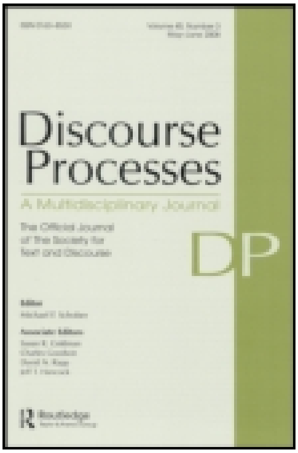


Discourse Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hdsp20

# Memory for Scientific Arguments and Their Sources: Claim-Evidence Consistency Matters

Brent Steffens [a], M. Anne Britt [a], Jason L. Braasch [b], Helge Strømsø [c] & Ivar Bråten [c]

[a] Department of Psychology, Northern Illinois University

[b] Department of Psychology, University of Memphis; Department of Educational Research, University of Oslo, Oslo, Norway

[c] Department of Educational Research, University of Oslo, Oslo, Norway

Accepted author version posted online: 28 Oct 2013.Published online: 09 Jan 2014.

*Discourse Processes*, 51:117–142, 2014

ISSN: 0163-853X print/1532-6950 online
DOI: 10.1080/0163853X.2013.855868

Routledge
Taylor & Francis Group

# Memory for Scientific Arguments and Their Sources: Claim–Evidence Consistency Matters

Brent Steffens and M. Anne Britt
*Department of Psychology*
*Northern Illinois University*

Jason L. Braasch
*Department of Psychology*
*University of Memphis;*
*Department of Educational Research*
*University of Oslo, Oslo, Norway*

Helge Strømsø and Ivar Bråten
*Department of Educational Research*
*University of Oslo, Oslo, Norway*



We investigated whether memory for scientific arguments and their sources were affected by the appropriateness of the claim–evidence relationship. Undergraduates read health articles in one of four conditions derived by crossing claim type (causal with definite qualifier, associative with tentative qualifier) and evidence type (experimental, correlational). This manipulation produced articles that overstated the results of a study and articles that understated their results, along with appropriate controls for each. We found that evidence and, to a lesser extent, source information was recalled more poorly for articles that overstated results (i.e., causal claims using correlational evidence) than for those where evidence was appropriate (i.e., causal claim with experimental evidence). Readers rejected these overstatements based on

Correspondence concerning this article should be addressed to Brent Steffens, Northern Illinois University, Department of Psychology, Psychology-Computer Science Building, Room 400, DeKalb, IL 60115, USA. E-mail: bsteffens@niu.edu

the study design rather than reprocessing the text. In contrast, understatements (tentative claim, experimental evidence) were recalled just as well as their appropriate control; however, the target content was reprocessed at a higher rate. These findings suggest that readers may remember an inappropriate definite conclusion but fail to recall the evidence used to support it.

## INTRODUCTION



People frequently read health and science news articles for entertainment and to learn about issues that affect their lives. Publishers of such articles, however, sometimes make it difficult for readers to obtain accurate information and draw evidence-based conclusions. Given the need to attract readership and sell advertising, authors and editors sometimes make overly strong generalizations and draw conclusions that go beyond the presented evidence or fail to mention important limitations of the study (Jensen, 2008; Yavchitz et al., 2012). A real-world example comes from an online news article that presented details from a correlational study (e.g., a survey finding that students who ate school lunch were much more likely to be obese compared with those who brought lunch from home) but drew a causal conclusion (e.g., "School lunches *cause* childhood obesity") rather than a less compelling but more appropriate conclusion ("School lunches *may* be *linked* to childhood obesity") (YGoY, 2011). Another related problem is that such conclusions might lack qualifiers or provide inappropriate ones that further contribute to the sense that a research result is more certain than it actually is. Such problems may be exacerbated by misapprehensions among lay science readers, both as to how one determines whether a conclusion is appropriate for a given type of evidence and in understanding scientific certainty and how it is appropriately expressed. In the former case, a reader's epistemology may lead to a preference for a simple view of causation (Bråten, Strømsø, & Samuelstuen, 2008). In the latter case, some readers view statements that use certainty hedges (e.g., "may," "likely") as meaning that a conclusion is unreliable. In either scenario, if readers are unable to notice when a claim is inappropriate based on the associated evidence, they may recall the inappropriate conclusion later and even use it to direct their future behavior.

The studies presented here investigated how well people were able to recall core elements of scientific articles that either overstated or understated their research results relative to an appropriate baseline. Two experiments varied evidentiary appropriateness and use of certainty hedges to produce claims that either overstated or understated the results of a reported study. Each study then measured participants' recall of the supporting evidence and source information. We considered predictions based on two different reading mechanisms. If the mismatch results in a coherence break, then to the extent that readers can create an inference to restore coherence, they will better recall the article details in the overstatement or understatement condition compared with the appropriate argument control. In contrast, if readers organize the texts' argument information according to a knowledge structure, then one would expect poorer recall in the cases where this is impeded by the claim–evidence mismatch.

## Reading Popular Science Articles

Popular reports of science and health issues frequently present research studies and conclusions from those studies (Jensen, 2008; Yavchitz et al., 2012). More specifically, such writing frequently takes the form of an argument in which the study provides support for a concluding statement. Often, an abbreviated form of the conclusion becomes the headline. For instance, in the above school lunch example, the reporter presents details of a correlational study of kids who ate school lunches and the incidence of childhood obesity. Many of the websites that reported this story correctly stated a headline that was consistent with the study: School lunches *may* be *linked* to childhood obesity. Some sites, however, used a headline that overstated the results: School lunches *cause* childhood obesity. Only a reader who was aware of the type of evidence required to support a causal assertion and who was reading carefully would be able to detect the inconsistency and recognize it as an overstatement.

## Reading Arguments

Research has found that during reading people monitor the truth value of information with respect to their prior knowledge and beliefs and against their model of the current situation derived from the text itself (Isberner & Richter, this issue; Maier & Richter, 2013; Richter, Schroeder, & Wöhrmann, 2009; Schroeder, Richter, & Hoever, 2008; Singer, 2006; Singer & Doering, this issue). Argument comprehension is a potentially useful genre for examining truth validation during comprehension. The purpose of an argument is to increase the believability of a claim by providing supporting information (Toulmin, 1958; Voss & Means, 1991), and the truthfulness of that support is an important criterion for judging support (Blair & Johnson, 1987; Shaw, 1996). Voss, Fincher-Kiefer, Wiley, and Silfies (1993) found that readers could make an agreement judgment about a claim as quickly as they could judge whether the statement was meaningful. Readers also very quickly activate knowledge and beliefs associated with a claim immediately after reading it. Therefore, the connection of comprehending and evaluating truth are naturally closely connected in arguments.

An argument may be minimally defined as an assertion that is perceived to be unfamiliar or controversial (relative to some audience) that is supported by at least one reason or evidentiary statement (Toulmin, 1958; Voss & Means, 1991). Schematically, one might envision the structure of an argument as a set of connected statements around a central claim. Prior research with other genres has shown that schema-consistent information leads to a more complete and coherent structure and that structure acts as a guide during retrieval (Bower, Black, & Turner, 1979; Bransford & Johnson, 1972; Dee-Lucas & Larkin, 1988; Dixon, Faries, & Gabrys, 1988; Mandler & Johnson, 1977; Meyer, Brandt, & Bluth, 1980; Meyer & Freedle, 1984; Staresina, Gray, & Davachi, 2009; Stein & Glenn, 1979). Researchers have found a similar pattern with arguments. The claim of an argument limits the type of supporting and opposing relationships and thus might be considered to guide a reader's comprehension and evaluation of arguments (Britt & Larson, 2003; Chambliss, 1995; Larson, Britt, & Kurby, 2009; Voss et al., 1993).

Such structural constraints have processing consequences. Britt and Larson (2003), for instance, found that short, two-clause arguments were read faster when they were presented in a claim–reason order compared with a reason–claim order. They also found that readers recalled claims better than reasons and frequently recalled arguments in a claim–reason order even when originally presented in a reason–claim order. The claim, especially the main predicate of the claim statement, has a guiding role in establishing the coherence of an argument (Britt & Larson, 2003; Larson et al., 2009). Attention to the main predicate of the claim statement is important because it constrains other parts of the argument and determines the types of statements that can be used to support or contradict the claim (Larson et al., 2009). Although readers often do not accurately recall the precise predicates of argument claims (especially value claims), those who can are better able to accurately discriminate whether or not a reason supports a claim (Britt, Kurby, Dandotkar, & Wolfe, 2008).

These experiments, however, focused on policy (take a stand on the superiority of a particular behavior, action, solution, or state) and value (take a stand on the superiority of the value, desirability, or morality of a state or action) claims. Scientific writing, in contrast, tends to involve claims that assert factual, causal, or associative relationships. Factual claims tend to be simple reports of data measurements (e.g., "The ratio of carbon dioxide reached 400 parts per million Sunday in readings taken by the two top monitors of greenhouse gases."). Factual claims may be challenged but typically only on technical, methodological grounds or outright error. Of greater interest are causal and associative claims. Scientific studies, and thus the texts that report their results, are frequently concerned with identifying the causes of phenomena or at least associations that suggest possible causal influences. Determining the appropriateness of evidence for scientific, causal claims requires an understanding of the basic scientific reasoning behind distinguishing causality from covariation. An assertion of a causal relationship typically requires evidence either of direct manipulation through a controlled experiment (possibly involving random assignment to treatment and non-treatment conditions) or an accumulation of indirect evidence (e.g., cytological or animal experiments along with epidemiological and longitudinal data, demonstration of a dose–response relationship, etc.). An associative claim requires only evidence of simple covariation. A demonstration of covariation alone does not rule out the possibility of random coincidence or association with another unknown factor that may be the actual cause. Thus, evidence of simple covariation is not sufficient to support a causal claim.

It is unclear how accurately readers represent causal claims. It may be that associative claims, like value claims, are not very accurately recalled and are actually systematically misrecalled. A recent study by Britt, Durik, Steffens, Bloss, and Baker (2012) that presented people with headline–study inconsistencies suggests that this may be the case. In their study, undergraduates read health news stories downloaded from the Web. Each article reported the results of a study, and the data were always clearly correlational. They manipulated whether the title and first sentence was consistent (correlational conclusion) or inconsistent (causal conclusion) with the findings of the reported study. They found a strong tendency for readers to summarize the study's correlational findings as causal when the title was causal compared with when it was consistent with the study (i.e., a correlational headline). Thus, readers appeared to sometimes restructure their representation of the text to resolve the inconsistency between the headline conclusion and the report of the study in the direction of the more definite causal claim.

One limit of the Britt et al. study was that the students may not have had the necessary knowledge to recognize what makes a study correlational. In fact, Vosniadou, Pearson, and Rogers (1988) found that familiarity and prior knowledge were important for detecting inconsistency. If participants are given an article that clearly states the boundary conditions on interpreting the study, they should be more likely to notice the inconsistency. Therefore, in the current studies, the article included an interpretation statement making clear the limits or benefits of the particular design employed.

## Use of Qualifiers in Scientific Writing

Another common feature of scientific writing is the use of hedges (e.g., may be, likely, possible) while often avoiding amplifiers (e.g., proves, absolutely) to communicate recognition of the limits of scientific inference. The nuances of this usage are not always understood by reporters of science and lay readers. Indeed, studies have shown that readers prefer certain conclusions (Bråten & Strømsø, 2010) and perceive various linguistic devices such as hedges, hesitations, and disclaimers as powerless language and find the associated message less persuasive (Blankenship & Holtgraves, 2005; Holtgraves & Lasky, 1999; Hosman, 1989; Hosman, Huebner, & Siltamen, 2002). For instance, Blankenship and Holtgraves (2005) gave students a counter-attitudinal argument with or without hedges. They found the argument with hedges led to less persuasion, more negative perceptions of the author and weaker evaluations of the argument than the version without hedges.

Although such interpretations of hedges and qualifiers hold for untrained readers, there are situational exceptions. Hedges and qualifiers are frequently used in academic writing to accurately capture the probabilistic nature of phenomena and conclusions (Butler, 1990; Horn, 2001; Hyland, 1998; Skelton, 1988) and are evaluated positively in these contexts (Jensen, 2008). There is some evidence that students do not prefer unhedged statements when the hedges are appropriate. In fact, Durik, Britt, Reynolds, and Storey (2008) found that as long as the hedges were professional and qualified interpretative statements, they did not lead to negative perceptions of the policy, source, or argument. This was especially true for readers who had taken science courses. Therefore, knowledgeable readers may be able to accurately evaluate the appropriateness of hedges on scientific argument conclusions.



## Detecting and Resolving Inconsistencies

Are readers sensitive to whether a study appropriately supports the author's claim (e.g., overstatements or understatements of evidence)? When detected, such inconsistencies are in essence a break in coherence, and research suggests readers will make inferences to resolve such breaks in narratives and, under some circumstances, in expository texts.

Researchers have used within-text inconsistencies to better understand how factors affect the establishment of global coherence in text processing. O'Brien and colleagues examined readers processing of short narratives in which an action (e.g., *Mary ordered a cheeseburger and fries)* either is character-consistent (fast food nut) or character-inconsistent (e.g., vegetarian health nut) (Albrecht & O'Brien, 1993; Hakala & O'Brien, 1995; O'Brien & Albrecht, 1992; O'Brien & Myers, 1985). Across several studies, readers took longer to read and had better memory for character-inconsistent action statements than character-consistent ones. Albrecht and O'Brien (1993) argued that readers reactivated the prior inconsistent information in an attempt to resolve the coherence break. This reprocessing and additional coherence-building reasoning results in better memory for the inconsistent texts. In fact, Blanc, Kendeou, van den Broek, and Brouillet (2008) used think-aloud protocols to show that readers actually produced a causal inference in the inconsistent version.

With expository texts, several studies have found that readers often fail to explicitly detect inconsistencies (Barton & Sanford, 1993; Glenberg, Wilkinson, & Epstein, 1982; Markman, 1979; Hannon & Daneman, 2004; Noordman, Vonk, & Kempff, 1992; Otero & Kintsch, 1992), although evidence shows that when they do notice them they attempt to make a coherence-building inference (Singer & Gagnon, 1999; Wiley & Myers, 2003). For example, Wiley and Myers (2003) found longer reading times for inconsistent information in an expository text, but only when multiple premises were presented close together in time. This could indicate that readers noticed the coherence break and generated an inference to repair the inconsistency (the authors did not verify that an inference had been made). Interestingly, when readers do generate a coherence-building inference that resolves a conflict, they may fail to explicitly report the presence of within-text factual inconsistencies (Otero & Campanario, 1990; Otero & Kintsch, 1992). The reader in effect "edits" the text to form a coherent representation.

The difference may lie with the nature of the inconsistency and type of resolution. One way to resolve an inconsistency is a content-based inference. Hakala and O'Brien (1995) argue that readers attempt to resolve global coherence breaks by integrating the inconsistent information but resolve local coherence breaks by modifying or distorting the inconsistent statement. Several other resolution strategies have emerged from research on the influence of inconsistencies on memory. Otero and Kintsch (1992) asked participants to recall texts containing inconsistencies and found several recall patterns that emerged when readers failed to report an inconsistency. These patterns included neglecting to recall either piece of conflicting information, recalling only one piece of conflicting information, or generating an inappropriate inference to resolve the conflict. Similarly, Chinn and Brewer (1993, 1998) proposed a taxonomy of reactions to inconsistent or anomalous information. Such reactions include rejecting the inconsistent information as incorrect, reinterpreting the inconsistent information, or making slight modifications to make the information consistent. However, unlike Otero and Kintsch (1992), the Chinn and Brewer taxonomy assumes the reader has explicitly detected that an inconsistency exists.

An entirely different way a reader might resolve an inconsistency is to preserve it but attribute it to a particular source. In some contexts, readers are thought to explicitly represent sources and relationships between sources and their content (Britt, Perfetti, Sandak, & Rouet, 1999; Perfetti, Rouet, & Britt, 1999; Rouet, 2006; Wineburg, 1991). Braasch, Rouet, Vibert, and Britt (2012) found that when claims conflict, readers appear to use sources to structure their memory representations of the text and resolve the discrepancy. The presence of source information may also affect how people read arguments. Tobin and Raymundo (2009) found that participants were more persuaded by stronger arguments but only when the source was a content expert. Similarly, Winter and Krämer (2012) found that readers preferred (i.e., more frequently selected, read longer, and evaluated more favorably) arguments written by expert authors that presented both sides of the controversy compared with lay authors. Finally, Stadtler, Scharrer, Brummernhenrich, and Bromme (2013) found that when conflicting information was presented in multiple documents written by experts as compared with a single document, readers were more likely to mention the conflict in their essay to a fictitious friend. These studies suggest that source information may function as an additional mechanism when evaluating argumentative texts. None of these studies, however, examined memory for sources.

## Current Study

The current study examines whether readers are sensitive to the truth value of a claim based on the evidence used to support it. The current study extends prior research in two ways. First, it examines readers' memories for argument evidence when presented with claim–evidence inconsistencies. In a set of health articles, we manipulated the definiteness of the claim by manipulating both the type of verb and the type of qualifier. Definite claims included both a qualifier that indicated high certainty (e.g., definitely) and a causal predicate (i.e., is caused by). Tentative claims, in contrast, included a qualifier that indicates some degree of uncertainty (e.g., likely suggests) and an associative predicate (i.e., is linked to). We then paired these claims with evidence that was either consistent or inconsistent to form overstatements, understatements, and their appropriate controls. *Overstatements* paired definite claims with correlational evidence. *Understatements* paired tentative claims with experimental evidence. We regard these two types of inconsistencies as qualitatively different. Overstatements are generally considered to be more egregious and, given readers' and writers' natural preference for strong language, more likely to occur in published articles (Jensen, 2008). Understatements may not typically be regarded so much as errors in reporting than as over-cautiousness or as a way to downplay a result that is not favorable to some interested party. Thus, we expect readers to react differently to each type of inconsistency. In the case of overstatements, readers' preference for strong language may lead them to attempt to preserve the definite conclusion while generating an inference that modifies the results of the study. However, they will not be able to make a content-based coherence-building inference to correct the study and as a result should reject the study as support for the claim made by the author. In the case of understatements, readers could create an inference, based on evaluation of the source, study or explanation provided, to resolve the inconsistency. Therefore, we expect different patterns of processing and recall for each inconsistency type.

The second research goal was to extend prior work on sourcing to within-text inconsistencies. All the health news articles we used have a single author with no embedded references to other authors. However, by constructing the articles as first-person accounts by the researcher who conducted the study, we sought to make source information more salient. This allowed us to examine the extent to which readers represent and use source information from a single source in argument comprehension. For overstatements, readers may make a source-based inference that the author is either not knowledgeable about the limits of a correlational study or presenting a biased account in an attempt to present a more persuasive article. For understatements, readers may examine the source more carefully to understand why the author would try to minimize a causal result. Because our articles present both evidence and a conclusion from the same source and prior work with sourcing has focused on competing sources presenting claims, we do not expect source recall differences to be very large.

In the studies we report, participants were asked to read the articles with the purpose of making a decision about their own behavior. Later, without warning, readers were tested on their memory for the document's content (claim, supporting evidence) and source information (e.g., the author, date). Based on past research, there are two plausible ways to explain what readers do and how it might affect their recall.

If we take an argument-reading perspective, readers will attempt to create a coherent representation of the argument and evaluate the extent to which the claim is an accurate conclusion from the study. As a result, reading an article in which the evidence fits neatly into an argument structure established by the claim ought to lead to better recall of the article relative to one in which the evidence cannot fill such a role. Thus, according to a *Structure Hypothesis*, claim-consistent information should lead to a more coherent structure that will guide retrieval of support for that claim. As mentioned earlier, people appear to structure their representations of arguments hierarchically around a central claim, and recall of argument information shows similar patterns to those observed by other studies of schema-like memory structures. For example, Britt and Larson (2003) found that readers were inclined to recall simple arguments in a claim–reason order even when presented in a reason–claim order, suggesting a restructuring of the information. If the claim acts as a schema-like organizer (Britt & Larson, 2003; Chambliss, 1995), then when the presented evidence is consistent with the constraints established by that structure, it should result in a more coherent and more complete representation. The source itself can be considered a form of support to the extent that the source has the knowledge to conduct such a study. Therefore, in the consistent case, both the source and the details of the study increase the believability of the claim (i.e., are bases of support). In the overstatement condition, because an inference cannot be made to create a coherent structure for the definite causal claim, readers will reject the claim as a reasonable conclusion from the support (i.e., a correlational study). As such, the evidence must be represented as an isolated part of the text and it becomes less accessible at time of recall. Thus, the prediction for the *Structure Hypothesis* is that readers in the *appropriate* evidence conditions will recall evidence and source statements better, will reject the overstated claims, and will not spend longer reading or revisiting prior information.

An alternative explanation and prediction for recall differences might be made by viewing claim–evidence appropriateness as another type of coherence break, like those previously studied with narrative and expository texts (Albrecht & O'Brien, 1993; Hakala & O'Brien, 1995; Singer & Gagnon, 1999; Wiley & Myers, 2003). That research would suggest a *Reprocessing Hypothesis* where detection of an inconsistency causes a reader to revisit the text and attempt to resolve the inconsistency with some type of inference (Albrecht & O'Brien, 1993; Hakala & O'Brien, 1995). The current study's materials afford a coherence-building inference by reprocessing the study details or the source information. This *Reprocessing Hypothesis* would predict that upon reading the inconsistency, readers would engage in additional processing and make coherence-preserving inferences. Thus, the prediction for the *Reprocessing Hypothesis* is that readers in the *inconsistent* evidence conditions will recall evidence and source statements better, will fail to reject the claim as reasonable, and will spend longer reading or revisiting prior information that would help resolve the coherence break.

## EXPERIMENT 1

To test the two hypotheses, participants read short news articles about health topics in which researchers presented the results of their own study. We manipulated the claim type (definite vs. tentative) and type of supporting evidence (experimental vs. correlational). This produced articles that overstated or understated their results along with ones that appropriately stated conclusions for their respective evidence type. For overstated claims, a definite claim was either supported by a correlational study (overstatement version) or an experimental study with random assignment (appropriate version). For understated claims, a tentative claim was either supported by the experimental study (understatement version) or the correlational study (appropriate version). After reading, participants' memories for the claims, sources, and supporting evidence (e.g., key design details) were measured.

### Methods

#### *Participants and Design*

One hundred thirty-two introductory psychology students (mean age $= 19.5$, $SD = 1.69$; 50% women) from a large Midwestern university participated for course credit. Demographic characteristics were 56% white, 29% African-American, 8% Hispanic, 2% Asian-American, and 5% mixed or self-classified as other. No participants had taken a psychology research methods course in college. Two participants did not complete the study so their data were not used. Participants were randomly assigned to one of four conditions: two claim types (definite vs. tentative) $\times$ two support types (experimental vs. correlational).

#### *Materials*

*Health articles.* Three short health articles ($\sim$200 words each) were written about different potential contributors to cancer in humans (cell phone usage, exposure to sun rays, and proximity to power lines). An example article is shown in Table 1. For each article, the title was followed by five elements of source

TABLE 1

Health Article Demonstrating Manipulation of Claim Strength (Definite vs. Tentative) and Type of Support (Causal vs. Correlation)

| *Title:* | *Cell Phones and Cancer* | |
|---|---|---|
| Source: | By Dr. Bradford Franks, Neurosurgeon at the University of Memphis, Tennessee. Published in *Newsweek Magazine*, Week 26, 2011. | |
| Introduction: | The rise in technology has led to greater cell phone use, but that doesn't come without concerns about potential risks associated with using them. When a cell phone is being used, it creates a radiofrequency (RF) current that is used to stay connected with the network. The RF levels vary from phone to phone; many have high levels, while a smaller set of phones have low levels. | |
| Claim statement: | ***Definite Claim*** | ***Tentative Claim*** |
| | My research has **clearly demonstrated** that cell phone use **causes** brain tumors. | My research **likely suggests** that cell phone use **is linked to** brain tumors. |
| Participant assignment: **Causal/correlational support** | In this study, I **randomly assigned 30 people to use** high RF-emitting phones, while 30 were given low RF-emitting phones to use. / In this study I **interviewed 60 patients with brain tumors**. | |
| Interpretation: **Causal/correlational support** | Random assignment is a gold standard for experiments, because **it rules out alternative explanations. /** This procedure **does not allow us to rule out alternative explanations.** | |
| Result statement: | The results showed that 90% of the people in the high RF-emitting phone group reported the presence of brain tumors after 5 years, while only 20% in the low RF-emitting phone group reported brain tumors. / The results showed that 90% of these cancer patients had been using high RF-emitting cell phones for the past 5 years. | |
| Claim statement: | ***Definite Claim*** | ***Tentative Claim*** |
| | **Clearly** my research **demonstrates** that cell phones cause brain tumors. | My research **likely suggests** that cell phone use **is linked to** brain tumors. |

*Note*. Only one version of each claim statement, participant assignment, and interpretation statement was presented to the participants depending on condition. Labels and bolded words in the text were not included in the versions presented to participants.

information. All sources were described as experts because prior research has found that argument evaluation was greatest when sources were experts (Stadtler et al., 2013; Tobin & Raymundo, 2009; Winter & Krämer, 2012). An eight- to nine-sentence article with a standard structure followed. Three to four sentences introduced the topic and were held constant across versions. The essential part of the article involved a first-person presentation of the author's study and main claim. The claim was either "definite" (causal with amplifiers) or "tentative" (correlational with hedges). Type of support was manipulated in the next two sentences. The first sentence reported on participant assignment (causal–random assignment or correlational–preexisting groups). The second sentence included an interpretation statement of the method of assigning participants. Finally, a simple statement of the results of the study was then presented followed by a restatement of the claim.

*Topic beliefs and demographic survey.* Participants rated their agreement with six initial health belief statements related to the documents (e.g., "Sun rays may cause skin cancer") on a 10-point scale (1 = strongly agree, 10 = strongly disagree). A demographic survey assessed age, gender, ethnic heritage, and number of science classes.

### Measures and Scoring

*Source recall.* For each article, participants were provided with the title and asked to recall five distinct source elements corresponding to the cues (name, occupation, affiliation, magazine, year). Responses were coded as correct if they matched or were close to the correct answers. For example, "neurologist" was considered acceptably close to "neurosurgeon" and scored as a correct response, whereas "doctor" was not. A sample of 20% of the source recall statements were independently scored, blind to condition, yielding adequate reliability ($\kappa = 0.90$). The number of sources recalled was calculated by counting each distinct article for which at least one source element was recalled (out of three).

*Claim and support recall.* Participants were given the title of the article and asked to recall the main claim ("What was the author's main conclusion?") and support ("Give as many details as you can about the study described in the article"). Claims were scored as tentative if they included a verbatim mention or synonym of a probable hedge (e.g., possibly, may) *or* a tentative predicate (is linked to). Claims were scored as definite if they used a verbatim mention or synonym of the definite predicate (e.g., causes, leads to). Support was manipulated across two idea units (see the participant assignment and interpretation statement in Table 1). A response was scored as experimental support if the participant mentioned random assignment *or* the strength of the interpretive statement (e.g., gold standard, rules out alternative explanations) and scored as correlational support if the participant mentioned anything noncausal (e.g., interviewed, surveyed) *or* about limits on interpretation (e.g., does not allow ruling out alternative explanations). All other responses were scored as other including other details of the study (e.g., number of participants), a statement of the results, or an explanation of the mechanism from the introduction. All scoring was completed blind to condition, but correct recall was later computed dependent on condition (e.g., a tentative claim recall was only accurate if the reader was in a tentative claim condition). A random sample of 20% of the recalls was scored by two independent raters and interrater reliability for those recall protocols was high ($\kappa = 0.94$).

### *Procedure*

In one 60-minute session, participants completed the topic beliefs and demographic surveys, read the health articles at their own pace, performed a 10-minute filler task, and finally completed the source recall and the claims and support recall for each article. The reading instructions for the health articles asked participants to "read carefully as if you were trying to make a decision about your own behavior. After you are done reading, you will be asked questions about what you read. So it is important that you try to remember what you read in these articles." Thus, the source recall task was a surprise task and the claims and supporting recall was less of a surprise.

## Results

Self-reported high school and college science classes and average topic beliefs did not reliably differ across groups (all $F < 1$, *ns*). Claim types (definite vs. tentative) × 2 consistency (appropriate vs. inconsistent) between-participant analyses of variance (ANOVAs) were conducted on the recall of claims, sources, and supporting evidence (alpha at .05 for all tests). The means and standard deviations for each condition are shown in Table 2. In some cases, there was a slight violation of normality; however, the sample size was large and the distributions were skewed in the same direction so no correction was needed. In analyses where there was heterogeneity of variance, Welch-Satterthwaite adjustment to degrees of freedom was used because the sample sizes were approximately equal and the results were the same when outliers were trimmed.

The claim type × consistency ANOVA on the average number of claims recalled found no significant differences across conditions, although claim type was marginally significant, $F(1, 127) = 3.142$, $MSE = 1.241$, $p = .079$, with a trend toward better recall for definite claims. The claim type × consistency ANOVA on average number of sources recalled revealed a significant effect of

TABLE 2
Average Number of Correctly Recalled Claims, Sources, and Supporting Evidence (of Three) as a Function of Experimental Condition for Experiments 1 and 2

| | *Definite Claim* | | *Tentative Claim* | |
|---|---|---|---|---|
| *Average Number of* | *Appropriate* | *Overstatement* | *Appropriate* | *Understatement* |
| | *Experiment 1* | | | |
| Claims recalled | 1.81 (1.25) | 1.82 (1.00) | 1.67 (1.16) | 1.27 (1.04) |
| Sources recalled | 1.06 (1.09) | 0.50 (0.71) | 1.15 (1.10) | 0.64 (0.93) |
| Supporting evidence recalled | 1.52 (1.23) | 0.91 (0.97) | 1.09 (1.10) | 1.30 (1.13) |
| | *Experiment 2* | | | |
| Claims recalled | 1.52 (0.99) | 1.45 (1.12) | 1.86 (1.11) | 1.38 (1.08) |
| Sources recalled | 1.72 (1.13) | 1.79 (0.98) | 1.50 (1.20) | 1.24 (1.12) |
| Supporting evidence recalled | 1.90 (1.23) | 1.21 (1.11) | 1.11 (1.17) | 1.28 (1.13) |

*Note*. Recall for each article element is out of three, and standard deviations are in parentheses.

consistency, $F(1,\ 128) = 10.136$, $MSE = .940$, $\eta^2 = .043$. Participants recalled sources better when the support was appropriate than when the support was inappropriate. No other effects were significant. The claim type × consistency ANOVA on supporting evidence recall revealed only a significant interaction, $F(1,\ 127) = 4.434$, $MSE = 1.230$, $\eta^2 = .016$. The follow-up $t$ tests found that only for definite claims did consistency of support matter, $t(56.726) = -2.184$, $d = .545$. Participants given supporting evidence that was claim appropriate recalled the target study details significantly more often than when the supporting evidence was claim inconsistent (i.e., an overstatement). The effect was not significant for tentative claims (i.e., understatements).

## Discussion

The results show that readers are sensitive to whether a study appropriately supports the strength of an author's claim. The results support the *Structure Hypothesis*. For definite claims, an overstatement led to worse recall of both sources and supporting evidence. The effect was not as strong for the more subtle tentative claims. For tentative claims, an understatement led only to worse recall of the sources. These results are consistent with other findings that the claim acts as an organizer for argument information (Britt & Larson, 2003; Chambliss, 1995).

Prior work with the inconsistency paradigm has shown better memory for inconsistent information due to the extra processing that is required by the need for a coherence-building inference (Albrecht & O'Brien, 1993; Hakala & O'Brien, 1995). That was not the case here. The lack of support for the *Reprocessing Hypothesis* could have been because the inconsistency was not clearly resolvable. In the Albrecht and O'Brien (1993) materials, the reader could make an inference even though the situation was also not absolutely resolvable. However, readers may be more promiscuous with coherence-building inferences in narratives than when reading arguments about research in science and health.

Because Experiment 1 looked only at recall, it is not possible to verify that participants' reading patterns were consistent with those of noticing and addressing a coherence break. Therefore, in Experiment 2 we examined whether readers encounter a coherence break with a claim–evidence inconsistency and whether they do indeed reprocess prior text segments. We also examine whether readers explicitly acknowledge the targeted design problem with the argument in the inconsistent conditions. These additional measures are intended to provide confirmation of detection of a problem in the argument and an attempt to resolve the problem.

## EXPERIMENT 2

The aim of Experiment 2 was to examine reading behaviors and explicit identification of the target problem as well as to replicate the results from Experiment 1. The same articles from Experiment 1 were presented to participants using the *Read&Answer* software tool to allow tracking of which segments students read and in what order. If readers attempt to resolve claim–support inconsistencies, then we would expect more disruptive reading patterns in the inconsistent arguments compared with the consistent arguments (i.e., rereading and longer reading times). A converging measure of readers' perceiving the claim–support inconsistencies as a coherence break would be the rate of rejecting the conclusion as reasonable on the basis of the provided study. In Experiment 2 we directly asked participants whether the conclusion the author made was reasonable and then to explain their answer. We expect that with the claim–support inconsistencies, participants would more frequently reject that the conclusion was reasonable and cite the target design problem as the reason. Based on Experiment 1, we expect that the detection of the inconsistency would be strong in the overstatement condition.

### Methods

#### *Participants and Design*

One hundred fifteen introductory psychology students (mean age = 19.76, *SD* = 2.65; 65% women) participated for course credit. None of these students participated in Experiment 1. Demographic characteristics were 57% white, 26% African-American, 4% Hispanic, 6% Asian-American, and 7% self-classified as other. Participants were randomly assigned to one of four conditions: two claim types (definite vs. tentative) × two support types (experimental vs. correlational). As in Experiment 1, an analysis of the type of support for the definite claims was conducted to show the effect of overstating results, whereas an analysis of the type of support for the tentative claims was conducted to show the effect of understating results.

### Materials

*Health articles.* The articles were presented on a computer using *Read&Answer* (Vidal-Abarca et al., 2011). A segment was masked with unreadable characters until the student clicked on it to display it. At that point, the previously readable item was masked again. Therefore, only one text segment was readable at any time. The text segments were the title, each source element (name, occupation, affiliation, magazine, year), and each sentence of the article for half of the participants. For the other half of the participants, all source elements were presented as a single region to more closely approximate Experiment 1. All articles from Experiment 1 were slightly modified to equate the number of syllables per sentence per region across conditions.

*Reasonable conclusion question.* To examine whether participants noticed a problem with the authors claim–support consistency, they were given the title of each article and explicitly asked "was the researcher's conclusion reasonable given the results of the study they presented (circle one)?" and were asked to explain their answer.

### Procedure

Participants completed the topic beliefs and demographic surveys and then were trained on the use of *Read&Answer*. Then, participants were given the same reading instructions and read the health articles, performed a filler task, and completed both the source and the claim and support recall for each article. Finally, they were given the reasonable conclusion question sheet.

## Results and Discussion

A preliminary analysis showed the presentation format of the source information (five vs. one segment) did not significantly interact in any of the analyses, so this factor was collapsed and the analyses conducted with a set of 2 claim type (definite vs. tentative) × 2 consistency (appropriate vs. inconsistent) between-participant ANOVAs. The means and standard deviations for the analyses can be found in Table 2.

The claim type × consistency ANOVA on the average number of claims recalled found no significant differences across conditions. In contrast to

Experiment 1, the claim type × consistency ANOVA on average number of sources recalled revealed no significant effect across conditions, although the effect of claim type was marginally significant ($F(1, 127) = 3.508$, $MSE = 1.233$, $p = .064$) with a trend toward better recall of sources for definite claims than tentative claims. The findings of Experiment 1 may not have been replicated because of the increased salience of the source. The presentation tool required readers to click to reveal text elements, so it may have been more difficult to skim or skip the source information compared with Experiment 1's paper format. In fact, an examination of the means across experiments shows much higher recall of sources in Experiment 2 (52%) than in Experiment 1 (28%).

Replicating the results from Experiment 1, the claim type × consistency ANOVA on supporting evidence recall revealed a significant interaction, $F(1, 111) = 3.920$, $MSE = 1.351$, $\eta^2 = .014$. Follow-up $t$ tests found that only for definite claims did consistency of support matter, $t(56) = -2.23$, $d = .586$. Participants given supporting evidence that was claim appropriate recalled the target study details significantly more often than when the supporting evidence was claim inconsistent (i.e., an overstatement). The effect was not significant for tentative claims (i.e., understatements). No other effects were significant. The rate of recall of support was similar across the two experiments.

To examine reading patterns, we conducted claim type × consistency ANOVAs on the average number of revisits to each region for each article and the average time spent on that region. The means and standard deviations are presented in Table 3. There were no significant effects of revisits to the claim or

TABLE 3
Average Number of Total Revisits to a Region and Total Time per Syllable per Article for Experiment 2

| | *Definite Claim* | | *Tentative Claim* | |
|---|---|---|---|---|
| | *Appropriate* | *Overstatement* | *Appropriate* | *Understatement* |
| | | *Total Revisits* | | |
| Source | 0.24 (0.44) | 0.39 (0.79) | 0.23 (0.46) | 0.18 (0.30) |
| Introduction | 0.27 (0.47) | 0.38 (0.56) | 0.19 (0.36) | 0.42 (0.67) |
| Claim | 0.03 (0.19) | 0.00 (0.00) | 0.04 (0.19) | 0.03 (0.19) |
| Study | 0.19 (0.43) | 0.29 (0.31) | 0.06 (0.16) | 0.29 (0.46) |
| | | *Total Time* | | |
| Source | 172.83 (84.81) | 165.39 (78.97) | 162.16 (100.77) | 166.18 (104.50) |
| Introduction | 205.31 (38.84) | 199.08 (62.82) | 200.04 (51.18) | 246.73 (82.65) |
| Claim | 274.95 (91.20) | 260.19 (95.02) | 278.25 (94.68) | 319.25 (103.21) |
| Study | 196.88 (49.31) | 195.98 (49.39) | 189.45 (54.83) | 241.64 (105.83) |

*Note*. Standard deviations are in parentheses.

source (all $p > .05$). For revisits to the introductory region, there was only a marginally significant effect of consistency, with more revisits in the inconsistent condition ($p = .083$). For revisits to the study region, there was a significant effect of consistency, $F(1, 111) = 5.805$, $MSE = .131$, $\eta^2 = .037$. Although the interaction between claim type and consistency was not significant, we examined the effect of type of evidence for each type of claim separately. Only for tentative claims did readers revisit the target study information more often when the supporting evidence was inconsistent than when it was consistent, $t(34.55) = 2.518$, $d = .667$. Therefore, the main effect of consistency on revisits to the study was driven by differences when given an understatement, not with an overstatement.

There were no effects for total reading time on the claim or source regions (all $p > .05$). For total reading time on the introduction region, there were marginal main effects ($p = .066$ for claim type and $p = .079$ for consistent) that were qualified by a significant interaction, $F(1, 111) = 5.387$, $MSE = 3{,}736.308$, $\eta^2 = .004$. Only when the claim was tentative did the inconsistency (i.e., understatement) lead to longer reading times on the introductory region, $t(46.975) = -2.574$, $d = .679$. Finally, for the total reading times on the critical study information, we found a significant effect of consistency, $F(1, 111) = 3.951$, $MSE = 20{,}254.596$, $\eta^2 = .003$, and a significant interaction, $F(1, 111) = 4.233$, $MSE = 20{,}254.596$, $\eta^2 = .004$. Only for tentative claims did readers spend longer reading the inconsistent study information compared with the appropriate control, $t(42.34) = -2.349$, $d = .619$.

To determine whether readers differentially detected the inconsistency, we classified the open-ended responses to the reasonable conclusion question as to those mentioning *Targeted* information (i.e., the manipulated parts of the claim or study), *Other Study* information (i.e., statements or evaluations of nonmanipulated study elements or findings), and *Beliefs* (i.e., opinions or prior knowledge). Examples are shown in Table 4. The means and standard deviations for each response type are presented in Table 5. To examine detection of a problem with the conclusion based on the important design issue or the claim statement, a claim type × consistency ANOVA was conducted on mentions of *Targeted* information when deciding that the conclusion was not reasonable. The analysis revealed a main effect of claim type, $F(1, 111) = 4.649$, $MSE = .619$, $\eta^2 = .02$, and consistency, $F(1, 111) = 5.762$, $MSE = .619$, $\eta^2 = .027$. These effects were qualified by a significant interaction, $F(1, 111) = 17.539$, $MSE = .619$, $\eta^2 = .158$. Simple effects revealed that the effect of consistency held only for definite claims. Participants were more likely to decide the conclusion was unreasonable based on the manipulated study elements when the supporting evidence was inconsistent than when the supporting evidence was consistent, $t(51.919) = 2.94$, $d = .772$. These results indicate that participants were able to explicitly notice the inconsistency when the claim strongly states a

TABLE 4
Example Student Responses to the Reasonable Conclusion Question for Experiment 2

| *Type of Information Mentioned* | *Example Student Responses* |
|---|---|
| Targeted info | "Just because there is a correlation does not mean it is a causation [sic]. No control group." |
| | "I said no because the researcher wasn't sure of his conclusion or the whole study for that matter because he kept using the word 'probably.'" |
| | "He did not use random assignment." |
| | "He said himself that the cancer some of the residents had could've come from other things. Plus he didn't do the whole population to compare." |
| Other study info | "The group of people was too small and there wasn't a specific time or age of the people mentioned." |
| | "The experiment showed that the less SPF you have on, the more likely you are to have cancer cells." |
| Beliefs | "Use cell phone all the time. don't [sic] see no problem at all with me." |
| | "I believe that electricity can effect [sic] our health." |
| | "Because the better sunscreen you have, the less risk you are at. I hear it everyday from everywhere." |

conclusion. In contrast, participants given an understatement were not more likely to say the conclusion was unreasonable based on the manipulated study elements than when the supporting evidence was consistent, $t(55) = 1.52$, $p > .05$.

When there was a clear overstatement, readers recalled the supporting evidence less well than when the claim and evidence was appropriate. This poorer recall was likely due to detecting a problem (as indicated by the rejection

TABLE 5
Average Number of Responses to the Reasonable Conclusion Question as a Function of Experimental Condition for Experiment 2

| | *Definite Claim* | | *Tentative Claim* | |
|---|---|---|---|---|
| *Mentioned* | *Appropriate* | *Overstatement* | *Appropriate* | *Understatement* |
| | | *Decide Conclusion Not Reasonable* | | |
| Targeted info | 0.48 (0.87) | 1.28 (1.16) | 0.61 (0.69) | 0.34 (0.61) |
| Other study info | 0.55 (0.87) | 0.41 (0.78) | 0.89 (0.92) | 0.69 (0.97) |
| Beliefs | 0.66 (0.97) | 0.31 (0.76) | 0.25 (0.52) | 0.24 (0.58) |
| | | *Decide Conclusion Reasonable* | | |
| Targeted info | 0.31 (0.66) | 0.10 (0.41) | 0.25 (0.52) | 0.48 (0.87) |
| Other study info | 0.55 (0.78) | 0.62 (0.90) | 0.57 (0.69) | 0.86 (0.95) |
| Beliefs | 0.44 (0.69) | 0.28 (0.65) | 0.43 (0.69) | 0.38 (0.73) |

*Note*. Each type of explanation is out of three, and standard deviations are in parentheses.

of the conclusion by mentioning the target study information) but not attempting to resolve the problem (as indicated by lack of revisits or longer reading times of specific regions). Thus, recall of the evidence is low with an overstatement because no additional reprocessing is taking place.

For understatements, participants not only noticed that something was wrong but returned to reexamine the content. By reprocessing the study information rather than outright rejecting the claim, it appears that the reprocessing did not lead to a successful inference for resolving the inconsistency. Otherwise, readers in this understatement condition would have had better recalled the inconsistency compared with the control condition.



## GENERAL DISCUSSION

Two experiments showed that overstated claims (definite causal claim backed by a correlational study) and understated claims (tentative associative claim backed by an experimental study) influence readers' comprehension of health articles. During argument comprehension, readers evaluate the truth of claims based on the evidentiary support provided.

For arguments with definite claims, readers recalled the sources (Experiment 1 only) and supporting evidence better when the claim was an appropriate interpretation of the study rather than an overstatement. Experiment 2 results help us better understand the cause of this memory disadvantage with an inconsistency. Readers in the overstatement condition were more likely to reject that the claim was reasonable based on limits of the study but did not systematically reprocess the prior text to resolve the coherence break. As predicted by the *Structure Hypothesis*, the consistent claim acted as a schema to guide recall of support. These results show that even untrained college students are creating a coherent representation of the argument presented by these types of health articles and have some ability to evaluate the quality of research statements. Although readers in the overstatement condition noticed the evidence was not appropriate, they failed to integrate it into their memory structure for the text. Because the evidence was represented as an isolated part of the text, it was less accessible at time of recall. Thus, these results do not support the *Reprocessing Hypothesis*. Readers did not recall, revisit, or read longer the inconsistent information in the overstatement condition. Readers did not attempt to work to resolve the problem and generally rejected the claim–evidence relationship. Thus, when the type of evidence is made clear, readers appeared to notice the inconsistency of making a definite causal claim from a correlational study and this influenced memory but not reading behavior.

For arguments with tentative claims, the effect of the study–claim inconsistency on memory was weak. Source information was better recalled when the study was consistent with the claim than when the claim understated it but only when the sources were not otherwise salient. This better recall of source information provides only very limited support for the *Structure Hypothesis*. The results also do not provide support for the essence of the *Reprocessing Hypothesis*. Although the understatement inconsistency seemed to be perceived as a resolvable coherence break, in terms of processing and accepting the claim, this reprocessing did not result in a better memory for the inconsistent information. Thus, when a researcher makes a hedged, associative claim based on an experimental study, readers did not appear to regard this as a problem and thus it did not affect reading behavior or recall (except for source recall in Experiment 1).

The finding that readers did not reject the conclusion in the understatement arguments is consistent with prior research that suggests that scientific conclusions containing a probabilistic hedge are not always considered weak or inappropriate (Durik et al., 2008). Reports of scientific research frequently begin with conclusions that are tentative, and only through replication and further research should the conclusions become more definite. Indeed, it is not necessarily unreasonable when given a single experimental study to draw a tentative conclusion. However, concluding an associative relationship from an experimental study may be more problematic if the presenter is biased. For example, researchers paid by some industries may understate results that are not favorable to their employers to make them seem less conclusive.

The current findings conflict with prior research with narratives and nonargument, expository texts. Several differences may have led to this outcome. Readers may have simply been unable to resolve the inconsistency because they lacked the ability or knowledge necessary to correct the conflict (Wiley & Myers, 2003). When evaluating arguments, the reader must understand the type of supporting evidence presented and evaluate whether the evidence supports the type of conclusion being drawn. Our interpretation statement may have helped readers identify whether or not the claim was appropriate given the evidence. However, to repair a claim–evidence inconsistency, readers must know how to modify the strength of the claim (e.g., realize "is linked to" would be more appropriate than "causes") or the type of evidence (e.g., awareness of design limits) to make the claim and support consistent. If readers lacked this knowledge, they would not be able to generate a corrective inference to resolve the inconsistency. As a result, readers could not generate a coherent representation. From the result that readers recalled the claim but did not as frequently recall the details of the study, it appears they chose to maintain coherence by rejecting the study.

Another difference with other studies was the presence of source information. This might have provided readers with another potential way to resolve the inconsistency: discrediting the source. Readers could have resolved the inconsistency through a negative source attribution (e.g., "this guy doesn't know what he's talking about"). In this respect, however, our articles were not very helpful because the sources were experts, publishing in an outlet that was not clearly biased. Based on these constraints, readers may have encountered difficulty using the source when trying to resolve the coherence break. We found that source recall was low for Experiment 1, where readers were allowed to skip or skim the source information. Under these naturalistic conditions, readers recalled only 28% of the three sources (or less than one per person on average). This level of recall was lower than recall for the evidence (41% of the three articles). When readers needed to click to reveal information, recall of source information rose to about half of the sources (52%), which was comparable with recall of evidence (46%).

Although the articles purported to be researchers' accounts of their own study, this did not lead to a highly salient and recallable source. Contrary to what one might expect from prior research with multiple sources (Braasch et al., 2012), we did not find evidence that readers recalled the source better when there was an inconsistency. For instance, Stadtler et al. (2013) found that multiple sources led to better memory for the conflict than single sources. The current results may suggest that lay readers do not expect to find inconsistencies within a source and as such do not make inferences about the source as unknowledgeable or biased. In the only other study we know of that investigated memory for sources with single source inconsistencies (de Pereyra, Rouet, Braasch, Le Bigot, & Britt, 2012), there was likewise no effect. It is unclear to what extent nonexperts will spontaneously attend to source information without the presence of multiple sources. The results of the current studies, however, do not provide an optimistic picture of attention to source information for inconsistencies within a single source.

The present study attempts to bridge several areas of research. One direction for future research could vary the quality (e.g., professional vs. lay authors) or the perceived bias (e.g., scientist, beverage company CEO) of a single source to determine which features lead readers use to successfully generate a corrective inference to explain the claim–evidence inconsistency. A second potential direction would be to examine how readers comprehend and resolve inconsistencies within texts that contain embedded sources. Editors may modify headline and initial paragraph to make a story more eye-catching to readers (Jensen, 2008). As a result, the appropriate conclusion being drawn by a cited scientist in the article may conflict with the inappropriate conclusion being drawn by the editor. Both of these future directions may identify conditions where the *Reprocessing Hypothesis* could receive support. If the source allows one to create such an inference, the reader may reread the source and this coherence-building inference could lead to better memory for the source connected to the claim.

The present study's main implication for readers of health studies on the Web is that overstating and understating results do not really affect the recall of the main claim of the author. Overstatement does, however, affect the level of recall for the critical information that could potentially be used to detect the problem (i.e., details of the study design) or resolve the inconsistency (i.e., source information). This is problematic for two reasons. First, students generally took away from the article an inappropriate conclusion. Second, because the source and evidence was not recalled as well in the inconsistent condition, it would not be available later for evaluating the accuracy of the conclusion or communicating this conclusion to others. Furthermore, they may act on these incorrect claims. Given that journalists and research press releases sometimes misrepresent study results and limits (Jensen, 2008; Yavchitz et al., 2012), it is important that readers read carefully to detect and resolve inconsistencies. The burden is most problematic for the lay reader who may not be sufficiently critical of science information in general (Phillips & Norris, 1999; Zimmerman, Bisanz, Bisanz, Klein, & Klein, 2001). In fact, in the present studies, a very clear design–interpretation statement was made. One would expect that readers' detection of the inconsistency would have been much lower if this statement had not been provided. This leaves open the question of whether removing this interpretative statement about the overstatements might have led the reader to better recall the details of the correlational study but to also conclude that a valid claim–support relationship exists. In our Web-based world, it is more important than ever to begin to better understand how readers represent and evaluate science and health articles on the Web, especially in higher stakes situations where the actual health of the information consumer might be on the line.

## ACKNOWLEDGMENTS

The authors thank Karina Pérez for her assistance with the analysis of the reading patterns in Experiment 2.

## FUNDING

This research was supported by a Norwegian Research Council grant to Ivar Bråten and Helge Strømsø, an Institute for Educational Sciences grant (No. R305F100007) to the University of Illinois at Chicago, and by the Center for the Interdisciplinary Study of Language & Literacy at Northern Illinois University. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not reflect the views of the sponsoring organizations.

## REFERENCES

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

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*, 477–487.

Blair, J. A., & Johnson, R. H. (1987). Argumentation as dialectical. *Argumentation*, *1*, 41–56.

Blanc, N., Kendeou, P., van den Broek, P. W., & Brouillet, D. (2008). Updating situation models during reading of news reports: Evidence from empirical data and simulations. *Discourse Processes*, *45*, 103–121.

Blankenship, K. L., & Holtgraves, T. (2005). The role of different markers of linguistic powerlessness in persuasion. *Journal of Language and Social Psychology*, *24*, 3–24.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220.

Braasch, J. L. G., Rouet, J. F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition*, *40*, 450–465.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726.

Bråten, I., & Strømsø, H. I. (2010). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes*, *47*, 1–31.

Bråten, I., Strømsø, H. I., & Samuelstuen, M. S. (2008). Are sophisticated students always better? The role of topic-specific personal epistemology in the understanding of multiple expository texts. *Contemporary Educational Psychology*, *33*, 814–840.

Britt, M. A., Durik, A. M., Steffens, B., Bloss, K., & Baker, J. (2012, April). *Students' interpretation of uncertainty in health news*. Presented at the annual meeting of the American Educational Research Association conference "Public Understanding of Science: The Educational Challenges of Scientific 'Uncertainty,'" Vancouver, British Columbia, Canada.

Britt, M. A., Kurby, C. A., Dandotkar, S., & Wolfe, C. R. (2008). I agreed with what? Memory for simple argument claims. *Discourse Processes*, *45*, 52–84.

Britt, M. A., & Larson, A. (2003). Construction of argument representations during on-line reading. *Journal of Memory and Language*, *48*, 749–810.

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J. F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209–233). Mahwah, NJ: Erlbaum.

Butler, C. S. (1990). Qualifications in science: Modal meanings in scientific texts. In W. Nash (Ed.), *The writing scholar* (pp. 137–170). Newbury Park, CA: Sage.

Chambliss, M. J. (1995). Text cues and strategies successful readers use to construct the gist of lengthy written arguments. *Reading Research Quarterly*, *30*, 778–807.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*, 1–49.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, *35*, 623–654.

Dee-Lucas, D., & Larkin, J. H. (1988). Attentional strategies for studying scientific texts. *Memory & Cognition*, *16*, 469–479.

De Pereyra, G., Rouet, J. F., Braasch, J. L., Le Bigot, L., & Britt, M. A. (2012, July). *Readers' memory for information sources in simple news stories: Effects of text and task features*. Presented at the 22nd meeting of the Society for Text and Discourse, Montreal, Canada.

Dixon, P., Faries, J., & Gabrys, G. (1988). The role of explicit action statements in understanding and using written directions. *Journal of Memory and Language*, *27*, 649–667.

Durik, A. M., Britt, M. A., Reynolds, R., & Storey, J. K. (2008). The effects of hedges in persuasive arguments: A nuanced analysis of language. *Journal of Language and Social Psychology*, *27*, 217–234.

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

Hakala, C. M., & O'Brien, E. J. (1995). Strategies for resolving coherence breaks in reading. *Discourse Processes*, *20*, 167–185.

Hannon, B., & Daneman, M. (2004). Shallow semantic processing of text: An individual-differences account. *Discourse Processes*, *37*, 187–204.

Holtgraves, T. M., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology*, *17*, 506–516.

Horn, K. (2001). The consequences of citing hedged statements in scientific research articles. *BioScience*, *51*, 1086–1093.

Hosman, L. A. (1989). The evaluative consequences of hedges, hesitations, and intensifiers: Powerful and powerless speech styles. *Human Communication Research*, *15*, 383–486.

Hosman, L. A., Huebner, T. M., & Siltamen, S. A. (2002). The impact of power of speech style, argument, strength, and need for cognition on impression formation, cognitive responses, and persuasion. *Journal of Language and Social Psychology*, *21*, 361–379.

Hyland, K. (1998). *Hedging in scientific research articles*. Philadelphia, PA: John Benjamin.

Jensen, J. D. (2008). Scientific uncertainty in news coverage of cancer research: Effects of hedging on scientists' and journalists' credibility. *Human Communication Research*, *34*, 347–369.

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education*, *77*, 339–366.

Maier, J., & Richter, T. (2013). Text belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction*, *31*, 151–175.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, *50*, 643–655.

Meyer, B. J., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, *16*, 72–103.

Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, *21*, 121–143.

Noordman, L. G. M., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, *31*, 573–590.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 777–784.

O'Brien, E. J., & Myers, J. L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 12–21.

Otero, J., & Campanario, J. M. (1990). Comprehension, evaluation and regulation in learning from science texts. *Journal of Research in Science Teaching*, *27*, 447–460.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what they read. *Psychological Science*, *3*, 229–235.

Perfetti, C. A., Rouet, J. F., & Britt, M. A. (1999). Towards a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Erlbaum.

Phillips, L. M., & Norris, S. P. (1999). Interpreting popular reports of science: What happens when the reader's world meets the world on paper? *International Journal of Science Education*, *21*, 317–327.

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

Rouet, J. F. (2006). *The skills of document use: From text comprehension to Web-based learning*. Mahwah, NJ: Erlbaum.

Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language*, *59*, 237–255.

Shaw, V. F. (1996). The cognitive processes in informal reasoning. *Thinking & Reasoning*, *2*, 51–80.

Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, *54*, 574–591.

Singer, M., & Gagnon, N. (1999). Detecting causal inconsistencies in scientific text. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 179–194). Mahwah, NJ: Erlbaum.

Skelton, J. (1988). Comments in academic articles. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 98–108). London, UK: CILT/BAAL.

Stadtler, M., Scharrer, L., Brummernhenrich, B., & Bromme, R. (2013). Dealing with uncertainty: Readers' memory for and use of conflicting information from science texts as function of presentation format and source expertise. *Cognition and Instruction*, *31*, 130–150.

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex*, *19*, 1198–1207.

Stein, N. L., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing: Volume 2. Advances in discourse processing* (pp. 53–120). Norwood, NJ: Ablex.

Tobin, S. J., & Raymundo, M. M. (2009). Persuasion by causal arguments: The motivating role of perceived causal expertise. *Social Cognition*, *27*, 105–127.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Vidal-Abarca, E., Martinez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., … Ferris, R. (2011). Recording online processes in task-oriented reading with Read&Answer. *Behavior Research Methods*, *43*, 179–192.

Vosniadou, S., Pearson, P. D., & Rogers, T. (1988). What causes children's failures to detect inconsistencies in text? Representation versus comparison difficulties. *Journal of Educational Psychology*, *80*, 27–39.

Voss, J. F., Fincher-Kiefer, R., Wiley, J., & Silfies, L. N. (1993). On the processing of arguments. *Argumentation*, *7*, 165–181.

Voss, J. F., & Means, M. L. (1991). Learning to reason via instruction in argumentation. *Learning and Instruction*, *1*, 337–350.

Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, *36*, 109–129.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*, 73–87.

Winter, S., & Krämer, N. C. (2012). Selecting science information in web 2.0: How source cues, message sidedness, and need for cognition influence users' exposure to blog posts. *Journal of Computer-Mediated Communication*, *18*, 80–96.

Yavchitz, A., Boutron, I., Bafeta, A., Marroun, I., Charles, P., Mantz, J., & Ravaud, P. (2012). Misrepresentation of randomized controlled trials in press releases and news coverage: A cohort study. *PLoS Medicine*, *9*, e1001308.

YGoY. (2011, February 8). School lunches cause obesity. *YGoY Health Science Updates*. Retrieved from http://ygoy.com

Zimmerman, C., Bisanz, G. L., Bisanz, J., Klein, J. S., & Klein, P. (2001). Science at the supermarket: A comparison of what appears in the popular press, experts' advice to readers, and what students want to know. *Public Understanding of Science*, *10*, 37–58.