulty [F (1, 95) = 58.65, MS = 10.25, $p < .001$, $\eta_p^2 = .38$]. Participants in the search allowed condition generated fewer non-answer responses to hard items (M = 10.37%, SD = 20.28) than did participants in the no search allowed condition (M = 47.48%, SD = 26.33). Participants in the search allowed condition also produced fewer non-answer responses to easy items (M = 2.83%, SD = 8.45) than did participants in the no search condition (M = 12.76%, SD = 17.36), but at an overall lower rate than for hard items. This interaction drove a main effect of search condition with participants allowed to search (M = 6.60%, SD = 15.97) producing fewer non-answers overall than did participants who were not

allowed condition (M = 30.12%, SD = 27.73), [F (1, 95) = 61.34, MS = 36.67, p < .001, η_p^2 =.39]. No other significant main effects or interactions were obtained (all F's < 2.53).

We also compared performance specifically on searched versus unsearched items. Participants in the search allowed condition numerically produced fewer no answer responses for searched items (M = 1.53%, SD = 4.01) than for unsearched items (M = 6.45%, SD = 13.37), although this difference was not significant. This may be attributable to the low rate of no answer responses overall, with 0.00% being the modal rate for both searched and unsearched items.

Screen recordings

The next analysis focused on participants' behaviors by examining the screen recordings of their searches. While all participants were screen recorded to maximize similarity across conditions, only the recordings of participants assigned to the search allowed condition (n = 48) were examined. Screen recordings from three participants were lost due to QuickTime Player file saving errors. Neither the study procedures nor any questionnaire data collected via Qualtrics were compromised as the errors occurred during the file saving process after participants had completed the study. Accordingly, all data from these participants were included in the study save for their screen recordings. Thus, the results reflect the content of the 45 available recordings.

As a first step, we developed a coding scheme for examining the screen recordings. The coding scheme reflected both a priori understandings of online search processes (e.g., we planned to code literal features of online search processes such as search term types from the outset of the experiment), and categorizations derived from a preliminary review of a subset of participants' screen recordings. 25% of the data set (i.e., twelve randomly selected screen recordings) was dual-coded by two independent raters trained in the coding scheme, with the

remaining screen recordings coded by one rater only. Interrater reliability for dual coded recordings was reliably high ($\kappa = .93$). Table 5 presents a summary of search codes.

The coding scheme contained four general categories, each with several possible sub-codes or specifications. The first category was *validity of search report*. This category assessed whether behaviors evident in the screen recordings matched the self-reports that participants provided for every questionnaire item. Each item ($N = 1{,}440$; 32 items for each of the 45 participants for whom we had screen recordings) was coded as either a *match* (i.e., a participant self-reported "Yes" for a search prompt with a search evident in the screen recording, or "No" with no search evident) or a *mismatch* (i.e., a participant self-reported "Yes" for a search prompt without any search evident in the screen recording for the item in question or vice versa).

We observed 495 searches coded as matches, and seven searches coded as mismatches. This indicates that participants' self-reports were accurate. Of the seven mismatches, six were false negatives for which participant self-reported "No" but actually conducted a search. In all six instances participants supplied their answer *before* conducting the search, and ultimately submitted that same answer, indicating that the search did not change their response. The other mismatch case was a false positive, with the participant self-reporting "Yes" without evidence of a search. This mismatch may have simply been an error. With the six unreported searches to add, and the one non-existent search to subtract from our total, we ultimately observed and applied the remaining coding categories to 500 searches.

The three categories of codes subsequently applied were *search type*, *search destination*, and *search role*, intended to identify features of search behaviors. In contrast to the questionnaire coding scheme, these coding categories were not mutually exclusive. Sub-codes from the three categories were applied to all 500 observed searches. The *search type* code was intended to

identify how participants formatted their search terms, whether as the *exact question* posed for a given questionnaire item (e.g., What is the name of the man who invented the steamboat "Clermont"?), as a *rephrased question* that differed from that of a questionnaire item (e.g., “Who invented the Clermont steamboat?), or as *keywords* (e.g., “Clermont steamboat”). If a participant conducted multiple, unique searches for a given question, we coded as many sub-codes as necessary to categorize the activity. Overall, keyword searches were more common than exact or rephrased questions. 415 of the 500 searches contained a keyword search (83.00%), 90 searches included rephrased questions (18.00%), and 20 searches included exact questions (4.00%).

The *search destination* category specified where participants “landed” after conducting a search. This category included three sub-codes: *search suggestions* (i.e., information appearing beneath the search bar while a search query was being entered), *Google search results page* (i.e., the page of listed results appearing after the search query), and *webpage* (i.e., specific websites). If a participant visited both a Google search results page and a specific webpage, both codes were applied. Each applicable sub-code for search destination was then quantified. For example, if a participant viewed a Google search results page, and navigated to two unique webpages from that search results list, corresponding sub-codes were quantified as one and two, respectively.

The *search suggestions* sub-code was applied if a participant began typing their query in the web browser address bar but stopped based on the emerging suggestions. This code was applied only if participants went no further in their query than to *begin* typing a search and reviewing suggested information from the web browser before entering their response in the questionnaire. As all participants saw search suggestions when conducting a search, we only applied this code if it was clear that a participant stopped searching after those suggestions appeared. This code was applicable to 39 searches (7.80%).

The *Google search results page* sub-code was applied if participants entered a search query, "landed" on a page of Google search results, and then entered their response in the questionnaire. We specify Google in this code because there were no instances of participants navigating to other search engines (e.g., Bing, Yahoo). This sub-code was the most commonly applied of the three possible sub-codes, fitting 457 of the observed searches (91.40%).

The *webpage* sub-code was applied if participants chose to visit a particular webpage in their search (e.g., Wikipedia). Any application of the webpage code was followed with a brief description of the name and content of the reviewed webpage (e.g., "a Smithsonian Air and Space museum webpage discussing the lunar landing"). This sub-code was applicable to 44 of the observed searches (8.80%).

Finally, we included a *search role* category to capture how search appeared to factor into a participant's response. This code was developed following review of a preliminary subset of recordings. The search role code contained four sub-codes: *augmentation*, *incompletion*, *verification*, and *emendation*. These sub-codes were mutually exclusive and fit the entirety of the data set ($N = 500$ searches). A fifth code of *retroaction* was also included to quantify searches that occurred any time after a questionnaire response was entered, and therefore had not been included in participants' self-reports.

The *augmentation* sub-code was applicable to any instance in which a participant (1) searched before entering any response to a questionnaire item and 2) did not search with terms that included any potential answer to the question at hand. As such, this code was applied to any case in which a participant appeared to rely solely on search to address a question, giving no indication of having a potential answer in mind. The name of this sub-code derives from the fact that in such searches, participants did not *indicate* having sufficient background knowledge

available for their response, as could be done by either by pre-writing a response before searching or by using a potential answer as a search term. Rather, in the cases to which the augmentation sub-code was applied, participants appeared to be expanding their existing knowledge base by searching for and ultimately using novel information found online. The augmentation code fit 423 of the 500 searches observed (84.60%).

The *incompletion* sub-code was used in instances in which participants searched without pre-supplying *or* providing a questionnaire response. In these cases, participants engaged in a search but responded with a non-answer (e.g., responding "no answer" or "I don't know"). This code fit 10 of the 500 searches observed (2.00%).

The *verification* and *emendation* sub-codes applied to instances in which participants appeared to formulate potential answers *prior* to their searches. The *verification* sub-code was applied to instances in which participants indicated having relevant prior knowledge before searching and ultimately used that indicated prior knowledge in their final response. This could be evidenced in two ways. First, participants could enter a response to a questionnaire item, conduct a search, and then submit their initial response without any changes beyond minor edits. Minor edits were defined as any superficial change to a response that did not alter its original meaning or reading (e.g., editing the capitalization or spelling of an initial answer). Second, participants could use their search term as their ultimate answer (e.g., using "Dover" as their search term for the questionnaire item "What is the capital of Delaware?" and ultimately using "Dover" as their final response). Both of these behaviors suggested that participants possessed and were able to apply relevant background knowledge *before* conducting their search. That the participant's original response remained unchanged (again, save for minor editing) suggests that

the search they conducted served the purpose of verification, perhaps to address uncertainty or to double-check their response. The verification code fit 32 of the 500 searches observed (6.40%).

Emendation codes were applied to instances in which participants formulated responses before conducting their search, but ultimately changed that initial response after completing the search. Any major change in an initial response following search completion was classified as an instance of emendation, suggesting the participant preferred a new answer identified via their search to their original response (e.g., initially answering "Benjamin Franklin" and then changing their response to "Samuel Morse" after searching for the questionnaire item "Who invented the telegram?"). The emendation code fit 35 of the 500 searches observed (7.00%).

Finally, participants occasionally looked up answers to items sometime *after* submitting a response. As the experiment did not allow for backtracking through the questionnaire items, these searches occurred after-the-fact, and neither impacted responses to the questionnaire items nor were reported in the questionnaire itself. We classified these as instances of *retroaction.* Eight total instances were identified across all participants, bringing the total number of searches observed to 508 (1.60%). Again, while these searches were rare and had no direct impact on questionnaire performance or corresponding self-reports, we sought to classify all searches evidenced in the screen recordings.

Experiment 2 Discussion

Experiment 2 replicated and extended the findings of Experiment 1. Participants completed the same tasks with the same between-subjects manipulation of search condition as in Experiment 1. Their responses to the questionnaire were analyzed using the same procedures. In addition, we collected and coded screen recordings to assess what participants in the search allowed condition did if they opted to look up information while completing the questionnaire.

Overall, participants' questionnaire performance closely aligned with the results of Experiment 1. Participants in both experiments routinely conducted online searches, particularly for hard items and after reading neutral statements. These patterns suggest that participants used searches to augment their prior knowledge and the information they acquired from the stories. While the same overall patterns emerged in both experiments, participants in Experiment 2 demonstrated higher search rates than did participants in Experiment 1 (see Tables 3 and 4 for comparisons of mean search rates). This may be attributable to the differences in setting and sample across experiments. Experiment 2 was conducted in-person with an experimenter present, and with collected screen recordings, which may have predisposed participants in Experiment 2 to work especially hard on our task in comparison to participants in Experiment 1. Additionally, the sample in Experiment 2 included young adult college students at a university setting, and these participants may have adopted a more school-oriented approach to the task relative to our general adult population sample from MTurk who participated in Experiment 1. For example, participants in Experiment 2 may have been more focused on obtaining valid information to answer test questions effectively. Critically, however, the same patterns of search rates obtained in both experiments. Moreover, the results of those searches for questionnaire performance were highly similar across the two experiments.

As critical evidence of this claim, we observed the same patterns of responses coded as *inaccurate - lure* across experiments. Participants in Experiment 2 reproduced inaccurate information when answering questions more often after having read that inaccurate information than they spontaneously produced those responses after reading accurate or neutral statements. Similarly, we observed an effect of statement difficulty, with participants more likely to reproduce inaccurate information from the stories to answer difficult as compared to easy

questions. These main effects were qualified by an interaction highlighting the salient effect of inaccurate statements when answering hard questions. Participants were less likely to hold applicable prior knowledge for hard relative to easy items, based on general knowledge norms, offering an explanation for this interaction. We also observed interactions between statement type and search condition as well as statement difficulty and search condition, highlighting the benefits of search with respect to the potential effects of inaccurate information. Participants assigned to the search allowed condition were significantly less likely to reproduce inaccurate information for hard items as compared to participants in the search not allowed condition, supporting our hypotheses and replicating the findings of Experiment 1. These interactions again drove a main effect of condition, demonstrating overall performance improvements when participants were given the option to search.

Analyses of responses fitting the *accurate* response coding category also yielded results similar to those in Experiment 1. We obtained a main effect of statement type, with participants more likely to produce accurate responses after reading accurate statements than after reading inaccurate or neutral statements. This highlights the obvious benefit of relying on text content when it is valid and relevant. We also observed a main effect of statement difficulty with participants generating fewer accurate responses for hard than easy items. This was qualified by an interaction with search condition, such that participants who were allowed to search were more likely to answer hard items accurately than were participants who were not allowed to search. Effects of difficulty emerged across the coding categories, foregrounding an influence of prior knowledge on participants' questionnaire performance. That statement difficulty interacted with search condition provides evidence that online searches can attenuate this influence. In line with this idea, we also observed a main effect of search condition, with participants in the search

allowed condition producing significantly more accurate responses than did participants in the search not allowed condition. This provides further support for our hypothesis that online searches, when enacted, can produce substantive performance benefits.

Turning next to the category of *inaccurate - other*, we observed a main effect of difficulty as in Experiment 1. Namely, participants across search conditions generated more responses fitting the inaccurate - other code for hard as compared to easy items. This appears to be another manifestation of the influence of prior knowledge on questionnaire performance. Another point of commonality across Experiments 1 and 2 is that in neither experiment did we observe an effect of search condition for this coding category. Rather, participants in the search allowed and no search allowed conditions performed similarly with respect to rates of inaccurate - other responses.

In the discussion section for Experiment 1, we described anecdotal instances of participants providing answers that, while technically incorrect, appeared to reflect consultation of outside resources. We observed similar instances in Experiment 2. Consider the questionnaire item “Which city is Michelangelo's famous "David" located in?” Several participants in the search allowed condition answered this item with “Galleria dell ‘Accademia.” While this is indeed the museum “David” is currently housed in, the question asks for the city rather than the name of the museum. We never observed any instances of this highly specific response in the no search allowed condition, suggesting it may be an instance of participants’ online search gone somewhat awry. With enough such instances, whether attributable to misreading the questionnaire items, using search terms with low correspondence to items, or difficulty evaluating search results, among other possibilities, responses fitting the inaccurate - other coding category could emerge at similar rates across conditions. The statement difficulty by

search condition interaction observed in Experiment 2 reinforces this explanation, as we found that participants in the search allowed condition actually generated more inaccurate - other responses for hard items, but fewer such responses for easy items than did their counterparts in the no search allowed condition. Searching for unfamiliar information, while often beneficial, could expose readers to a variety of potential answers, not all of which would be suitable for successfully answering the questionnaire items.

In addition to the statement difficulty by search condition interaction, we identified one additional difference between Experiments 1 and 2 for the inaccurate - other coding category. In Experiment 1, we observed a main effect of statement type, with participants producing more inaccurate - other responses after reading neutral as compared to accurate or inaccurate statements. This did not replicate in Experiment 2, perhaps due to the observed differences in search rates. Participants in Experiment 2 more frequently engaged in searches more than did participants in Experiment 1, potentially reducing the influence of statement type on participants' responses.

Finally, we examined the coded category of *no answer*. As in Experiment 1, we obtained main effects of statement difficulty and search condition. Participants produced more non-answer responses for hard relative to easy items, speaking to the pervasive influence of prior knowledge on questionnaire performance. The effect of condition, with participants in the search allowed condition producing fewer non-answer responses than did participants in the no search allowed condition, highlights searches as supporting informed responses. We also observed a statement difficulty by search condition interaction, with participants less likely to provide non-answer responses for hard items when allowed to search than when not allowed to search. This interaction demonstrates that search can be especially beneficial for unknown information. With

regard to differences across the experimental results, a main effect of statement type obtained only in Experiment 1, with participants producing more non-answer responses after reading neutral and inaccurate statements, relative to accurate statements. This again suggests that the influence of statement type on response patterns was attenuated in Experiment 2, likely attributable to the increased likelihood of searches observed in the second experiment.

We also observed somewhat different interactions with statement type across the experiments. While increased rates of no answer responses for hard items following neutral statements were observed in both experiments, no interaction between statement type, statement difficulty, and search condition was observed in Experiment 2, contrasting with Experiment 1. This may be attributable to more consistent differences in rates of no answers responses across conditions in Experiment 2, with very low rates of no answer responses in the search allowed condition across the levels of the other independent variables. The lack of statistically significant differences in no answer responses for searched and unsearched items in Experiment 2 may likewise be attributable to the low frequency of no answer responses for participants.

Overall, we identified benefits of searches in three of the four coding categories in Experiment 2, as in Experiment 1. Participants in both experiments were less likely to produce inaccurate responses, more likely to provide answers in general (rather than choosing "no answer"), and more likely to provide accurate responses in particular when allowed to search as compared to when searches were not allowed.

Beyond attempting to replicate the findings from Experiment 1, we also extended our examination to evaluate whether and how participants engaged in searches. To do this, in Experiment 2 we collected screen recordings. Analyses of these recordings focused on features of participants' search behaviors across four coding categories distinct from the categories

applied to questionnaire responses. First, we examined whether participants' self-reports of searches were reliable indicators that searches had actually been enacted. Comparisons of self-report data to actual behaviors evidenced in the screen recording data indicated that their reports were indeed reliable. Out of 500 observed searches, only seven instances indicated disconnects between reports of searches and actual search behavior. Six of those seven reports involved participants supplying a response *before* searching and ultimately submitting their initial response. Search thus did not influence participants' responses in these six cases, as participants generated and ultimately submitted answers that they determined independently. Participants may have viewed these instances of double-checking or confirming known responses as distinct from cases in which they looked up answers to questions that they could not address independently. That said, participants' self-reports matched their recorded behaviors in the overwhelming majority of cases.

The reliability of participants' reports may of course be a function of the screen recording activity. Knowing that a researcher could later view their behaviors may have promoted a heightened degree of honesty amongst participants. That said, there was little if any motivation inherent to the experiment that could have incited inaccurate search reporting. Searches were neither penalized nor rewarded, and performance on the questionnaire itself had no ramifications for participants.

The next coding category applied to search results was *search type*, used to designate how participants formatted their searches (i.e., as the *exact* question posed, a *rephrased* version of the question, or as one or more *keywords*). Participants used keyword searches for more than 80% of their searched items, which seems reasonable given that keywords constitute an efficient and well-practiced means of retrieving information from online resources (Walhout, Oomen,

Jarodzka, & Brand-Gruwel, 2017). What is noteworthy about this coding category is the margin for error associated with straying from the exact question posed in the questionnaire (which participants relied on for only 4% of the searches conducted). The success of any given keyword search hinges on how well it corresponds to the question posed. A low degree of correspondence could be detrimental to search outcomes. Consider the questionnaire item "What is the name of the man who invented the steamboat Clermont?" Multiple participants used "Clermont" as their only search term for this item, yielding a wide range of information about Clermont, Florida, none of which was relevant to answering the item. Again, low correspondence between search terms and questionnaire items could contribute to why participants in the search allowed and no search allowed conditions produced *inaccurate - other* responses on the questionnaire at commensurate rates.

We next considered where participants "landed" after conducting a search. The corresponding se*arch destination* code contained three sub-codes: search suggestions, Google search results page, and webpage. Google search pages were by far the most popular search destination, figuring in more than 90% of the searches conducted. This might suggest that participants were engaging in rather superficial searches, satisfied with whatever information appeared most readily. However, Google search pages offer a notable advantage over other potential search destinations, as they provide information from multiple sources deemed especially relevant to the search in question. In many cases, the Google search page resulting from a participant's query included the same answer to the query across the sources listed. Of course, just because *we* noticed that multiple sources listed on a given search page often suggested the same answer does not mean participants chose to focus on Google search pages for this reason, or even attended to this feature in their searches. While some anecdotal observations

(in the form of cursor movements from source to source across a page of search results) could suggest some participants may have considered consistency across sources, alternative methodologies are necessary to explore this issue. We revisit this empirically in Experiment 4.

The final coding category was *search role*, used to classify the possible intention of a search. The search role code contained four sub-codes: *augmentation*, *verification*, *emendation,* and *incompletion*. Augmentation, which was applicable to any instance of a search preceding entry of a response to a questionnaire item, was the most prevalent sub-code, fitting more than 80% of the observed searches. While the augmentation sub-code was the most commonly applied, this need not mean participants lacked relevant prior knowledge to those respective questions. Rather, application of this sub-code was based on whether or not prior knowledge of an answer was *directly evidenced* prior to search. There could certainly be many cases in which participants had a potential answer in mind but simply elected not to produce that response before searching.

Regarding our other search role sub-codes, we obtained similar rates of verification (6.4%) and emendation (7%) across searches. These sub-codes pertained to cases in which participants provided a response before searching, either keeping it the same after the search (verification) or formulating a new response (emendation). These codes were not especially prevalent in the data set, but they shed light on one of the ways in which the online search option factored into performance improvements. Namely, the option to search for responses allowed participants to double-check and/or revise responses they derived independently.

Of course, search was not always helpful. In a small number (2%) of cases, participants did not provide an answer even after completing a search. Given that this code was uncommon, fitting only 10 trials, detecting any reliable pattern in what went awry in these instances was difficult. A variety of factors, including difficulty understanding the question, difficulty

conducting a corresponding search, and a lack of motivation to identify an answer, among other possibilities, may have driven these cases. Another interesting albeit small set of cases involved *retroactive* searches. Retroaction was a unique sub-code added to encapsulate the eight cases in which participants went back to search for answers to questions after having already submitted responses. The experimental design did not allow participants to backtrack to previous questions, so these searches did not offer any direct performance benefit here.

The overall results again highlighted that while participants reproduce inaccurate information after reading it, searches can ameliorate those effects. Are there methods of further enhancing the benefits of searches? We examine this issue in the next chapter.

Chapter IV: Experiment 3

The results of Experiments 1 and 2 indicated that opportunities to search decreased the likelihood participants would reproduce inaccurate information. These effects were not completely eliminated, as participants still reproduced inaccuracies at statistically significant rates. Moreover, searches were conducted at a relatively low rate. For any topics about which participants are uncertain or for which they lack background knowledge, more frequent searching could have benefits.

While a wide variety of interventions could potentially influence search behaviors, we focused on one that could readily be applied in everyday reading situations. The manipulation involved the use of hypertext which refers to links to relevant information and resources. Hypertext is used frequently online, for a variety of reasons (e.g., to support comprehension; to provide ads; to entertain). Experiment 3 assessed search rates and questionnaire performance for participants assigned to an "augmented" online search condition that included links to relevant information for every questionnaire item. Their performance was compared to that of a new sample of participants assigned to the search allowed condition from the two previous experiments that did not include links.

We predicted that participants assigned to the augmented search condition would be more likely to search for additional information than would participants assigned to the search allowed condition. This was based on the idea that providing links would make searches easier and therefore more likely to occur. We further hypothesized that if participants in the augmented search condition conducted more searches than did participants in the search allowed condition, they would reproduce inaccuracies from the stories at comparatively lower rates, and produce

accurate responses at comparatively higher rates. This prediction was based on the findings from Experiments 1 and 2 demonstrating that engagement in search resulted in performance benefits.

Experiment 3 Methods

Participants

96 participants (25 male, 62 female, and 9 participants who either did not identify with the gender binary or chose not to report gender identity) were recruited from Northwestern University's Introduction to Psychology subject pool to participate in a single session lasting approximately one hour. Course credit was awarded as compensation for participation. Participants ranged in age from 18-23 (M = 18.70).

Materials

The stories and questionnaire materials from Experiments 1 and 2 were used in Experiment 3. Participants assigned to the "standard" search condition received the same instructions as did participants in the search allowed conditions in Experiments 1 and 2. Participants assigned to the "augmented" search condition were allowed to search, but with three additional components. First, when participants were told they could search they were further instructed: "You may either open a separate tab in this browser window to do so or select the link provided for each question. These links will take you directly to search results pertinent to each question to help you determine the correct answer." Second, a line of text appeared beneath each questionnaire item stating: "View search results for this question here" with the word "here" offering a link to Google search results for that particular item. Those search results reflected the exact phrasing of the question posed (e.g., the link for the item "What is the capital of Delaware?" produced search results for "What is the capital of Delaware?" rather than for related keywords

or any rephrased version of the question). Third, participants were asked to indicate if they used the links or did an independent search for each item.

Procedure

The procedure was identical to Experiment 2.

Design

The design of Experiment 3 was the same as Experiments 1 and 2 with one difference. Half of the participants (n = 48) were randomly assigned to the standard search condition, and half (n = 48) were randomly assigned to the augmented search condition. Experiment 3 followed a 3 (statement type: accurate, inaccurate, or neutral) x 2 (statement difficulty: easy or hard) x 2 (search type: standard online search or augmented online search) design.

Experiment 3 Results

Online search rates

We first assessed how frequently participants in the two search conditions self-reported looking up answers to questionnaire items using a repeated measures ANOVA. On average, participants across conditions reported searching for 46.30% (SD = 18.65) of the 74 total questionnaire items (both filler and critical) and 43.00% (SD = 19.36) of the 32 critical items. Mean individual search rates ranged from 0.00% to 93.75% of the critical items. In the augmented search condition, participants reported searching for 49.87% (SD = 15.84) of the critical items on average, with individual search rates ranging from 9.37% to 93.75%. 86.59% of the searches conducted by participants in the augmented search condition involved only use of the provided links. In the standard search condition, the average self-reported search rate for critical items was 36.13% (SD = 20.26), with a range of 0.00% to 71.87%.

We submitted search rates for the critical items to analysis using a repeated measures ANOVA, with statement type (i.e., accurate, inaccurate, or neutral) and statement difficulty (i.e., easy, hard) entered as within-subjects variables, and search condition (i.e., augmented search, standard search) entered as a between-subjects variable. We observed a main effect of statement type. Participants were less likely to search for information from previously presented accurate statements (M = 39.39%, SD = 33.49) than they were to search for previously presented inaccurate (M = 45.44%, SD = 32.43) or neutral statements (M = 45.66%, SD = 33.69). A main effect of statement difficulty also emerged, with participants more likely to search for hard items (M = 63.02%, SD = 25.56) than for easy items (M = 22.98%, SD = 16.95), [F (1, 95) = 375.35, MS = 20.14, $p < .001$, $\eta_p^2 = .80$]. These effects replicated findings from Experiments 1 and 2. Importantly for our hypothesis, we observed a main effect of search type with participants in the augmented search condition (M = 49.87%, SD = 15.84) significantly more likely to search than were participants in the standard search condition (M = 36.13%, SD = 20.26), [F (1, 95) = 14.12, MS = 2.99, $p < .001$, $\eta_p^2 = .13$].

These main effects were qualified by two interactions. First, we observed an interaction between statement difficulty and search type [F (1, 95) = 6.40, MS = 0.34, p = .013, $\eta_p^2 = .06$]. Participants in the augmented search condition (M = 72.66%, SD = 18.66) were more likely to report searching for hard items than were participants in the standard search condition (M = 53.38%, SD = 27.98), ($p < .001$). Participants in the augmented search condition (M = 27.08%, SD = 16.97) were also more likely to search for easy items than were participants in the standard search condition (M = 18.88%, SD = 16.07), (p = .01). Additionally, we observed an interaction between statement difficulty and statement type [F (1, 95) = 5.11, MS = 0.16, p = .007, $\eta_p^2 = .05$]. For hard items, participants searched significantly more often for information

previously presented in neutral statements (M = 66.21%, SD = 28.92) as compared to in accurate statements (M = 59.54%, SD = 30.88), (p = .045). Search rates following inaccurate presentations (M = 60.82%, SD = 30.22) did not differ from either of the other statement types (p's > .16). For easy items, participants searched significantly more often for information previously presented in inaccurate statements (M = 30.06%, SD = 26.88) as compared to accurate statements (M = 19.24%, SD = 21.92), (p = .002). Search rates following neutral presentations (M = 25.10%, SD = 24.33) did not differ from either of the other statement types (p's > .14). This interaction helps distinguish ways in which information from the stories influenced search behaviors. No other significant effects or interactions were observed (all F's < 0.26).

Questionnaire coding

Responses to the critical questions (N = 3,072) were categorized using the same four-category coding scheme from Experiments 1 and 2. The *inaccurate - lure* code fit 4.39% of responses in the data set. The *accurate* code fit 82.73% of responses in the data set. The *inaccurate -other* code fit 9.66% of responses in the data set. Finally, the *no answer* code fit 3.20% of responses in the data set. All response categories were mutually exclusive and fit the entirety of the data set. Two raters independently categorized 50% of the questionnaire responses with the remaining half coded by one rater only. Interrater reliability was reliably high (κ = .94), with all disagreements resolved through discussion. Table 6 presents response rates across all coded categories as a function of search.

Questionnaire response rates

We assessed response rates (i.e., the frequency of responses fitting a given code) for each of the four coded categories with separate repeated measures ANOVAs as in Experiments 1 and 2 with Bonferroni corrections. Statement type (i.e., accurate, inaccurate, or neutral) and

statement difficulty (i.e., easy, hard) were entered as within-subjects variables. Search type (i.e., augmented search, standard search) was entered as a between-subjects variable. Response rate was the dependent variable.

Inaccurate - lure responses

Participants were more likely to reproduce inaccurate information after reading inaccurate statements (M = 10.83%, SD = 15.67) than they were to spontaneously produce those inaccuracies after reading accurate (M = 1.30%, SD = 4.95) or neutral statements (M = 1.03%, SD = 4.18), [F (1,95) = 31.09, MS = 1.46, $p < .001$, $\eta_p^2 = .25$]. Participants were also significantly more likely to produce inaccurate responses to hard (M = 7.06%, SD = 13.35) as compared to easy items (M = 1.71%, SD = 6.35), [F (1,95) = 79.04, MS = 3.28, $p < .001$, $\eta_p^2 = .46$]. This again replicates Experiments 1 and 2.

We also observed a statement type by statement difficulty interaction [F (1,95) = 20.16, MS = 0.77, $p < .001$, $\eta_p^2 = .40$]. For hard items, participants were more likely to reproduce inaccurate information from the stories after reading inaccurate statements (M = 16.73%, SD = 17.93) than they were to spontaneously produce inaccuracies after reading accurate (M = 2.60%, SD = 6.76) or neutral statements (M = 1.85%, SD = 5.44), with rates of spontaneous productions no different following accurate or neutral statements ($p > .99$). Similarly, albeit at an overall lower rate, participants were more likely to reproduce inaccurate information for easy items following inaccurate statements (M = 4.93%, SD = 1.01) than they were to spontaneously produce those inaccuracies following accurate (M = 0.00%, SD = 0.00) or neutral statements (M = 0.21%, SD = 2.04). There were no differences in spontaneous productions following accurate and neutral statements ($p > .99$). There was also no difference in inaccurate reproductions across

our augmented (M = 4.16%, SD = 10.99) and standard search conditions (M = 4.62%, SD = 10.60), (p = .51). No other significant effects or interactions were observed (all F's < 1.28).

Accurate responses

We observed a main effect of statement type, with participants generating more accurate responses after reading accurate statements (M = 87.99%, SD = 18.32) than after reading inaccurate (M = 77.41%, SD = 23.77) or neutral statements (M = 82.78%, SD = 21.89), [F (1,95) = 29.92, MS = 3.20, p < .001, η_p^2 = .24]. We also observed a main effect of statement difficulty. Participants were more likely to produce accurate responses for easy (M = 91.61%, SD = 13.88) than for hard items (M = 73.85%, SD = 24.59),[F (1,95) =173.94, MS = 29.01, p < .001, η_p^2 = .65]. These results replicated previous findings.

The main effects of statement type and statement difficulty were qualified by a significant interaction [F (1,95) = 7.60, MS = 0.69, p < .001, η_p^2 = .07]. For hard items, participants were more likely to produce accurate responses after reading accurate statements (M = 80.84%, SD = 21.94) than after reading neutral statements (M = 74.66%, SD = 25.50), (p < .001). Rates of accurate responses following both accurate and neutral statements were also higher than following inaccurate statements (M = 66.05%, SD = 24.15), (p's < .015). For easy items, participants were more likely to produce accurate responses following accurate statements (M = 95.14%, SD = 9.48) than following inaccurate statements (M = 88.77%, SD = 17.09), (p = .005). In contrast to hard items, no differences were observed for easy items in rates of accurate responses following presentations of accurate and neutral statements (p = .07) or neutral and inaccurate statements (p = .93).

Critically for our hypothesis, we obtained a main effect of search condition. Participants in the augmented search condition were more likely to generate accurate responses (M = 87.96%,

SD = 15.37) than were participants in the standard search condition (*M* = 77.50%, *SD* = 25.78), [*F* (1,95) = 17.33, *MS* = 6.22, $p < .001$, $\eta_p^2 = .16$]. This was qualified by an interaction between statement difficulty and search condition [*F* (1,95) = 14.63, *MS* = 2.44, $p < .001$, $\eta_p^2 = .13$]. Participants in the augmented search condition were more likely to produce accurate responses for hard items (*M* = 82.45%, *SD* = 17.04) than were participants in the standard search condition (*M* = 65.25%, *SD* = 27.82), ($p < .001$). At an overall higher rate, participants assigned to the augmented search condition (*M* = 93.47%, *SD* = 11.08) also produced more accurate responses for easy items than did participants in the standard search condition (*M* = 89.74%, *SD* = 16.04), ($p = .041$). Finally, we observed a three-way interaction between statement type, statement difficulty, and search condition. Participants in the augmented search condition were more likely to produce accurate responses after reading *inaccurate* statements for easy items (*M* = 91.95%, *SD* = 12.17) than were participants in the standard search condition (*M* = 85.59%, *SD* = 20.53), [*F* (1,95) = 6.12, *MS* = 0.56, $p = .003$, $\eta_p^2 = .06$]. The augmented search condition here appears to have afforded some degree of "protection" against the influence of exposure to inaccuracies. No other significant effects were obtained (all *F*'s < 1.39).

Inaccurate - other responses

We observed a main effect of statement type with participants more likely to provide inaccurate - other responses after reading neutral statements (*M* = 12.42%, *SD* = 14.86) than after reading accurate (*M* = 7.88%, *SD* = 11.41) or inaccurate statements (*M* = 8.66%, *SD* = 12.96), [*F* (1,95) = 8.38, *MS* = 0.60, $p < .001$, $\eta_p^2 = .08$]. We also observed a main effect of statement difficulty. Participants generated more responses fitting the inaccurate - other category for hard (*M* = 13.81%, *SD* = 14.38) as compared to easy items (*M* = 5.50%, *SD* = 10.57), [*F* (1,95) = 140.97, *MS* = 7.61, $p < .001$, $\eta_p^2 = .60$]. We also observed a main effect of search condition.

Participants assigned to the augmented search condition (M = 7.74%, SD = 10.56) produced significantly fewer responses fitting the inaccurate - other category than did their counterparts in the standard search condition (M = 11.67%, SD = 15.31), [F (1,95) = 8.93, MS = 0.97, p = .004, η_p^2 = .09].

These main effects were qualified by an interaction between statement difficulty and search condition [F (1,95) =17.17, MS = 0.93, p <.001, η_p^2 = .15]. Participants assigned to the standard search condition were more likely to provide inaccurate - other responses for hard items (M = 17.52%, SD = 16.74) than were participants assigned to the augmented search condition (M = 10.10%, SD = 10.36). In contrast, participants generated inaccurate - other responses for easy items at similarly low rates across conditions (p = .95). No other effects were significant (all F's < 3.22).

No answer responses

We observed a main effect of statement difficulty, with participants more likely to produce no answer responses for hard (M = 1.13%, SD = 4.85) than easy items (M = 5.28%, SD = 16.14), [F (1,95) = 15.86, MS = 1.84, p < .001, η_p^2 = .14]. We also observed a main effect of condition. Participants in the augmented search condition (M = 0.09%, SD = 1.04) were significantly less likely to provide no answer responses than were participants in the standard search condition (M = 6.32%, SD = 16.50), [F (1,95) = 2.88, MS = 15.60, p < .001, η_p^2 = .14]. These main effects were qualified by a significant interaction [F (1,95) = 9.93, MS = 1.15, p = .002, η_p^2 = .10]. Participants in the augmented search condition (M = 0.17%, SD = 1.47) were less likely to generate no answer responses for hard items than were participants in the standard search condition (M = 10.38, SD = 21.64), (p < .001). At an overall lower rate, participants in the augmented search condition (M = 0.00%, SD = 0.00) were also less likely to generate no answer

responses for easy items than were participants in the standard search condition (M = 2.26%, SD = 6.68), ($p < .001$). No other significant effects or interactions were obtained (all F's < 1.16).

Experiment 3 Discussion

The purpose of Experiment 3 was to assess an ecologically valid means of increasing participants' search rates and, in turn, reducing rates of inaccurate reproductions and increasing rates of accurate responses. Hyperlinks to additional resources are commonly embedded in texts presented online, facilitating access to information that readers might not know of or seek out on their own. We predicted that offering hyperlinks to relevant information during the questionnaire task would foster searching. We further predicted that rates of reproductions of inaccurate information would decrease and that rates of accurate responses would increase when these hyperlinks were offered as compared to when participants were merely given the option to search independently.

Analyses of search rates supported our hypothesis. Participants in the augmented search condition, presented with the option to use hyperlinks for every questionnaire item, conducted more searches overall, and for hard items in particular, as compared to participants in the standard search condition. This suggests that decisions to search were positively influenced by the inclusion of hyperlinks in the materials. This may be attributable to the increased ease of completing searches, as the links took participants directly to relevant information without requiring additional typing, and additional consideration of what to type.

Having identified a beneficial influence of the hyperlink manipulation on search rates, we examined whether this benefit influenced performance on the questionnaire task. Overall, the standard patterns of influence of inaccurate information were again obtained in Experiment 3 at a low rate. Participants produced significantly more inaccurate responses after reading inaccurate

statements relative to accurate or neutral statements. They also produced more inaccurate responses for hard as compared to easy items. These effects were qualified by an interaction such that participants were more likely to reproduce inaccurate information from the texts for hard items specifically. However, contrary to our hypothesis, participants in the standard and augmented search conditions reproduced inaccuracies from the texts at equivalent rates.

To what might we attribute the lack of difference between the search conditions? One possibility relates to the fact that neither the standard search condition nor the augmented search condition directed participants' attention towards searching for items related to the text contents. That is, participants were not given any indication that information from the texts might be inaccurate or that search might be especially useful for addressing text-relevant items. Manipulations that increase search rates generally, rather than specifically for text-relevant items, may be insufficient for achieving greater reductions in inaccuracy use than those observed in our standard search conditions in Experiments 1 and 2. To achieve larger reductions, instructions directing participants' attention specifically to text-relevant items may be necessary. Future work should consider how people might be encouraged to more closely direct their search behaviors to detect and reject inaccuracies presented in what they read.

A related possibility pertains to the timing of our hyperlink manipulation. Recall that we provided participants with hyperlinks on the questionnaire, after they had finished reading. This timing was suitable for the current project as the hyperlinks were intended to change behavior at test rather than during reading. However, providing participants with hyperlinks during reading could actually afford heightened opportunity for them to verify and "tag" pieces of information from the text as potentially inaccurate. Future work should thus also consider how the timing of manipulations targeting inaccuracy use might be optimized.

Our next set of analyses examined rates of accurate responses. Participants were more likely to produce accurate responses after reading accurate statements, as compared to inaccurate or neutral statements, again indicating a beneficial influence of previously read text content when it was valid. Participants were also more likely to produce accurate responses for easy than for hard items. Additionally, participants were more likely to produce accurate responses after reading accurate statements specifically for hard items. This interaction evidences how participants used text content to supplement gaps in knowledge. In addition to differences observed across levels of statement type and statement difficulty, we also observed differences across the two search conditions. Participants in the augmented search condition produced significantly more accurate responses than did participants in the standard search condition. This exemplifies a benefit of the hyperlink presentations, providing support for our hypothesis.

We next considered response rates for our *inaccurate - other* coding category. Participants were more likely to produce inaccurate - other responses after reading neutral statements as compared to after reading accurate or inaccurate statements. This can be attributed to the lack of relevant and useful information in the neutral statements. Participants were also more likely to produce inaccurate - other responses for hard as compared to easy items. Participants in the augmented search condition also produced significantly fewer inaccurate - other responses than did participants in the standard search condition, exemplifying an additional benefit of the hyperlink manipulation. We also observed a significant interaction between condition and difficulty, such that participants were more likely to provide inaccurate - other responses for hard items specifically when assigned to the standard search condition. This outcome, in combination with differences observed in rates of accurate responses, suggests that participants in the augmented search condition tended to provide responses that were more

closely targeted to the questionnaire items. The analyses of inaccurate - other responses thus provided additional evidence that hyperlink presentations were beneficial for participants, albeit not directly localized to previously read inaccuracies.

Our final set of analyses examined rates of no answer responses. Participants were more likely to give no answer responses for hard as compared to easy items. Participants in the standard search condition were also more likely to produce no answer responses overall than were participants in the augmented search condition. These effects were qualified by an interaction between statement difficulty and search condition, with participants more likely to produce no answer responses for hard items when assigned to the standard search condition as compared to the augmented condition. No answer responses were virtually eliminated in the augmented search condition, with only two instances observed in the 1,536 total responses. Participants in the standard search condition, in contrast, produced 96 responses fitting this category (again out of 1,536 total). This highlights a notable feature of task performance for participants in the augmented search condition, in that they produced specific answers for virtually every questionnaire item. This orientation towards task completion may be attributable to the ease with which participants could retrieve information for each item via the hyperlinks.

Of course, the answers participants produced still contained inaccuracies from the stories, with inaccurate productions commensurate across the search conditions. While use of text content can certainly yield benefits when those contents are valid, the aim of our intervention in the current experiment was to elicit use of more consistently reliable information. This objective was achieved to some degree, with participants in the augmented search condition displaying higher search rates and accurate response rates, lower rates of wrong guesses, and fewer non-answer responses than did participants in the standard search condition. However, this

experiment again demonstrated the overall persistence of inaccurate information use. Many studies to date have consistently obtained this result, raising questions about why these effects are so difficult to eliminate. In the fourth and final experiment, we sought to better understand the thought processes underlying people's use of inaccuracies as a means of understanding the durability of this phenomenon.

Chapter V: Experiment 4

Experiments 1, 2, and 3 demonstrated that searches improved participants' performance on the general knowledge task. In Experiment 4, we sought to examine the nature of the obtained benefits and the processes underlying those benefits by collecting concurrent verbalizations and responses to a semi-structured interview protocol regarding the questionnaire task. Unlike the previous experiments, Experiment 4 followed a case study design rather than a randomized experimental design. A case study format afforded the opportunity to examine participants' thoughts using a grounded approach, without generating or testing hypotheses regarding the types of verbalizations or interview responses participants might produce. This in-depth examination involved studying a small set of participants' thoughts and responses to the tasks incorporated in prior experiments. The goal was to derive a detailed account of how participants went about answering test questions with particular attention to (1) any potential use or mention of text contents and (2) engagement in and decisions regarding search. This account could inform existing theoretical explanations for people's use of previously read inaccuracies, and also potentially provide insight into methods of further reducing that use.

Experiment 4 Methods

Participants

10 participants (5 male and 5 female) were recruited through Northwestern University's Introduction to Psychology subject pool to participate in a single session lasting approximately one hour. Participants ranged in age from 18-21 (M = 19.40). They received course credit as compensation.

Materials

We utilized a subset of the materials from the previous three experiments. Two of the four stories and half (37) of the questionnaire items were presented to participants. We utilized half of the materials to reduce fatigue, repetitions, and task demands. Table 2 denotes which items were excluded from Experiment 4. In addition to these materials, a semi-structured, seven question interview protocol was developed. The interview questions probed aspects of the behaviors that participants may have enacted while completing the questionnaire. A semi-structured interview was used as participants could engage in a variety of behaviors (e.g., searching online or not; providing an answer or no response), such that not every question of interest might be applicable to every participant. Each of the questions in the protocol, as well as their requisite contingencies, are delineated in Table 7.

Procedures

Participants began by reading the two stories, one at a time in a self-paced manner. They were then given the same instructions for the questionnaire task provided in the search allowed conditions from Experiments 1 and 2 with one change. Participants were asked to provide verbalizations while completing the questionnaire, with instructions stating: “As we are particularly interested in the process by which people go about answering these questions, we would like you to speak your thoughts about each question out loud as you complete this task. That is, we want to know what you are thinking as you go about answering each question. Please speak your thoughts aloud as you answer every test item.” If a participant did not speak their thoughts aloud for any questionnaire item, they were prompted by the experimenter with “What led you to this response?” before they continued to the next item. The instructions and prompts were intended to elicit descriptions of the different factors and considerations that influenced participants’ choices and responses.

After participants completed the questionnaire task, the experimenter initiated the interview protocol. Interview questions were asked one at a time in the order listed in Table 7 with as many of the seven total questions being asked as was applicable for each participant. Participants' verbalizations and interview responses were audio recorded for later transcription.

Design

Experiment 4 used a 3 (statement type: accurate, inaccurate, or neutral) x 2 (statement difficulty: easy or hard) design, with no manipulation of search condition (i.e., all participants were allowed to search).

Experiment 4 Vignettes

We present data from Experiment 4 here as descriptive vignettes. Participants' responses are summarized in turn, first with a numeric overview of how each participant went about completing the questionnaire task (e.g., how many responses they answered accurately, how many items they conducted searches for). We then detail ways in which the story contents figured in participants' concurrent verbalizations and interview responses to elucidate factors underlying or protecting against people's reproductions of inaccuracies. We also discuss participants' search behaviors.

Participant 1

Participant 1 provided an accurate response to all 16 of our critical items, engaging in search for half of them. Six of these eight searches were for hard items and two were for easy items. Participant 1 completed the first six critical questionnaire items without referencing the text, either citing prior knowledge as the basis of their answer (e.g., "Physicians who specialize in skin ailments are dermatologists. My sister is a dermatologist") or engaging in online search to determine an answer (e.g., "Oh geez I have no idea. Invented the telegraph… [*conducts online*

search] Samuel Morse. Oh is that the guy that made the Morse Code?"). On the seventh critical item, which asked "What is the medical term for the inability to sleep?" the participant stated:

> "That was in one of the stories um I have no idea what it was though. Medical term for inability to sleep… [*conducts online search]*. No that's not the word they used in the story. They used a more complicated sounding word that started with like an n. Maybe I'm imagining things. Huh. Strange."

This highlights one route by which searches may benefit people after exposure to inaccurate information. While their memory of the inaccurate information in the story was imprecise, the participant nonetheless expressed confusion regarding a discrepancy between the information in the story and the information derived from their search. That search ultimately led them to provide an accurate response. Participant 1 went on to mention the story two more times, each time revealing confusion about that same type of discrepancy. For the question "What is the name of the scientist who discovered radium?", Participant 1 noted: "Well in the story it had his name in it." Interestingly, the participant actually read the neutral version of this information. Participant 1 searched for and found a correct answer to this item, subsequently noting that: "[The story] didn't say that right?" Participant 1 also referenced the stories for the question "What is the name of the first doctor to successfully perform a heart transfer?" The participant initially stated: "I don't know. It mentioned it in the story but…I feel like it would have been a long time ago. I feel like that story was wrong," before conducting a search. Participant 1 actually read a neutral statement version of this information as well, but his apparent doubt in the stories' veracity is notable.

That Participant 1 was confused by and doubtful of the stories' contents was reinforced by their responses to the interview questions. When asked why they chose to search for some answers but not others, Participant 1 volunteered the first instance in which they noted a

discrepancy between story information and what they believed was an appropriate answer. They stated:

> "Um if I didn't know the answer I'd look it up. I think there were a couple where um like with the insomnia one where I was pretty sure I knew the answer was insomnia but I thought I remembered seeing in the story that they'd called it something else um and I wasn't reading super closely so I didn't remember what it was called but I thought it started with the letter N. So I was looking for something that followed that and then that's kinda why I looked it up just to make sure I was right."

This response hints at a means by which people may attempt to avoid the undesirable consequences of exposure to inaccurate information. First, Participant 1 noted searching for a match, either between the story and prior knowledge or between the story and information obtained via search, as a strategy for answering questions. This validation process may lead the reader to identifying content from the story as inaccurate. Obtaining matches between false text contents and either prior knowledge or online resources, for example, would be unlikely to occur. Secondly, the participant "wasn't reading super closely." While close readings can benefit comprehension, people also benefit from limiting the attention they allocate towards unreliable information. Participant 1 reinforced this idea when answering whether they thought the stories were applicable to the questionnaire task:

> "...as I probably hinted at as I was describing my reactions to the questions um I think that there were a lot of details in the stories that were not accurate and that could have thrown someone off that did pay attention to the specific details and stuff."

While we were interested in responses regarding the texts, we were also interested in examining search decisions. As evidenced in the quotes above, Participant 1 cited uncertainty or a lack of prior knowledge as reasons for searching. With respect to search procedures, Participant 1 stated that they "just looked at the first answer that would come up," further noting that "if something didn't make sense I'd click like Wikipedia or something to see if it was right." They

also added they didn't "think any of these questions were like super hard like you needed a PhD to answer," suggesting they did not perceive it to be especially hard to determine valid answers for this task.

Participant 2

Participant 2 provided accurate responses to 14 of the critical items, and reproduced inaccurate information from the stories for two items. Participant 2 searched for seven of the 16 critical items, with five searches for hard items and two for easy items. They did not search for either of the items for which they reproduced inaccurate information from the texts. In contrast to the majority of their peers, Participant 2 did not refer to the texts at all while completing the questionnaire. They either simply stated an answer if they had one in mind, sometimes adding details about how they knew it (e.g., "Photosynthesis. I remember that from biology but I don't remember how to spell it"), or they stated that they would be looking up a response.

When asked about their approach to completing the questionnaire, Participant 2 stated: "I mean mostly just relying on knowledge that I probably had or should know and then googling anything I was kind of unsure of or wanted to double check." Participant 2 further indicated that they elected to search when she "just wasn't sure right away." This suggests that they thought the inaccuracies from the story were valid, given that in both applicable cases, they used that information without searching or indicating uncertainty. Participant 2 then qualified this statement by adding that: "A few of them I just didn't feel like googling like the one about specialization in cutting body parts cause that seems really general too."

With respect to searching, Participant 2 noted that the questions were "pretty straightforward" such that when they "copied the question or the keywords... the first thing that showed up was what was probably right." When asked whether they found the stories applicable

to the questionnaire task, Participant 2 noted they "were applicable in terms of the idea. I don't think they were great in terms of the exact answer. They didn't necessarily name all the names." This suggests Participant 2 may have been more aware of a connection between the stories and the questionnaire than their concurrent verbalizations would suggest. However, Participant 2 did not indicate detecting any discrepancies between their prior knowledge and the story information. Further, when presented with one of inaccuracies they used during the post-experiment debriefing, Participant 2 did not indicate awareness of nor comment on the inaccurate content as a key element of the study. That said, it is always possible that some awareness or realization may have occurred but simply was not shared.

Participant 3

Participant 3 provided accurate responses to 12 of the 16 critical items. For the four items they completed incorrectly, Participant 3 wrongly guessed on three items and used an inaccuracy from the stories for the remaining item. They searched for four of the 16 critical items, with three of these searches for hard items and one search for an easy item. Participant 3 referenced the texts for six items in the questionnaire task, starting with the first question. As they completed the question "Who wrote the play *Romeo & Juliet?"* Participant 3 noted: "So the dad in one of the stories said it was Jonson which is really funny because everyone knows it was Shakespeare." Participant 3 thus appeared to have detected a discrepancy between their prior knowledge and the story information, and discounted the falsehood while answering the question. Shortly thereafter, however, Participant 3 used an inaccuracy from the stories to answer the question "What is the name of the man who invented the steamboat 'Clermont'?" While addressing this item, Participant 3 noted: "Whitney. That was in one of the stories, what was his first name? John?" Immediately after answering, however, Participant 3 began to express doubt in their response.

Instead of moving to the next question, they pondered: “Well the story was wrong about Shakespeare so maybe it wasn’t Whitney. It had one factual error…. *At least* one factual error. Is it okay if I look it up?” After being told that they could search, Participant 3 stated: “See, it was Robert Fulton. So the narrator was unreliable, can’t trust Billy’s dad.”

Participant 3 subsequently noted and refuted inaccuracies from the texts for three additional critical items, rejecting the inaccurate information for one on the basis of prior knowledge and using searches to verify that the others were inaccurate. In the former case, when answering the question “What is the name of the project which developed the atomic bomb during World War II?”, Participant 3 noted: “It’s the Manhattan project. The story said it was Los Alamos which doesn’t sound as nice. Manhattan project really rolls off the tongue. And I know it’s the right answer.” In the latter cases, Participant 3 initially mulled over the information in the story for answering the question “What is the name of the man who invented the smallpox vaccine?”, stating: “So the story said Jonas Salk. Jonas? But wasn’t he polio? I’m going to look it up.” Participant 3 then verified Salk’s connection to polio before investigating the smallpox vaccine and ultimately providing the correct answer (Edward Jenner). Finally, for the question “What is the system by which plants make food for themselves?”, Participant 3 noted that the stories had mentioned the “Krep’s cycle.” They then conducted a search for “Krep’s cycle,” which Google corrected to Krebs cycle, and read aloud the dictionary definition from the first page of search results. Participant 3 contemplated the Krebs cycle as something “that all of us do” whereas photosynthesis was specific to plants. They then entered photosynthesis as their final answer.

Participant 3 also mentioned that the stories were useful for answering at least one of the items accurately. For the item about the most famous Greek doctor, Participant 3 stated:

> "One of the stories mentioned the Hippocratic oath. Hippocrates? But was he a philosopher? His name sounds like a philosopher. But I guess if it was mentioned in a physician's oath then he must have been involved in medicine… I'm going to look it up. There you go, Hippocrates."

Participant 3 thus took the additional step of searching to verify their initial thoughts, rather than trusting the story content outright. Nevertheless, that Participant 3 pointed to a portion of the story potentially being applicable to answering a question suggests they did not consider the texts to be wholly unreliable sources.

In the interview portion of the experiment, Participant 3 stated that they primarily relied on prior knowledge and prior experiences to address questionnaire items, and were confident in their final answers. When asked why they chose to search for some questions and not others, Participant 3 stated:

> "I love learning new things so I looked up things that I wanted to know more about. So if I was really curious to know the answer I would look it up. For some questions though I just didn't really care to be honest."

This points to two interesting individual differences potentially implicated in how people completed the questionnaire task. *Trait* curiosity, which involves being inquisitive and interested in learning and information exploration generally, could predispose people to engaging in search. *State* curiosity, which pertains to interest arising for particular topics or in particular contexts, could likewise motivate searches. In discussing how they approached searching, Participant 3 stated:

> "So usually Google provides the answer in bold at the top and then has some more description underneath it, usually part of an article from Wikipedia. For these types of questions this would usually work. If the questions were something else I would have checked a few articles but for this Google was fine."

Participant 3 thus echoed sentiments expressed by Participants 1 and 2, indicating that search strategies are task dependent, and that the appropriate search strategy for *this* task was

relatively straightforward. Finally, with regard to the stories, Participant 3 stated: “I don’t know what the right word is but the stories were kind of… harmful to me doing well on the task,” and then reiterated that Billy’s dad was an unreliable narrator. This suggests Participant 3 held a rather negative view of the story contents, deriving from their discovery that some information from the texts was presented inaccurately.

Participant 4

Participant 4 provided accurate responses to 15 of the 16 critical questions and a wrong guess for the remaining question. They searched for 10 of the 16 critical items, with seven searches for hard items and three for easy items. Participant 4 referenced the texts four times throughout the questionnaire task, starting with the first critical item presented. For this item, Participant 4 stated: “Shakespeare but then in the thing that we read it said someone else so I don’t know what that was but…” Participant 4’s thoughts then trailed off and they moved to the next item without further comment. This statement indicated that Participant 4 detected a discrepancy between the story content and their prior knowledge, ultimately relying on the latter.

Participant 4 next mentioned the texts when answering the question “What is the name of Dorothy's dog from *The Wizard of Oz*?” This information had been presented accurately in the story, with Participant 4 saying “Toto” as they entered their response before adding “That was in the reading thing.” It is difficult to ascertain with certainty whether they were merely remarking on a match between their prior knowledge and the story content or actively using the story content to augment their existing prior knowledge (among other possibilities). Given how frequently participants provided “Toto” as a response to this item on their own, however, the former possibility seems plausible. Participant 4 next mentioned the stories while addressing the question “What is the name of the first doctor to successfully perform a heart transfer?”

Participant 4 stated: “Oh that was definitely in that story but uh. I don’t remember,” before conducting a search and answering the question correctly. Similarly, for the question “What are physicians who specialize in cutting the body called?” Participant 4 remarked: “I think that was in the thing too haha,” without mentioning what the story had stated. They then searched to answer the question, albeit ultimately responding with “incisionist” rather than “surgeon.”

In the interview portion of the experiment, Participant 4 described their general approach to the task as them knowing “some of them” and there being a “a few that [they] didn’t feel confident about” but in fact did know. They also mentioned that for the questions there were “some [they] definitely learned in the past but didn’t remember” and also “some [they] had never heard of and looked up.” Participant 4 stated that they were confident in their answers for all but the item pertaining to physicians who specialize in cutting the body as they didn’t think they were “searching the right thing.” With respect to how they searched in general, Participant 4 stated:

> “Yeah I mean hahah if I was doing this for a class I would probably look at the sources more so I probably should have been doing that. But I guess I’ll like look at the result and then kind of read a little - cause a lot of times Google will recommend a name if you search who is so and so but then I like to read the little summary to make sure that that person is like - like cause sometimes it’ll match you with the wrong thing so I would read a little briefly. I think there were a few where I was like I can’t tell if they are giving me the right results so then I like clicked on a page to read more. Most of the time I trusted the first result which probably isn’t the right thing to do haha. There were also some that I like knew the answer but had to recall it so when I saw the right answer I was like oh okay that’s it.”

This response encapsulates two noteworthy search issues. First, Participant 4 noted that if the task had been for a different purpose, they would “probably” have attended more closely to sources, also adding that they “should” have done that here. This suggests the possibility of a participant approaching the task differently in other contexts. That Participant 4 consulted

specific webpages when uncertain about the initial search results points to greater involvement in the task than she may have given herself credit for. Second, they noted reading at least slightly further than just the first name that appeared in the search results page as a means of validating what the most suitable answer seemed to be, while still trusting the first result "most of the time." This suggests that their searches at times were more involved than they might appear to an observer.

In response to the last interview question regarding perceptions of the stories' relevance to the questionnaire task, Participant 4 first noted that they "definitely recognized some names and like some concepts" referencing within-story discussion of different types of doctors as an example. This suggests that Participant 4 may have found the stories useful to some degree for answering the questions. However, they then added that there were "some that I was like, I don't know, I didn't trust [the stories] as much I guess cause like um one of them mentioned *Romeo and Juliet* being written by someone who wasn't Shakespeare." This was the only inaccuracy they mentioned noticing, suggesting detection of a single falsehood was sufficient for them to become skeptical of the validity of the stories. With reference to this issue, Participant 4 stated that they were "reading more for like the like the plot than the specific trivial knowledge questions." As participants were not informed about connections between the stories and the questionnaire, this is certainly not a surprising response. Participant 4's comment highlights that orientations to the texts can influence the information people attend to and rely upon later.

Participant 5

Participant 5 provided accurate responses to 15 of the critical questions and an inaccurate guess for remaining question. They engaged in online search for nine of the 16 critical questions. Seven of these searches were for hard items and two were for easy items. Participant 5

referenced the texts for seven items in the questionnaire task, starting with the first critical question. Participant 5 first provided the answer, stating: “Shakespeare, read that in high school. Did not look that up.” They then added: “ I also remember in the short story he said the wrong name for who wrote it and that kinda was interesting.” Like previous participants, Participant 5 detected and discounted this inaccurate presentation from the stories.

Participant 5 responded quite similarly to a subsequent question of “What is the medical term for the inability to sleep?” They stated: “I did catch that as well in the short story. She said narcolepsy but that is not right, that’s when you sleep too much. It’s actually insomnia. I did not look that up.” Here we see a clear refutation of a story inaccuracy on the basis of prior knowledge. Participant 5 also cast doubt on the veracity of the story content for a topic they were unfamiliar with. For the question “What is the name of the man who invented the steamboat ‘Clermont’?” they stated: “I remember that from the short story and I think I remember the name that the dad said but he was wrong about *Romeo and Juliet* so I’m gonna look that up.” Like other participants, Participant 5 contemplated the dad in one of the stories as being an unreliable narrator. After searching, Participant 5 added: “I don’t think that was the name that the dad said or was it?” This suggests that Participant 5 held some degree of doubt in the veracity of what the story said about this topic, albeit tempered with uncertainty. Participant 5 also stated knowing that the topics for three other items came up in the stories, but that they were not entirely certain as to whether a name or concept was specified. Each of these items was in fact presented neutrally, such that Participant 5’s statements accurately reflected the contents of the texts. Finally, they noted one instance in which story content was useful. For the question “Who was the most famous Greek doctor?” Participant 5 stated: “ I’m pretty sure it’s Hippocrates, also because it’s the Hippocratic oath *and* it was in the short story, I’m going to go with that.”

For the interview portion of the experiment, Participant 5 described their general approach to the task as: “I guess if I know, if I know it I’d just go with it. If I… I knew a lot of them just wasn’t really sure so I had to look it up to affirm uh which maybe I shouldn’t have done.” They also noted that they should have “trusted [themselves] more” in completing questionnaire items. Regarding confidence in their responses, Participant 5 stated that they were “pretty um confident because if I was absolutely sure off the bat I would go with it and after I looked it up I was definitely sure because … the Internet said so [laughs].” Participant 5 described their approach to search as “just like briefly looking at each of the tabs and if they each are saying the same thing then it’s probably gonna be the right answer.” Here Participant 5 appeared to be invoking matches between different pieces of information appearing in search results as a means of determining a suitable response.

Finally, in response to the interview question regarding perceptions of the stories’ relevance to the questionnaire task, Participant 5 remarked:

> “I mean they were almost all pretty err a lot of them were really related to it and I noticed that because I actually yeah when I was reading the first story I was like ooh narcolepsy is not the right term to use there I wonder what’s going on here and then I realized oh they had a question about it um and so I guess they had, the questions - some of them were completely unrelated I think but yeah I did notice a lot of relation between those kinds of things.”

In this response, Participant 5 recounted an instance in which they detected and discounted an inaccuracy during reading and expressed curiosity about what the implications of that inaccurate information might be. They then noted their recognition of a questionnaire item related to this inaccurate statement, as well as other relations between the stories and questionnaire. This suggests that Participant 5 may have developed some ideas regarding the purpose of the study, although they never made any explicit statements in this regard.

Participant 6

Participant 6 provided accurate responses to 15 critical items and a no answer response for one item (although they verbalized the correct response to that item aloud several times). They searched for seven critical items, with five searches for hard items and two for easy items. Participant 6 referenced the texts for eight items in the questionnaire task, starting with the first critical question. Participant 6 actually detected *every* inaccuracy presented (six total across the two stories). For four of the six items presented inaccurately in the stories, Participant 6 specified the inaccuracy and then refuted it either on the basis of prior knowledge or with a search. In addressing the question "What is the medical term for the inability to sleep?", for example, Participant 6 stated: "They also said narcolepsy in the story. That's insomnia. Narcolepsy is when you sleep uncontrollably." For the question "What is the name of the man who invented the smallpox vaccine?" Participant 6 said: "It's not Salk like they said cause that's polio. I don't know who invented the smallpox vaccine." After searching, Participant 6 concluded that Edward Jenner was the most suitable response. Participant 6 then added: "They are just feeding me lies," with "they" seeming to reference the characters in the stories.

For the other two items presented inaccurately in the stories, Participant 6 noted that the story contents deviated from their expectations, albeit without explicitly naming the inaccuracies. For instance, in response to the question "Who wrote the play *Romeo & Juliet*?" Participant 6 stated: "It was different in the reading comprehension but I believe it's Shakespeare." Similarly, for the question "What is the name of the project which developed the atomic bomb during World War II?" Participant 6 stated: "That's the Manhattan Project, they said something bizarre in the story though." For the two other instances in which they mentioned the story contents, Participant 6 noted one instance in which the story did not specify information necessary to

answer a question, and one instance in which the information in the story was correct (i.e., "Those are the Wright brothers. That was right in the story.").

During the interview portion of the experiment, Participant 6 emphasized their love of trivia while describing their approach to completing the task, stating:

> "I just felt like, I love trivia, I felt like it was just like trivia questions so that's how I kind of approached it. If I knew it, I knew it. Especially with the stories they told you a lot of wrong stuff so I'm like, eh, and I knew that it wasn't true and for some of them I knew right away what it was but others I had to make sure so it just kind of depends on my previous knowledge."

Two aspects of this statement are of particular interest here. First, Participant 6 noted a general affinity for the kind of information probed in the questionnaire. This general propensity towards developing and validating a body of general knowledge appears to have factored into Participant 6's exceptional performance. Secondly, Participant 6 reemphasized having detected a number of errors within the contents of the story, which suggests that noticing errors influenced their approach to the questionnaire task, which involved useful searches. That said, Participant 6 may of course have conducted the same searches even without having noticed the errors.

In response to the subsequent interview questions, Participant 6 estimated their confidence to be "like 85%." They noted "a lack of confidence in the answer if I thought I was missing more information" to be the key determining factor in whether or not they searched. As for how Participant 6 determined suitable answers from the information made available in their searches, they said:

> "Um it was kind of… it was mostly the frequency of which, of what I saw the answers I put. If I, especially for the smallpox vaccine, I did smallpox vaccine inventor or inventor of smallpox vaccine and even before I entered it into google Edward Jenner's name came up and then after I searched it Edward Jenner was the first name and then I saw Edward Jenner a bunch of other times so I was like okay. The more frequent it is the more likely I think that's the answer because it's more sources agreeing on it."

Thus, similarly to Participant 5, Participant 6 appeared to be invoking correspondence across sources as a component of their searching strategy. In response to the interview question probing the applicability of texts to the questionnaire task, Participant 6 provided a surprising answer in light of their frequent error detection:

> "I think they were pretty applicable just because there was information in the stories there but even though it was sometimes the wrong information it was kind of all right in my mind that I was thinking about it just because when I was reading the stories like Salk didn't do smallpox he did polio and um Whitney I don't think... At that point I was like I don't think he invented the steamboat cause I knew it was the cotton gin but that… I had it on my mind already."

This statement suggests that Participant 6 found utility in thinking through various topics during reading that were later probed during the questionnaire task. While, as they noted, the information presented was not always useful or valid, it prompted them to think about prior knowledge they possessed and whether it was sufficient. These contemplations may have facilitated performance by activating and thus increasing the availability of relevant prior knowledge.

Participant 7

Participant 7 provided accurate responses to 13 critical items and no answer responses to the remaining three. They searched for eight critical items, with six searches for hard items and two for easy items. They referenced the stories three times during the questionnaire task, albeit without any apparent detection of the falsehoods. Their first reference was in regard to the question "What is the last name of the brothers who flew the first plane at Kitty Hawk?" for which they said: "Oh my god. Whoa wait. I think I know from that story but I also… I don't feel like I'm right." After searching, they added: "Wright Brothers, Jesus got that wrong." Their second reference to the text pertained to the question "What is the name of the first doctor to

successfully perform a heart transfer?" They mentioned that the topic at hand "was in the story but his name wasn't in it." In Participant 7's final reference, they struggled to recall what exactly was stated in the story, noting: "Oh um it was in the story, it starts with an A. Atomic bomb project. Oh there was a project. It wasn't the Manhattan Project or… it seems like it. K. I thought it started with an A." Notably, the inaccuracy in the stories was the "Los Alamos Project" with "Alamos" thus likely being the term Participant 7 was attempting to recall.

When asked to describe their general approach to the questionnaire task, Participant 7 said:

> "I answered the ones I knew immediately. If I felt like they were easy to look up I would look them up and if they were like more ideas like the best doctor in Greece I think was one of them and that seemed, it could've - it probably could've had an answer online but it seemed like a waste of time if it wasn't."

This statement highlights two factors that influenced Participant 7's decisions to search: ease of search completion and the likelihood of finding an answer for a question. They reiterated that their search behaviors were driven "mostly" by "like the simplicity of looking them up" and that they answered, "no answer" when a search query was "too long to type." With respect to search strategies, Participant 7 said the most suitable responses were "probably the first one or if it was in the Wikipedia page." They thus appeared to have invoked a "top hit" strategy for determining appropriate answers while also relying on trusted sources. When asked about the relevance of the stories to the questionnaire task, Participant 7 replied:

> "It gave me some...it offered information that was asked about in these… about the readings like the plane one and it even mentioned like even though I knew the dog's name from the *Wizard of Oz* it mentioned that."

This statement reinforces that Participant 7 detected connections between the stories and the questionnaire task, in accordance with their verbalizations. That said, they did not indicate

noticing any of the inaccurate presentations in the stories specifically. During debriefing, Participant 7 expressed surprise at learning of the inclusion of false information in the texts, asking: “Really?! What were they?” After reviewing the inaccuracies they had been presented with, Participant 7 did not indicate recognition of them, and continued to express surprise and interest.

Participant 8

Participant 8 provided accurate responses to fourteen critical items, a no answer response to one critical item, and an inaccuracy from the stories for the remaining item. They searched for five items, all of which were hard. They did not reference the stories at all while answering the questions. Thus, the summary presented here is based entirely on their interview responses.

In describing their general approach to the questionnaire task, Participant 8 emphasized prior knowledge, stating: “The questions I just sort of - they….all of them sort of either had a memory associated with them or something.” Regarding their confidence, Participant 8 stated it “depends…Almost all of the time - I’d say about like 75 - I’d say about 60% of the time I was pretty confident in most of the answers.” However, they also added that they “never really guessed - if I wasn’t confident then I looked it up just cause yeah.” While they usually felt sure on the basis of prior knowledge, Participant 8 appeared to have used searches to boost confidence in their responses when needed. They further emphasized confidence as a key determinant of whether they searched, stating that their search behavior was driven “just [by] the degree of which I was confident.” They also noted a strong desire to know or feel certain about their answers, stating it was “gonna be really frus - like this is gonna be on my mind if I can’t figure it out [l*aughs*].”

With regard to their search strategies, Participant 8 said: “A lot of the times [the best response] would be the first thing that popped up.” They also validated the information they received from searches against their prior knowledge, stating: “A lot of the times it would be um, I would see that and I would go oh yeah.” However, they indicated that they were not always able to do so, with “like three or four questions where I was like [I’ve] never even heard of that.” Finally, in regard to their perceptions of the stories’ relevance to the questionnaire task, Participant 8 stated: “There was some directly themed questions like the smallpox, the - a lot of the ones about the inventions, the Kitty Hawk.” Thus, while they did not mention the readings at all while answering the questions, this response indicates that they did in fact detect correspondence between the two tasks. However, they did not provide any indication of detecting inaccuracies in the stories specifically. In fact, they concluded their interview by noting: “I think the only one that [the stories] really helped me on was the steamboat,” which was the item for which they reproduced a story inaccuracy. During debriefing, Participant 8 did not express any realization or prior contemplation of information from the stories having been inaccurate. Accordingly, we did not obtain any indication that Participant 8 was aware of having read inaccuracies.

Participant 9

Participant 9 provided accurate responses to fifteen critical items and a no answer response to the remaining item. They searched for six critical items (all hard) and referenced the stories four times over the course of the questionnaire task. In two of these references, they detected and rejected a story inaccuracy while answering the related question. For the question “What is the medical term for the inability to sleep?” they stated: “Insomnia but in the story it was called narcolepsy. But that’s wrong.” In response to the question “What is the system by

which plants make food for themselves?", they similarly stated: "Photosynthesis but in the story it said the Krebs cycle." In their other two references to the story content, they remembered reading about the topic but were unable to recall the specific information provided. In response to the question "What is the name of the man who invented the steamboat 'Clermont'?", for example, they noted: "I think this was in the story but I don't remember." Similarly, when asked "What is the name of the man who invented the smallpox vaccine?" they stated: "Uh it was just in the story and I know it uhh it starts with an S. Um. I'll just look it up. Oh Edward Jenner. It does not start with an S." In this instance, Participant 9 may have been trying to retrieve "Salk" which was the inaccuracy presented in the story.

When asked to describe their general approach to completing the questionnaire task, Participant 9 said: "I feel like if I... I either kind of knew it immediately or knew that I didn't know it cause they're like knowledge so you either know it or you don't." Participant 9 therefore indicated that they could readily determine their capability for answering questions. Regarding their confidence, Participant 9 said: "I think… for most of them I was pretty confident or I looked it up so then I would be also confident." This statement suggests that, similar to other participants, Participant 9 used searches to augment their knowledge and confidence. In describing why they chose to search for some questions and not others, Participant 9 stated: "If I just didn't know it I would search for it or if I wasn't like 100% sure... actually that's not true because some of the ones I wasn't 100% sure I just put it in anyway [*laughs*]." This statement reinforces uncertainty as a determining factor in their decisions to search.

With regard to search strategies, Participant 9 said, "Um I don't know I feel like I just like you kind of learn how to like sift through Google results when you use the computer a lot which I do." This suggests that their appraisals of search results were guided by tacit knowledge

acquired through internet use. Participant 9 then added: “I mean usually like the first result is pretty accurate. Google is good at answering questions [*laughs*].” Participant 9 thus evoked a “top hit” strategy for identifying appropriate responses, similar to other participants. When asked about the potential relevance of the stories for the questionnaire, they noted that accurate *and* inaccurate information had appeared in them:

> “I think that they had some information that was right and like helped you with the story like I’m pretty sure it said Hippocrates in the doctor story and then that was one of the questions and I know that that’s right. But then like they also had some inaccurate information in them.”

Participant 9 thus appeared to have detected more overlap between the two tasks than they initially noted during the questionnaire task. In addition, they appear to have developed a nuanced view of the story information, rather than focusing exclusively on either accuracies or inaccuracies.

Participant 10

Participant 10 provided accurate responses to seven critical items. For the remaining the nine critical items, they provided a no answer response to four items, an incorrect guess for two items, and inaccurate information from the stories for three items. Participant 10 did not search for any items. They referenced the stories twice during the questionnaire task, albeit not with respect to inaccuracies. Instead, Participant 10 mentioned that they recalled reading about the topic, but could not recall the specific information relevant to answer the questions. While addressing the question “What is the name of the first doctor to successfully perform a heart transfer?”, for example, they stated: “Didn’t remember that from the reading... I remember reading about heart transfer but not the name.”

In describing their general approach to the questionnaire task, Participant 10 stated: “Uh most of it just kind of knew it, just came to mind but for the stuff that didn’t I felt it ... was okay not to know.” The latter portion of this statement provides an explanation for their lack of searches. Their responses to subsequent interview items provide further explanation as to why they did not elect to search. They noted that “the ones that I answered I felt pretty confident on,” except for “like 2 or 3 that I was iffy on.” As such, Participant 10 appears to have felt little need to seek a confidence boost by validating their responses. Additionally, they expressed disinterest in the topics being queried, which may further help explain both their dearth of searches and their low rate (> 50%) of accurate responses. Participant 10 described the items for which they answered, “no answer” as “either related to fields of knowledge that I don’t find a lot of interest in or know a lot about” or knowledge that has “lost relevance to me and so I don’t remember it.”

With respect to their perceptions of the stories’ relevance to the questionnaire task, Participant 10 indicated awareness of correspondence between the story content and the questions, but without noting inaccuracies, stating:

> “Um when I was reading them there were like a couple of dates and names and stuff and I was like oh this is probably going to come back up at some point just cause - I don’t know I guess standardized tests really trained me for that um but even after that I didn’t pay enough attention to the reading I guess to memorize all of it.”

As such, Participant 10 did not indicate any awareness of having been presented with or utilizing false lures from the stories. At debriefing, they expressed surprise that they had read false information, without any apparent realization or recollection of having read false content.

Experiment 4 Results Summary & Discussion

The purpose of Experiment 4 was to investigate the choices and considerations underlying participants’ performance on the questionnaire task. While a number of experiments

both within and outside of this dissertation have examined the products associated with task completion (e.g., elevated error rates for items for which participants read inaccurate information), few have examined the processes leading to those outcomes. This fourth experiment served to supplement prior experiments by providing new insights into performance on previously examined tasks. In particular, we sought to investigate factors that influence the ways in which participants answer questionnaire items with a focus on inaccurate information use and online search. Participants in Experiment 4 were tasked with reading a set of short stories containing potentially inaccurate information. They then completed a post-reading questionnaire while verbalizing their thoughts. Think-alouds can provide insight into participants' thought processes otherwise obscured from the view of the researchers (Ericsson & Simon, 1993). Finally, participants were asked a series of interview questions to further probe their perceptions of story contents and their search behaviors.

Participants' think-alouds and interview responses were transcribed and summarized in a series of vignettes. Each vignette systematically presents a basic quantitative portrait of participants' task performance, an account of when and how (if at all) participants referenced the previously read texts, and how they responded to each of our interview items. Here we present and discuss the key themes and patterns that emerged across participants.

All participants recognized at least some connection between the stories and the questionnaire task. Given the number of critical items (16), there were numerous opportunities for participants to infer these connections. Eight of the 10 participants referenced the texts in their think-alouds for the questionnaire task and in the interview portion of the experiment. The frequency of their referents ranged from two to eight items, with no single participant referencing the stories for more than half of the trials. The remaining two participants did not

reference the stories at all while completing the questionnaire, but noted connections during the interview.

With respect to memory for story content, we observed considerable variance within and across participants' thirty-eight total textual references. These references included agreement with story contents (e.g., stating that a story contained the correct response), uncertainty about story contents (e.g., expressing an inability to recall what exactly was stated in a story *or* accurately noting that a story did not specify a given idea), disagreement with story contents (e.g., noting that an idea in a story was inaccurate *or* different from what a participant had determined to be the best response), and generic referents (i.e., simply stating that a topic was in the stories, without indicating agreement, uncertainty, or disagreement). These categorizations provide an overview of the findings. Given the variable nature of participants' responses, these categories were not mutually exclusive, with a given response potentially fitting multiple categories.

Overall we observed four instances of agreement, seventeen instances of uncertainty, twenty instances of disagreement, and four generic referents. This range of responses is to be expected as the stories contained a variety of accurate and inaccurate information, which could reasonably lead to agreements and disagreements respectively. The texts also contained neutral information, which invoked topics without providing specifications, and could be expected to generate uncertainty with regard to the stories' relevance to particular questions. Finally, generic references, while not aligning with any particular type of information presented in the stories, are well within expectations for the kinds of responses participants might produce while thinking aloud. As participants were merely asked to speak their thoughts out loud while answering each questionnaire item, vague or simplistic references to the text align perfectly well with these instructions.

While participants' inclusions of these four types of referents may not be surprising, their differing frequencies are informative. With respect to the relatively low rate of agreements with stories, readers may consider accurate information to be the norm for story content and thus unworthy of significant attention or discussion. Readers may thus be less likely to note accurate content relative to inaccurate or unspecified content as only the latter two categories may violate expectations, disrupt comprehension, or motivate additional attention and memory activity. The relatively higher rates of disagreements and expressions of uncertainty relative to rates of agreements provide some evidence for this possibility.

We also observed a low rate of generic referents, with only four responses fitting this category, all of which came from the same participant. Other participants consistently produced more specific references to the texts. Therefore, we refrain from discussing this category further. As compared to agreements and generic referents, we observed disagreements with story contents much more frequently both within and across participants. Disagreements are of particular interest to the current project as these responses can critically inform our understandings of people overcoming the potential influence of inaccurate information. One group of participants (1, 3, 4, 5, 6, and 9) noticed and rejected at least some (if not all) of the false information in the stories when answering the post-reading questions. In most cases, these participants did not reproduce inaccurate information from the stories. The sole exception was Participant 3, who continued to ponder a question after submitting his inaccurate response and subsequently searched to check whether the information from the story had indeed been inaccurate.

The second group of participants (2, 7, 8, and 10) did not appear to detect or reject any inaccurate information presented in the stories. Three of the four participants in this group

reproduced inaccurate information at least once. Participants' detection of inaccuracies thus stands out as a critical factor in determining whether they might reproduce false information. While this may not be surprising, the performance-based contrasts between the two groups and their detection is quite clear. One possibility for the experiment was that all participants would have noticed at least some inaccuracies and failed to detect others. However, we observed multiple participants overlooking even blatant inaccuracies, demonstrating no awareness or suspicion of having read falsehoods during the interview or debriefing.

These results might be taken to suggest that manipulations designed to increase noticing of inaccuracies could help attenuate the influence of inaccurate information. Prior work, however, suggests that attempts to increase noticing in and of itself are insufficient. Warning readers about the possible inclusion of inaccuracies in recently read texts has not been shown to reliably reduce inaccurate reproductions (Marsh & Fazio, 2006). Similarly, manipulations designed to increase noticing of and attention to errors (e.g., instructions to re-read or read more slowly) have not been shown to attenuate the use of inaccurate information (Fazio & Marsh, 2008; Marsh et al., 2003).

How might we reconcile the discrepancy between the importance of people's noticing inaccuracies, and the lack of benefit associated with increasing opportunities to notice inaccuracies in lab tasks? In the statements and actions of participants who noticed inaccuracies, there was evidence not only of noticing, but also of evaluation. Participants who noticed inaccuracies or doubted the veracity of story contents but lacked the background knowledge necessary to evaluate that information frequently utilized searches to address relevant questions. Noticing may thus have been an essential first step towards avoiding the use of inaccuracies that, when coupled with search engagement, led to performance benefits.

We were also interested in understanding when, why, and how participants searched for additional information. The interview questions were useful in this respect, as participants rarely verbalized information about their search behaviors while answering the questions. Five general explanations for searching emerged from participants' responses. The most common explanation invoked uncertainty about an appropriate answer or a lack of confidence in an answer under consideration. Seven participants (1, 2, 4, 5, 6, 8, and 9) noted these factors in their interview responses. The prevalence of this explanation makes intuitive sense, as searches routinely serve the function of supplementing or validating information people are currently contemplating. Searches for information that someone felt certain about or had high familiarity with would, quite often, be a poor use of time and resources.

The four other categories fit responses from smaller numbers of participants. Explanations included a general love of learning (participants 3 and 6) and subject matter specific curiosity (participants 3 and 8). While these explanations are conceptually similar, they could have distinct influences on search decisions. A general love of learning could encourage higher search rates overall, regardless of content. Subject matter curiosity, in contrast, could be associated with greater variability in search behaviors depending on topic. Another category pertained to the ease or simplicity of search (participants 7 and 8). The more effort required to conduct a useful search, the less likely it might be enacted. Similarly, the lower the perceived likelihood of finding information suitable for a response, the less likely a participant might be to search. These explanations speak to a key consideration for the search manipulations in this dissertation. Searching required additional time and cognitive resources above and beyond what was *already required* by the experimental tasks. That participants may not have always, or even

often, found it worthwhile to expend those additional resources to engage in online search is entirely reasonable given their (lack of) investment in the experimental activity.

A final explanation emerged only in the responses provided by Participant 4. She invoked search as being a natural or expected behavior, essentially asking *why wouldn't I search*? Given easy and reliable access to online information, some people might certainly take the view that search is a "natural" byproduct of needing or wanting more information and to be accurate. This contrasts with a view of search as dependent on the willingness to expend resources, and constitutes another perspective worthy of consideration in future work.

Other interview questions queried how participants determined the most suitable answer from their search results. Recall that Experiment 2 analyzed participants' search decisions by using screen recordings. Here we supplemented those findings by explicitly asking about those potential behaviors and identifying five categories of responses. The most prevalent category invoked the idea of a "top hit," with eight of the nine participants who searched noting that they often used and trusted the first Google result. This is consistent with prior work describing search behaviors in similar samples (Granka, Joachims, & Gay, 2004; Pan et al., 2007).

While this heuristic may appear overly simplistic, it can actually work quite well in many situations as evidenced by participants' collectively high rate (84.38%) of accurate responses to critical items. As some participants noted, Google excels at identifying answers to questions like those in the experimental task. While not infallible, a "top hit" heuristic can be a useful search strategy. This heuristic could also be augmented in a variety of ways, some of which participants described in their interviews. Four participants (1, 3, 7, and 8), for example, noted that they read the actual web content supplied by the first result as a means of verification, with three participants mentioning reading Wikipedia content specifically. Reading details contained in the

"top hit" could usefully provide readers with greater certainty that they had indeed found the most suitable information for their purposes.

Additionally, several participants used a "matching" strategy to increase certainty in their responses. Participants 5 and 6, for example, noted checking for consistency across sources. If multiple search results provided matching information, they felt confident that the response in question was correct. Consulting and cross-referencing multiple sources can be an effective strategy for searches and validation. Beyond just the search results, participants 4 and 8 identified matches or correspondence between searches and their prior knowledge as a means of identifying accurate responses. This strategy was applicable when participants struggled to retrieve applicable information from memory but felt they could recognize it via search. This demonstrates that searches can help ameliorate retrieval failures. Finally, participant 9 noted a tacit "feel" for useful search results acquired over frequent internet use. This "feel" may involve strategies explicated by other participants (e.g., identifying consistency across sources or reading details in search results).

As a final point, participants 1, 2, 3, and 4 added a qualifier to their explanations of how they determined suitable responses. They specified their articulated strategies as useful on *this* task but not for *every* situation. While we focused on short-answer questions probing general knowledge in these experiments, other tasks could certainly yield different results with respect to when, why, how, and to what effect people search. This will be considered at greater length in the General Discussion.

The results from Experiment 4 thus provide insight into factors that can influence people's use of inaccurate information. Noticing appeared to play a key role in determining whether inaccurate information exerted an influence on post-reading task performance. With one

exception, inaccurate reproductions were observed amongst participants who did *not* notice or suspect anything amiss in the stories. The data here also speak to participants' search strategies and considerations. We observed variability in the uptake of this manipulation across experiments, with some participants searching for many items and others opting not to search at all. Participants' interview responses shed light on the kinds of individual differences and situational factors that can influence the likelihood of search behaviors, as well as how participants think about searches.

Chapter VI: General Discussion

Many projects have shown that people will rely on the inaccurate information they read to complete post-reading tasks (e.g., Ecker, Lewandowsky, & Tang, 2010; Fazio, Barber, Rajaram, Ornstein, & Marsh, 2013; Fazio & Marsh, 2008; Fazio, Rand, & Pennycook, 2019; Frenda, Nichols, & Loftus, 2011; Gerrig & Prentice, 1991; Johnson & Seifert, 1994; Loftus, 2005; Loftus & Hoffman, 1989; Marsh & Fazio, 2006; Marsh, Meade, & Roediger, 2003; Okado & Stark, 2005; Rapp & Braasch, 2014; Rapp, Hinze, Kohlhepp, & Ryskin, 2014). Given the undesirable consequences associated with using false information, a variety of interventions designed to attenuate these effects have been tested. However, people's use of inaccurate information often persists despite interventions focused on instructions, procedures, and experimental materials. This dissertation seeks to contribute to the literature on people's reliance on inaccurate information by testing the efficacy of an activity commonly employed in the real world. It examined whether affording participants the opportunity to search online while answering questions might reduce their use of previously read inaccurate information. While people can often consult outside resources during everyday reading, this access is typically restricted in experimental examinations. Examining people's use of searches, and the consequences of searching, with previously applied methodologies and materials, extends our understandings of how reliance on inaccurate information might be attenuated, and offers a more externally valid estimate of reliance than has been reported in previous work.

Review of experimental outcomes

Experiment 1 constituted our first examination of whether offering the opportunity to search might influence people's use of previously read inaccurate information. 216 participants recruited through Amazon's Mechanical Turk read fictional stories containing potentially

inaccurate information about real world concepts and events. After reading they were tasked with answering general knowledge questions relevant to the information manipulated in the stories. Half of the participants were instructed that they could search online for answers to the questions if they wished, while the remaining half were instructed not to search. We predicted that participants allowed to search would reproduce inaccurate information less frequently and would produce correct responses more frequently, than would participants who were not allowed to search. Because participants' searches were entirely at their own discretion, support for this prediction depended on whether participants opted to search at all.

Analyses of participants' self-reports indicated that they routinely chose to search when answering questions, and did so especially often for hard items. Supporting our hypothesis, participants in the search condition reproduced significantly less incorrect information from the stories, and significantly more correct responses overall, than did participants who were not allowed to search. Participants in the search allowed condition also produced significantly fewer no answer responses than did participants in the no search allowed condition. Follow-up analyses indicated that the benefits of searches were attributable to actual search activity rather than to a more evaluative or critical mindset instantiated by the opportunity to search, as benefits were specifically observed for questions for which participants searched. Experiment 1 provided promising evidence that readers can usefully utilize searches, without any external incentive or extended prompting to do so, and that searches support post-reading performance.

We sought to replicate these patterns in Experiment 2 with a different sample and experimental setting. Participants in Experiment 2 were recruited from Northwestern University and completed the study in a lab setting. This allowed for assessing whether beneficial patterns similar to those in Experiment 1 would emerge in another context. The experiment also assessed

participants' activity in more depth by screen recording their search behaviors. Analyses of search rates and questionnaire performance obtained outcomes closely mirroring those in Experiment 1. Participants in Experiment 2 exhibited the same general search pattern of seeking out information more often for hard than for easy items, with their overall rates of search higher than in the previous experiment. We observed significant reductions in reproductions of inaccurate information when participants were allowed to search as compared to participants who were not allowed to search. Participants who were allowed to search also produced more correct responses than did participants who were not allowed to search, and were less likely to produce no answer responses.

Participants' self-reported searches seemed to offer a reliable estimate, based on their corresponding screen recordings. They typically used keyword searches and consulted Google search result pages rather than individual webpages when considering their responses. The apparent purpose of search most often involved participants seeking to augment their existing knowledge. Less often, participants' searches suggested other purposes such as double-checking or correcting answers. These outcomes indicated the search instructions generally functioned as intended, with participants using searches to access and certify valid information.

Having replicated key response patterns across Experiments 1 and 2, we next examined whether search behaviors, and their accompanying benefits, might be further promoted. To do this, we offered some participants direct links to relevant information, rather than requiring them to engage in a self-directed search. Participants in Experiment 3 were randomly assigned to either an augmented search condition, in which they received links to information relevant for each question in turn, or a standard search condition similar to that used in Experiment 2. The participant sampling and methods were otherwise the same as in Experiment 2. We hypothesized

that participants in the augmented search condition would conduct more searches than would participants in the standard search condition, and as a result, reproduce fewer inaccuracies and provide more correct responses.

Analyses of search rates and questionnaire performance partially supported the hypotheses. Participants in the augmented search condition were more likely to search than were participants in the standard search condition. Participants in the augmented search condition also produced more correct responses, as well as fewer no answer and inaccurate - other responses. While we did not have specific hypotheses for these latter two response types, they highlight benefits associated with the search support provided via the links. However, reproductions of inaccurate information did not differ across search conditions, with participants' use of inaccuracies persisting at equivalently low rates. Participants thus benefited from our augmented search manipulation, albeit not with respect to the particular response type of greatest interest to the project. This suggests the need for evaluating other manipulations that might further enhance the efficacy of search.

Finally, Experiment 4 examined the choices and considerations that might underlie participant's performance on the questionnaire task. While many projects have examined the implications of exposures to inaccurate information on post-reading tasks, few have examined participants' thought processes during task completion. Participants in Experiment 4 completed a subset of the same tasks used in the previous experiments (i.e., reading two of the four stories, and answering 37 of the 74 questionnaire items). All participants were given the option to search online for support on the questionnaire task if they wished. To collect participants' thought processes, all participants were asked to speak their thoughts aloud while answering the

questions. They were also asked a series of interview questions after the task to further their use of text information and search behaviors.

Analyses for Experiment 4 provided insights into factors that might contribute to people's use of inaccurate information, as well factors underlying search considerations. Whether or not participants detected story inaccuracies, for example, was shown to be a key determinant of later reliance. All instances of inaccurate reproductions, save for one, obtained with participants who did not report detecting that the stories contained falsehoods. Participants routinely cited uncertainty or a lack of prior knowledge as key determining factors in decisions to search. Some participants also cited the ease of searching or the likelihood of obtaining useful information as influential factors in their search decisions. When they chose to search, participants often used a "first result" or "top hit" heuristic for identifying information to use, with some participants further noting correspondences between listed search results, and between those results and prior knowledge, as methods of evaluating the suitability of potential responses.

Implications

These findings have implications for contemporary discussions of the influence of inaccurate information during everyday reading, for explanations as to why that influence obtains, and for accounts of validation during reading. The results obtained across the experiments indicate that people may be more effective at avoiding the potential use of inaccurate information than previous studies have demonstrated. Evidence of people's use of inaccuracies has primarily been derived from lab studies in which participants complete post-reading tasks on the basis of memory alone. Restricting participants' access to outside information and resources can be valuable for understanding the underlying cognitive processes associated with exposure to inaccurate information (Rapp & Braasch, 2014; Rapp, 2016).

However, those restrictions may offer overestimates of how often reliance on inaccurate information occurs in everyday reading, given people often can consult supplemental resources rather than relying solely on memory.

Critically, the current project indicates that readers *do* routinely search for information when given the opportunity, and *without* any external incentives. Moreover, searches proved beneficial to readers in several ways, with participants less likely to reproduce inaccuracies when given the option to search relative to when searches were disallowed. Given, again, that outside resource access is characteristic of contemporary everyday reading, findings from the current project indicate that rates of reliance on inaccurate information may actually be lower than has previously been described.

Another key implication of this work pertains to explanations of inaccurate information use. Explanations often appeal in part to the heightened availability of recently processed inaccurate information in short-term memory, in contrast to accurate information in long-term memory that may have been acquired in other contexts (e.g., Oppenheimer, 2008). The ready accessibility of recently read text information can promote its use on subsequent tasks. Recently read story information may have less of an effect if people can in turn readily access accurate information. Participants in the current project often choose to do just that, particularly with respect to topics for which they lacked sufficient background knowledge. The increased availability of valid information obtained through outside resources in turn led to improvements in post-reading performance, with participants producing accurate responses for the majority of corresponding items for which they elected to search. When participants did not search for additional information, they were more likely to supply inaccuracies as answers

The current project also has implications for contemporary accounts of validation. Historically, such accounts have focused on the use or neglect of prior knowledge. Prior knowledge is certainly integral to validation (Cook & O'Brien, 2014; Richter, Schroeder, & Wöhrmann, 2009) as the findings reported here consistently indicate. Within and across experiments, we observed large effects of item difficulty, with participants performing better on easy items for which they held prior knowledge than on hard items for which they lacked prior knowledge. However, these findings demonstrate that validation need not be limited to information from prior knowledge. People can consult outside resources to verify information, which is especially useful when they lack the prior knowledge necessary for independent validation. Given that everyday reading situations typically allow for outside resource access, accounts of validation processing may benefit from greater inclusion and consideration of when, how, and how often people consult outside resources in the service of validation. Doing so could both enhance the external validity of these accounts and usefully inform associated interventions.

Limitations and future directions

Findings from these experiments can inform descriptions of, explanations for, and conceptualizations of validation relevant to inaccurate information use. There is nevertheless a need for continued work on these issues, building off of and informed by the current project. Across the four experiments, we used the same set of stories, experimentally manipulated statements, and questionnaire items. This design decision was intentional, as it allowed for conceptual comparisons across experiments within and outside of this dissertation. Other types of information could be tested using the same general research design to determine the generalizability or specificity of response patterns and effects. The information manipulated in the experiments was also characterized by well-defined answers readily available via online

search. This was a similarly intentional feature of the current project allowing for search to be a manageable and reasonably successful task if and when participants chose to undertake it. Topics with greater ambiguity could yield differences in search propensities, strategies, and outcomes. As participants in Experiment 4 reported, the perceived difficulty of a search might influence the likelihood of conducting it. Conversely, if participants perceive task topics to be especially relevant, interesting, or important, search frequency and depth might be expected to increase, as suggested by findings from Experiment 4 highlighting the role of curiosity in search decisions.

A related limitation pertains to the nature of the experimental task. Short-answer questions were used to align with prior work. There are a variety of other means by which potential use of inaccuracies could be assessed, including validity judgments, multiple-choice measures, and discussion prompts, among other possibilities. Moreover, the nature of the assessment task could have important implications for both estimating rates of inaccuracy use and of search. For example, if participants were tasked with not only answering test items but also providing an explanation for their responses (e.g., how they arrived at an answer) more careful responding and more frequent engagement in search could reasonably be expected to obtain. Utilizing other types of tasks in future work could enhance and expand our understandings of how engagement in search can impact post-reading performance.

Finally, we observed significant variability participants' decisions about searching. What accounts for that variability, however, was neither directly nor systematically addressed here. The results of Experiment 4 suggest individual differences that might underlie this variability with state, trait, and task dependent curiosity, information seeking habits, and perceptions of the task all related to participants' responses. Investigating how these and other pertinent individual differences might help explain variability constitutes another fruitful avenue for future

work. High levels of curiosity, for instance, could be hypothesized to be associated with high rates of search engagement.

Conclusion and summary

The experiments in this dissertation provide evidence of the benefits of search with respect to people's exposures to and use of inaccurate information. Search is commonly studied in research on consumer decision making (e.g., De Los Santos, 2018; Johnson, Moe, Fader, Bellman, & Lohse, 2004) and school-based learning (see Rieh et al., 2016 for a review), among other topic areas. In studies of inaccurate information use, however, examinations of searches are rare. The findings from this dissertation suggest this should change. Granting participants the option to engage in search reliably attenuated their reproductions of previously read inaccuracies. We also identified reliable patterns in search engagement, with participants searching more often for unfamiliar than familiar information, and searching more often when information accessibility was enhanced with direct links. These patterns corresponded to performance benefits for participants on post-reading tasks. Future studies examining exposures to and use of inaccurate information should consider the implications of people's access to and use of outside resources. Search is a prevalent aspect of contemporary reading and information processing that, as the current findings show, should be incorporated into accounts of the consequences of exposures to inaccurate information.

Table 1

Sample Story Passages and Corresponding Test Items

Sample Passage 1

Accurate Version: “Here’s what I need you to do. Go to Heathrow airport in London and use my plane ticket to fly back to Kentucky. That’s all. I just need it to look as if I left Europe. Then, if anyone should even think to track me, the last place they will look for me is here.”

Inaccurate Version: “Here’s what I need you to do. Go to Heathrow airport in Dublin and use my plane ticket to fly back to Kentucky. That’s all. I just need it to look as if I left Europe. Then, if anyone should even think to track me, the last place they will look for me is here.”

Neutral Version: “Here’s what I need you to do. Go to Heathrow airport and use my plane ticket to fly back to Kentucky. That’s all. I just need it to look as if I left Europe. Then, if anyone should even think to track me, the last place they will look for me is here.”

Corresponding Test Item: Which city is Heathrow Airport located in?

Sample Passage 2

Accurate Version: Billy had thought about telling his father that he was embarrassed, but he just couldn’t bring himself to do it. Besides, he had known what his father would say- “Embarrassed?! Why in the world would you be embarrassed by new inventions? Sure, I know it’s scary to show the world something it has never seen before, but you should be proud! Like when Fulton was so happy with the steamboat he created, he gave it a name, the Clermont.”

Inaccurate Version: Billy had thought about telling his father that he was embarrassed, but he just couldn't bring himself to do it. Besides, he had known what his father would say- "Embarrassed?! Why in the world would you be embarrassed by new inventions? Sure, I know it's scary to show the world something it has never seen before, but you should be proud! Like when Whitney was so happy with the steamboat he created, he gave it a name, the Clermont.

Neutral Version: Billy had thought about telling his father that he was embarrassed, but he just couldn't bring himself to do it. Besides, he had known what his father would say- "Embarrassed?! Why in the world would you be embarrassed by new inventions? Sure, I know it's scary to show the world something it has never seen before, but you should be proud! Like when that guy was so happy with the steamboat he created, he gave it a name, the Clermont."

Corresponding Test Item: What is the name of the man who invented the steamboat "Clermont"?

Sample Passage 3

Accurate Version: Finally, it was the night before my first day at medical school. I was plagued with insomnia, I just couldn't get to sleep no matter how hard I tried. So I took a sleeping tablet, which put me to sleep but meant that I had the hardest time getting up in the morning.

Inaccurate Version: Finally, it was the night before my first day at medical school. I was plagued with narcolepsy, I just couldn't get to sleep no matter how hard I tried. So I took a sleeping tablet, which put me to sleep but meant that I had the hardest time getting up in the morning.

Neutral Version: Finally, it was the night before my first day at medical school. I was plagued with an inability to sleep, I just couldn't get to sleep no matter how hard I tried. So I took a sleeping tablet, which put me to sleep but meant that I had the hardest time getting up in the morning.

Corresponding Test Item: What is the medical term for the inability to sleep?

Table 2

Critical Item List

Easy Items

1. Who wrote the play *Romeo & Juliet*?
2. What is a horse covered in black and white stripes called? *
3. What is the medical term for the inability to sleep?
4. What is the name of the skirt worn by men in Scotland? *
5. What is the term for stones that contain the remains of plants and animals?
6. What is the name of the precious red stone?
7. What is the system by which plants make food for themselves? *
8. What is the ailment which exhibits symptoms including chronic headaches and nausea?
9. What are physicians who specialize in skin ailments called? *
10. What is the molten rock that originates from a volcano called? *
11. What is the capital of France?
12. What is the last name of the brothers who flew the first plane at Kitty Hawk? *
13. What is the name of Dorothy's dog from *The Wizard of Oz*? *
14. What is the name of Tarzan's girlfriend? *
15. What is the name of the first man on the moon? *
16. What are physicians who specialize in cutting the body called?

Hard Items

17. What is the name of the man who invented the steamboat "Clermont"?
18. What is the name of the largest desert on Earth? *

19. What is the capital of Kentucky? *

20. What is the name of the man who invented the smallpox vaccine?

21. What is the name of the project which developed the atomic bomb during World War II?

22. What is the name of the man who invented the telegraph?

23. What is the capital of Delaware?

24. Who published the idea that the Earth revolves around the sun in 1543? *

25. What is the name of the scientist who discovered radium?

26. What was the name of the first lunar module on the moon? *

27. What is the name of the first doctor to successfully perform a heart transfer?

28. What is the name of the last planet to be discovered? *

29. Which city is Michelangelo's famous "David" located in? *

30. Who was the most famous Greek doctor?

31. Which city is Heathrow Airport located in? *

32. What is the name of the painter of "American Gothic"? *

* Denotes items included only in Experiments 1-3. Experiment 4 utilized only half of the item list.

Table 3

Response Rates as a Function of Search, Experiment 1

Item Difficulty	Self-Reported Search Behavior	Inaccurate – Lure	Accurate	Inaccurate - Other	No Answer
Easy	Searched	0.85%	80.90%	15.04%	3.21%
	Unsearched	1.00%	90.65%	4.82%	3.58%
Hard	Searched	0.82%	90.00%	8.54%	0.64%
	Unsearched	17.19%	42.27%	18.83%	21.73%

Table 4

Response Rates as a Function of Search, Experiment 2

Item Difficulty	Self-Reported Search Behavior	Inaccurate – Lure	Accurate	Inaccurate - Other	No Answer
Easy	Searched	0.00%	94.33%	3.43%	2.24%
	Unsearched	1.25%	91.49%	5.12%	2.14%
Hard	Searched	2.23%	84.30%	12.23%	1.24%
	Unsearched	25.13%	45.47%	18.41%	10.99%

Table 5

Search Codes Applied to Screen Recordings, Experiment 2

Search Code 1	**Code Description**	**Count**
Validity of Search Report *Match*	Accurate correspondence with self-reported search	495 cases
Validity of Search Report *Mismatch*	Inaccurate correspondence with self-reported search	7 cases

Search Code 2	**Code Description**	**Count**
Search Type *Exact Question*	Search framed as exact question posed in questionnaire	20 cases
Search Type *Rephrased Question*	Search framed as question similar to that in questionnaire	90 cases
Search Type *Keywords*	Search consisted of keywords rather than a question	415 cases

Search Code 3	**Code Description**	**Count**
Search Destination *Search suggestions*	Search contents were limited to search bar suggestions	39 cases
Search Destination *Google search results*	Search contents consisted of search results pages(s)	457 cases
Search Destination *Webpages*	Search contents consisted of one or more individual webpages	44 cases

Search Code 4	**Code Description**	**Count**
Search Role *Augmentation*	Search involved retrieving and supplying new information	432 cases
Search Role *Incompletion*	Search ended without an answer being supplied	10 cases
Search Role *Verification*	Search involved checking a pre-supplied response	32 cases
Search Role *Emendation*	Search involved correcting a pre-supplied response	35 cases
Search Role *Retroaction*	Search occurred after response submission	8 cases

Table 6

Response Rates as a Function of Search, Experiment 3

Augmented Search Condition

Item Difficulty	Self-Reported Search Behavior	Inaccurate – Lure	Accurate	Inaccurate - Other	No Answer
Easy	Searched	0.72%	89.78%	9.50%	0.00%
	Unsearched	1.30%	95.78%	2.92%	0.00%
Hard	Searched	1.57%	87.61%	10.50%	0.32%
	Unsearched	23.57%	62.73%	13.70%	0.00%

Standard Search Condition

Item Difficulty	Self-Reported Search Behavior	Inaccurate – Lure	Accurate	Inaccurate - Other	No Answer
Easy	Searched	0.51%	88.40%	9.47%	1.61%
	Unsearched	2.63%	92.09%	3.73%	1.55%
Hard	Searched	1.57%	81.77%	14.94%	1.72%
	Unsearched	15.52%	50.38%	20.61%	13.49%

Table 7

Interview Protocol, Experiment 4

Question 1. How would you characterize your general approach to completing the questionnaire task?

(*Asked of all participants*)

Question 2. To what extent were you confident in the accuracy of your answers?

(*Asked of all participants*)

Question 3. You went about answering questions in different ways. What led you to search for some questions and not others?

(*Asked if participant searched for any questions*)

Question 4. Thinking back to instances in which you chose to search online for answers, how did you determine the "right" answer from the search results?

(*Asked if participant searched for any questions*)

Question 5. Tell me about the instances in which you chose to answer, "no answer." What made determining a suitable answer challenging?

(*Asked if participant had any unanswered questions*)

Question 6. During the questionnaire task, you mentioned that the reading helped you determine an answer. To what extent did you consider the readings applicable to the questionnaire task?

(*Asked if participant previously referenced the texts*)

Question 7. We had you read a series of stories before completing the questionnaire task. To what extent did you consider the readings applicable to the questionnaire task?

(*Asked if participant made no references to the texts*)

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1061-1070.

Appel, M., & Richter, T. (2007). Persuasive effects of fictional narratives increase over time. *Media Psychology*, *10*, 113–134.

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition*, *3*, 77-88.

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility. *Psychological Science, 20*, 66-73.

Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes*, *51*, 26-49.

Dhillon, M. K. (2007). Online information seeking and higher education students. In Chelton, M. and Cool, C. (Eds.), *Youth information- seeking behavior II: Context, theories, models, and issues* (pp. 165–205). Lanham, MD: Scarecrow Press.

Donovan, A. M., Theodosis, E., & Rapp, D. N. (2018). Reader, interrupted: Do disruptions during encoding attenuate misinformation effects? *Applied Cognitive Psychology*, *32*, 775-786.

Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 813-825.

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, *6*, 185–192.

Ecker, U. K. H., Lewandowsky, S., Cheung, C. S. C., & Mayberry, M. T. (2015). He did it! She did it! No, she did it! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, *85*, 101-115.

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*, 1087-1100.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Eslick, A. N., Fazio, L. K., & Marsh E. J. (2011). Ironic effects of drawing attention to story errors. *Memory*, *19*, 184-191.

Fazio, L. K., Barber, S. J., Rajaram, S., Ornstein, P. A., & Marsh, E. J. (2013). Creating illusions of knowledge: Learning errors that contradict prior knowledge. *Journal of Experimental Psychology: General*, *142*, 1–5.

Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, *15*, 180–185.

Forgas, J. P., Laham, S. M., & Vargas, P. T. (2005). Mood effects on eyewitness memory: Affective influences on susceptibility to misinformation. *Journal of Experimental Social Psychology*, *41*, 2005.

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science*, *20*, 20-23.

Gerrig, R. J., & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science*, *2*, 336–340.

Granka, L., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 478– 479). New York: ACM Press.

Hao, Q., Wright, E., Barnes, B., & Branch, R. M. (2016). What are the most important predictors of computer science students online help-seeking behaviors? *Computers in Human Behavior*, *62*, 467-474.

Hinze, S. R., Slaten, D. G., Horton, W. S., Jenkins, R., & Rapp, D. N. (2014). Pilgrims sailing the Titanic: Plausibility effects on memory for misinformation. *Memory & Cognition*, *42*, 1-20.

Higham, P. A., Blank, H., & Luna, K. (2017). Effects of postwarning specificity on memory performance and confidence in the eyewitness misinformation paradigm. *Journal of Experimental Psychology: Applied*, *23*, 417-432.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When discredited information in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1420-1436.

Johnson, H. M., & Seifert, C. M. (1998). Updating accounts following a correction of misinformation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1483-1494.

Kendeou, P., & O'Brien, E. J. (2014). The knowledge revision components (KReC) framework: Processes and mechanisms. In Rapp, D. N., and Braasch, J. L. G. (Eds.), *Processing*

inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences (pp. 353–379). Cambridge, MA: MIT Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review*, *95*, 163-182.

Lai, K., Lee, Y. X., Chen, H., & Yu, R. (2017). Research on web search behavior: How online query data inform social psychology. *Cyberpsychology, Behavior, and Social Networkin*g, *20*, 596-602.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*, 106-131.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*, 550-572.

Loftus, E.F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning and Memor*y, *12*, 361–366.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: the creation of new memories. *Journal of Experimental Psychology: General*, *118*, 100-104.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 19-31.

Marsh, E. J. (2004). Story stimuli for creating false beliefs about the world. *Behavior Research Methods, Instruments, & Computers*, *36*, 650–655.

Marsh, E. J., Cantor, A. D., & Brashier, N. M. (2016). Believing that humans swallow spiders in their sleep: False beliefs as side effects of the processes that support accurate knowledge. *Psychology of Learning and Motivation: Advances in Research and Theory*, *64*, 93-132.

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, *34*, 1140–1149.

Marsh, E. J., Meade, M. L., & Roediger, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, *49*, 519–536.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247-287.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, *19*, 338–368.

Okado, Y., & Stark, C. E. L. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory*, *12*, 3-11.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Science*, *12*, 237-241.

Ozkara, B. Y., Ozmen, M., & Kim, J. Y. (2016). Exploring the relationship between information satisfaction and flow in the context of consumers' online search. *Computers in Human Behavior*, *63*, 844-859.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, *12*, 801–823.

Prentice, D. A., & Gerrig, R. J. (1999). Exploring the boundary between fiction and reality. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 529-546). New York, NY, US: Guilford Press.

Prentice, D. A., Gerrig, R. J., & Bailis, D. S. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin & Review*, *4*, 416–420.

Rapp, D. N. (2016). The consequences of reading inaccurate information. *Current Directions in Psychological Science*, *25*, 281-285.

Rapp, D. N., & Braasch, J. L. G. (Eds.). (2014). *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*. Cambridge, MA: MIT Press.

Rapp, D. N., & Donovan, A. M. (2017). Routine processes of cognition result in routine influences of inaccurate content. *Journal of Applied Research in Memory and Cognition*, 6, 409-413.

Rapp, D. N., Hinze, S. R., Kohlhepp, K., & Ryskin, R. A. (2014). Reducing reliance on inaccurate information. *Memory & Cognition*, *42*, 11-26.

Rapp, D. N., Jacovina, M. E, & Andrews, J. J. 2014. Mechanisms of problematic knowledge acquisition. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing Inaccurate Information: Theoretical and Applied Perspectives from Cognitive Science and the Educational Sciences* (pp. 181-202). Cambridge, MA: MIT Press.

Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension: An integrative view of reading. *Current Directions in Psychological Science*, *14*, 276–279.

Rich, P. R., & Zaragoza, M. S. (2016). The continued influence of implied and explicitly stated misinformation in news reports. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 62-74.

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, *96*, 538–558.

Rieh, S. Y., Collins-Thompson, K., Hansen, P., & Lee, H. (2016). Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science*, *42*, 19-34.

Rieh, S., Kim, Y. & Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing and Management*, *48*, 1136–1150.

Storm, B. C. (2011). Retrieval-induced forgetting and the resolution of competition. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 89–105). New York, NY: Psychology Press.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from Nelson and Narens (1980) norms. *Behavioral Research Methods*, *45*, 1115-1143.

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communicate*, *33*, 460-480.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352-373.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, *39*, 299–316.

Walhout, J., Oomen, P., Jarodzka, H., & Brand-Gruwel, S. (2017). Effects of task complexity on online search behaviors of adolescents. *Journal of the Association for Information Science & Technology*, *16*, 1449-1461.

Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology*, *53*, 134-144.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattribution and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 934-945.

Zhou, M. (2015). SCOOP: A measurement and database of student online search behavior and performance. *British Journal of Educational Technology*, *46*, 928-931.

Vita

The author of this dissertation, Amalia Donovan, was born in Arlington, Virginia in 1991. She attended the University of Southern California for undergraduate study, graduating with a Bachelor of Arts degree in Psychology in 2013. She completed a Master of Arts degree in the Learning Sciences program at Northwestern University in 2014. In 2015, she enrolled in the Learning Sciences doctoral degree program, also at Northwestern University. During graduate school, she published several journal articles and book chapters with her advisor, Professor David Rapp, and other colleagues within and outside of the university. A list of these publications appears below.

Published journal articles & book chapters

- Rapp, D.N., Donovan, A.M., & Salovich, N.A. (In press). Assessing and modifying knowledge: Facts vs. constellations. In P. Kendeou, P. Van Meter, A. List, & D. Lombardi (Eds.), *The Handbook of Learning from Multiple Representations and Perspectives*. New York, NY: Routledge.

- Donovan, A.M., Theodosis, E., & Rapp, D.N. (2018). Reader, interrupted: Do disruptions during encoding attenuate misinformation effects? *Applied Cognitive Psychology, 32,* 775-786.

- Donovan, A.M., Zhan, J., & Rapp, D.N. (2018). Supporting historical understandings with refutation texts. *Contemporary Educational Psychology*, *54*, 1-11.

- Donovan, A.M., & Rapp, D.N. (2018). Updating of character information when reading multiple texts for pleasure. In J. Braasch, I. Bråten, & M. McCrudden (Eds.), *Handbook of multiple source use* (pp. 303-319). New York, NY: Routledge.

- Rapp, D.N., & Donovan, A.M. (2018). The challenge of overcoming pseudoscientific ideas. *Behavior Therapist*, *41*, 4-11.

- Rapp, D.N., & Donovan, A.M. (2017). Routine processes of cognition result in routine influences of inaccurate content. *Journal of Applied Research in Memory and Cognition*, *6*, 409-413.