

# Reducing reliance on inaccurate information

David N. Rapp · Scott R. Hinze · Kristine Kohlhepp ·  
Rachel A. Ryskin

Published online: 13 June 2013  
© Psychonomic Society, Inc. 2013

**Abstract** People learn from the texts that they read, but sometimes what they read is wrong. Previous research has demonstrated that individuals encode even obvious inaccuracies, at times relying on the misinformation to complete postreading tasks. In the present study, we investigated whether the influence of inaccurate information might be reduced by encouraging the retrieval of accurate knowledge. Participants read an extended text that contained both accurate and inaccurate assertions, after which they evaluated the validity of statements associated with those assertions. In general, participants made more mistakes in their evaluations of statements after having read inaccurate as compared to accurate assertions, offering evidence of the influence of misinformation. However, when participants were tasked with correcting inaccuracies during reading, their mistakes were substantially reduced. Encouraging the retrieval of accurate knowledge during reading can reduce the influence of misinformation. These findings are discussed with respect to the contributions of episodic traces and prior knowledge on learning, as well as to the conditions that support successful comprehension.

**Keywords** Learning · Memory · Prior knowledge · Reading comprehension · Persuasion · Text processing

People learn about the world from what they read. They encode and rely upon the information presented in fictional and nonfictional sources, applying the acquired knowledge

to solve problems, make decisions, build opinions, and motivate future activity (Britton & Black, 1985; Gerrig, 1993; Graesser, Singer, & Trabasso, 1994; Johnson-Laird, 1983; Kintsch, 1998; van Dijk & Kintsch, 1983). This is a good thing when texts provide accurate information that is based on meticulously conducted research, rigorously developed arguments, and carefully constructed prose. However, texts often contain inaccuracies, both intentional and unintentional, which people can also rely upon from their readings (e.g., Gerrig & Prentice, 1991; Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993; Marsh, Meade, & Roediger, 2003). Consider the following illustrations of this reliance: When participants are asked to read fictional stories that contain *valid* declarative facts (e.g., “The capital of Russia is Moscow” or “Bannister ran the first sub-4-minute mile”), their accuracy on postreading tests of those facts improves relative to when they read versions of the stories that leave out the statements (Marsh et al., 2003; Marsh & Fazio, 2006). When stories contain *inaccurate* statements (e.g., “The capital of Russia is St. Petersburg” or “Owens ran the first sub-4-minute mile”), participants also reproduce those inaccurate ideas on postreading tests to a greater degree than if they read versions leaving out the inaccuracies (Marsh, 2004; Marsh & Fazio, 2007).

Several projects have delineated the scope of readers’ use of inaccurate information. First, such use can emerge even when participants should know that what they are reading is incorrect. Similar effects of inaccurate information have been obtained whether or not the presented information was well-known (Eslick, Fazio, & Marsh, 2011; Fazio & Marsh, 2008a, 2008b; Marsh & Fazio, 2006; Marsh et al., 2003), as confirmed by general knowledge norms (e.g., Nelson & Narens, 1980) and through direct tests of participant knowledge (Fazio, Barber, Rajaram, Ornstein, & Marsh, 2013). Participants sometimes report having acquired their inaccurate understandings prior to reading experimental materials, even

---

D. N. Rapp (✉) · S. R. Hinze · K. Kohlhepp  
Northwestern University, 2120 Campus Drive,  
Evanston, Illinois 60208, USA  
e-mail: rapp@northwestern.edu

R. A. Ryskin  
University of Illinois at Urbana-Champaign,  
Urbana, Illinois 61820, USA

when the information is patently wrong and unlikely to have been seen or known beforehand (Marsh et al., 2003). These findings signify a crucial failure in the application of prior knowledge for comprehension.

Second, the problematic effects of reading inaccurate information have been demonstrated with explicit tests of participants' knowledge about specific facts, as above, as well as with recognition-based judgments of the validity of statements (Appel & Richter, 2007; Gerrig & Prentice, 1991; Prentice, Gerrig, & Bailis, 1997; Wheeler, Green, & Brock, 1999) and with measures of processing latencies for relevant information (Rapp, 2008). These convergent findings indicate that inaccuracies have the potential to create difficulties not just for the products of reading experiences, but also for moment-by-moment processing of unfolding text.

Third, interventions intended to encourage careful evaluation of text contents have proven unsuccessful at eliminating the influence of inaccurate information. Warning participants about potential inaccuracies in texts, both prior to and after reading, fails to reduce use of these inaccuracies (Marsh & Fazio, 2006). Similarly, requiring participants to retrieve accurate knowledge one week before misinformation is presented (Fazio et al., 2013), or just prior to reading misinformation (Rapp, 2008), fails to enhance the production of accurate facts, recalibrate judgments about the validity of related statements, or correct the reading patterns that exemplify an influence of misinformation. Asking participants to read stories multiple times (Marsh et al., 2003), presenting stories more slowly (Fazio & Marsh, 2008b), and explicitly identifying potential inaccuracies for readers (Eslick et al., 2011) also fail to improve participants' performance. These manipulations sometimes even result in *greater* reliance on text inaccuracies.

Why do people use obviously false information? One account of this propensity appeals to the role of prior knowledge and episodic memory traces during text experiences (but see also Richter, Schroeder, & Wöhrmann, 2009, for other accounts). Concepts wax and wane in memory during reading, in correspondence with the encoded text information (Gerrig & McKoon, 1998, 2001; Myers & O'Brien, 1998; O'Brien, 1995; Rapp & van den Broek, 2005; Ratcliff, 1978). For example, reading the word "Kentucky" automatically activates semantic associates from long-term memory, which are then available to the reader as the text unfolds. New episodic associations are built, and past semantic associations are strengthened or weakened, through these dynamic activations (van den Broek, Rapp, & Kendeou, 2005). By this account, incorrect information could become encoded into memory as new associations (e.g., Kentucky's capital is Louisville). Old, accurate associations (e.g., Kentucky's capital is Frankfort) also might not be sufficiently activated or strategically considered. In fact, if a new association (e.g., Louisville) is more familiar, available, or has more semantic associates related to the

concept being encoded (e.g., Kentucky), accurate knowledge can be blocked, inhibited, or ignored (J. R. Anderson, 1981; Gallo, 2010; Storm, 2011). The result is a stronger influence of inaccurate information than of prior knowledge on postreading judgments, recall, and comprehension.

Additionally, as individuals read inaccurate information that they have not encountered before (or have encountered infrequently), they encode new episodic traces. These traces may be readily available when participants are subsequently tested on related information, particularly if cues in the test questions or task requirements help reactivate those episodic details. General knowledge tests, judgment queries, and sentence comprehension tasks all provide cues that can foster such retrieval. For example, when participants read "What is the capital of Kentucky?," the question can act as a retrieval cue for episodic traces that include the misinformation. Additionally, retrieval of these inaccuracies can be influenced by the task demands if participants expect that they should use the information that they had previously read on subsequent tasks (Rapp & Mensink, 2011).

Because prior knowledge and episodic traces are necessarily involved in learning, and have been invoked in accounts of the acquisition of inaccurate information in particular, interventions might usefully target the encoding of these traces. Activities that do this could require participants to specifically evaluate and interact with material as it is read, to influence the episodic traces encoded during reading, as well as the complementary retrieval of accurate prior knowledge. One intervention applied in this manner (Marsh & Fazio, 2006) asked participants to explicitly track misinformation (i.e., press a key when you notice an inaccuracy), which resulted in slight decreases from the usual production of misinformation on postreading tests. The effectiveness of this intervention should be enhanced by directly encouraging retrieval from prior knowledge during the encoding of potential inaccuracies.

Doing this involves asking participants to monitor and act upon inaccuracies by applying accurate knowledge to identify and correct them. These evaluative criteria are routinely enacted during peoples' everyday editing of texts. Proofreading involves the detection of spelling and grammatical problems, whereas fact-checking involves establishing coherence within and across sentences and identifying and correcting inaccuracies in content (L. Anderson, 2006; Brunyé, Mahoney, Rapp, Ditman, & Taylor, 2012). These activities encourage the evaluation of text through comparisons of prior knowledge with what is being read, but for the purposes of reducing reliance on inaccuracies, fact-checking most usefully prompts monitoring and correction of text content. If fact-checking encourages readers to recruit prior knowledge to correct inaccuracies, we should see reductions in the pattern of incorrect judgments traditionally obtained when participants are asked to evaluate the validity of text-derived statements.

But this is not a foregone conclusion. Recall that explicitly identifying inaccuracies, slowing text presentations, and asking participants to reread texts—all of which draw attention to inaccuracies—have failed to decrease the influence of misinformation. Asking participants to act upon the false information that they read, even with the goal of correcting it, might encourage the encoding of inaccuracies that are later retrieved on postreading tasks; that is, asking participants to think about the misinformation, even in the service of correcting it, might increase later use of the misinformation. This could also occur as new associations derived from the inaccuracies are encoded into memory, blocking correct knowledge from being usefully applied on subsequent tasks. The result would be continued reliance on inaccurate information.

Across four experiments, we examined the consequences of reading a fictional<sup>1</sup> text that included assertions consistent either with an inaccurate idea (e.g., “Mental illness is contagious,” or “Seat belts do not save lives”) or with the accurate version of the same idea (e.g., “Mental illness is not contagious,” or “Seat belts save lives”). After reading, the participants were presented with single-sentence summarizations of the ideas and asked to judge whether the sentences were true or false. In past work, participants’ judgments of the validity of statements were less correct after reading inaccurate as compared to accurate assertions related to the statements (Gerrig & Prentice, 1991). We examined whether fact-checking would reduce these incorrect judgments.

## Experiments 1A and 1B

Previous projects documenting readers’ use of misinformation have tended to test information that is declarative in nature, specifically involving single-statement facts integrated into stories (e.g., the names of people or events, geographical trivia, etc.; Eslick et al., 2011; Fazio & Marsh 2008a, 2008b; Marsh, 2004; Marsh & Fazio, 2006, 2007; Marsh et al., 2003). As an important extension, we tested participants’ acquisition of assertions about the world (Gerrig & Prentice, 1991). These assertions represent general categories of information around which individuals organize their understandings of events, behaviors, and attitudes, often as informed by real-world experiences. Examples include the notions that brushing your teeth can prevent gum disease and that a college education can help one get a better job. These assertions are learned

through experience and through explicit or implicit reminders of their relevance and applicability to our lives, in formal and informal settings. This contrasts with declarative facts, which are often directly learned from textbooks, teachers, or trivia games. Assertions offer an additional set of materials to test participants’ use of inaccurate information, and to date, such assertions have not been subjected to instances intended to encourage evaluations of their content as misinformation. Thus, we began by testing whether these materials would obtain the patterns that have been reported in other studies.

In Experiment 1A, participants read a story containing accurate and inaccurate assertions (normed to ensure that the general population was aware of the accuracy of the assertions), and afterward judged the validity of test statements that summarized either the accurate or inaccurate forms of the assertions. If participants inappropriately relied on the misinformation conveyed in the story, they should be more likely to make incorrect judgments of test statements after reading misinformation than after reading correct information.

Without a baseline measure of performance on the judgment task, though, the predicted effects would be open to interpretation. The predicted pattern could reflect an increased reliance on inaccurate information in contrast to how participants might respond outside of the story’s influence, or an increased reliance on accurate information from the story in contrast to what their normal judgments would reflect. In Experiment 1B, we presented a new story that contained absolutely no information related to the assertions, followed by the judgment task. The degree to which participants’ responses patterned similarly to the accurate and inaccurate judgments from Experiment 1A would provide insight into the specific influence of story content.

## Method

**Participants** A group of 28 Northwestern University undergraduates participated in Experiment 1A, and a different sample of 28 Northwestern University undergraduates participated in Experiment 1B, all for course credit. All were native speakers of English.

**Apparatus** Experiment 1A utilized one of two versions of a 19-page printed booklet containing the experimental story; Experiment 1B utilized only one version of a 19-page printed booklet containing the control story. Across both experiments, the distractor booklets contained visuospatial and math puzzles, one puzzle printed per page. The second part of both experiments used a Pentium PC computer running Superlab software. Participants sat in front of a Dell color monitor with their hands on the keyboard to make appropriate responses. Test sentences

<sup>1</sup> We used a fictional text for three reasons: (1) previous work has focused on fictional texts to demonstrate that participants rely on their contents even though they seemingly should encourage more skeptical evaluation than do expository materials; (2) interactions with fiction represent an important set of experiences that prove directly relevant to everyday life; and (3) fiction, as opposed to nonfiction, affords the opportunity to design materials in which inaccuracies can be presented without creating inordinately awkward content and structure.

were displayed in the center of the screen in standard upper- and lowercase black type.

*Materials* For Experiment 1A, the 19-page fictional story “The Kidnapping” from Prentice et al. (1997) was adapted for the project. The story setting was changed from Yale University to Carleton College to make it less familiar to participants. The story described events over the course of a day, with characters talking to each other about various topics. Sixteen of these topics constituted the experimental assertions, conveyed in either accurate or inaccurate forms (see Appendix A for sample assertions). The assertions were described in conversations lasting one to six short conversational paragraphs in length (each approximately one to four sentences long). The assertions always appeared in the same order in the story.

Before beginning, we conducted a norming study to ensure that the accurate and inaccurate versions of the assertions would be, a priori, well known to participants, as the original materials were not normed in this manner (Prentice et al., 1997). Additionally, any expectations for what people should believe to be accurate might have changed since the original implementation of the materials. We asked 28 Northwestern University undergraduates, none of whom participated in any other part of the study, to read statements summarizing the inaccurate and accurate forms of each assertion (e.g., “Not brushing your teeth enough can lead to gum disease” vs. “Brushing your teeth can lead to gum disease”; “Having a college education can help you get a better job” vs. “Having a college education won’t help you get a better job”). The two-page printed survey presented the accurate and inaccurate forms of each statement side by side. Participants were asked to put a check next to the statement from each of the 16 pairs that they believed to be the most accurate. Two forms of the survey were created to counterbalance the placement of each statement in a pair, with pairs being randomly ordered in each version of the survey.

The average agreement for the correct assertions across all items was 82.37 %, with the accuracy for each assertion ranging from 21.43 % to 100 %. Three of the assertions obtained agreement rates that fell below a reasonable expectation of 65 % agreement, so we opted to remove those items and replace them with new assertions. We conducted a second norming study to assess expectations for the 13 original and three new assertions. A second group of 30 Northwestern undergraduates, none of whom participated in any other part of the study, completed the survey. The average agreement for the correct assertions across all items was now 91.77 %, with the accuracy for each assertion ranging from 67.86 % to 100 %. These results gave us confidence that participants would have correct expectations for the assertions to be presented in the story.

For Experiment 1B, the fictional story “The Raven” by Robert Twohy was shortened to equate it for length with the story from Experiment 1A. The story was 19 pages long (7,291 words), describing a private investigator’s attempt to solve a mystery. The text did not discuss topics that were related in any way to the assertions to be tested.

For the experiments, we used the normed assertions as the basis for the 16 test statements in the judgment task. Each statement could be presented in its accurate or inaccurate form. The statements were equated for length [ $M = 10.5$  words for accurate statements and 10.8 words for inaccurate statements;  $t(15) < 1$ ]. We also included 16 filler statements unrelated to information in the story in order to make the task less obviously related to the text. All of the fillers described urban legends that, although false, are often debated, and that were so selected as to be less clear with respect to their truth value (e.g., “Chewing gum takes seven years to pass through the digestive system” or “The average person swallows eight spiders a year”), in contrast to the experimental statements.

*Design* For Experiment 1A, story assertion (accurate vs. inaccurate) and test statement (accurate vs. inaccurate) varied as within-participants variables. There were two versions of the story, with eight assertions appearing in their accurate form in one version and in their inaccurate form in the other, and vice versa for the remaining eight assertions. The two versions were 7,197 words and 7,099 words, respectively. There were also two versions of the subsequent judgment task, with eight test statements presented in their accurate form in one version and in their inaccurate form in the other, and vice versa for the remaining eight statements. The 16 filler statements were added to each list. We used a Latin-square design to construct four sets of materials, with each set containing one version of the story and one version of the judgment task. Participants were presented with one set of materials in the counterbalanced design. The test statements in the judgment task were presented in a different random order for each participant.

For Experiment 1B, the design was identical, except that it included only one version of the story. Thus, only test statement (accurate vs. inaccurate) was varied within participants.

*Procedure* Participants completed the experiments individually. They were instructed to read the story at their own pace and to inform the experimenter when they had finished. Participants then worked on the distractor booklet for 7.5 min. Next they used a computer to complete the judgment task. The participants were instructed as follows:

In this part of the experiment you will be presented with statements. Your task is to decide whether each statement is true or false. Please respond to each of the following statements by pressing either the TRUE–YES or FALSE–NO keys. We would like you to answer according to whether or not the statement is true in everyday life.

Participants responded by pressing the “J” (labeled YES) or “K” (labeled NO) keys on the keyboard. The “A” key was labeled NEXT and used to begin the task when participants were ready. When done, participants were debriefed and thanked for participating.<sup>2</sup>

Results and discussion

The experimental analyses were conducted using analysis of variance (ANOVA) with participants ( $F_1$ ) and items ( $F_2$ ) as random variables. We also conducted generalized linear mixed-models (GLMM) analyses for error rates. For this latter analysis, we included both participants and items as random factors in a single model, to address fundamental issues with their separate evaluation (Baayen, Davidson, & Bates, 2008; Quené & van den Bergh, 2008; Richter, 2006). The lme4 package for R was used to conduct these analyses (Bates, Maechler, & Bolker, 2011). For simplicity, rather than presenting the models in detail, we present significance tests for the fixed effects of interest ( $z$  scores for GLMM analyses), as was recommended by Baayen et al. Prior to the analysis, we eliminated responses falling more than 2.5 standard deviations above the mean response time for each participant, resulting in a loss of less than 1 % of the data for Experiment 1A, and 2.44 % in Experiment 1B. Due to experimenter error, one story item was only presented in its accurate form, so we eliminated the item in this and all subsequent analyses. Table 1 presents the mean percentages of incorrect judgments for Experiment 1A; Table 2 depicts the same data for Experiment 1B.

For Experiment 1A, the accuracy analyses focused on the percentages of incorrect judgments produced by each participant as a function of the information that had been read in the story. Incorrect judgments included cases in which participants identified inaccurate test statements as being true and accurate test statements as being false. Overall, participants made twice as many incorrect judgments after reading inaccurate information ( $M = 29.39\%$ ) than after reading accurate information ( $M = 14.52\%$ ), supported by a significant main effect of story assertion [ $F_1(1, 27) = 6.99$ ,

**Table 1** Mean percentages of incorrect judgments, with standard deviations, from Experiment 1A

Test Statement	Accurate Assertion in Story		Inaccurate Assertion in Story		Mean
	Mean	SD	Mean	SD	
Accurate	13.61	20.90	31.14	28.74	22.38
Inaccurate	15.43	24.43	27.64	31.91	21.54
Mean	14.52		29.39		

$MSE = .39, p < .05, \eta_p^2 = .21; F_2(1, 14) = 13.39, MSE = .12, p < .01, \eta_p^2 = .49; GLMM, z = 2.83, p = .005$ ]. Neither the effect of test statement nor the interaction between story assertion and test statement was significant (all  $F_s < 1; GLMM, z_s < .40$ ).

These results replicated previous findings on readers’ use of misinformation, this time with assertions rather than declarative stimuli. After participants read assertions that were inaccurate, they were twice as likely to make incorrect judgments on a subsequent test of the validity of those assertions, relative to after having read accurate assertions in the story. However, these findings do not indicate whether the effects were due to participants having learned inaccurate information, which would be associated with increases in incorrect judgments, or whether participants benefited from reading the accurate information, which would be associated with decreases in incorrect judgments. Experiment 1B tested the direction of the effect.

In Experiment 1B, participants produced similar percentages of incorrect judgments, regardless of whether or not the test statement was accurate ( $M = 14.45\%$  overall) (both  $F_s < 1; GLMM, z < 1$ ). Similar performance for both types of statements following this control story provided confidence that the effects observed in Experiment 1A were due to the story contents, rather than to a priori differences in the agreement rates for accurate and inaccurate outcomes.

We performed cross-experimental comparisons to directly test whether the judgment patterns from Experiment 1B resembled those following accurate or inaccurate story assertions from Experiment 1A. First, we compared the judgment patterns following the control story to the judgment

**Table 2** Mean percentages of incorrect judgments, with standard deviations, from Experiment 1B

Test Statement	Mean	SD
Accurate	13.95	14.29
Inaccurate	14.94	12.33
Mean	14.45	

<sup>2</sup> The participants also completed three personality and individual-difference surveys as part of the laboratory procedures at the conclusion of the experiment, including measures of narrative transport and need for cognition. The scores on these surveys did not correlate with performance on the task and are omitted from further discussion.

patterns associated with accurate assertions that had appeared in the experimental story. Inspection of the means suggested little difference, with participants making 14.45 % incorrect judgments after reading the story in Experiment 1B containing no assertions, and 14.52 % incorrect judgments after reading accurate assertions in Experiment 1A. No differences were obtained as a function of test statement, experiment, or the Test Statement  $\times$  Experiment interaction (all  $F$ s  $< 1$ ; GLMM,  $z$ s  $< 1$ ). This suggests that participants were not more likely to learn or use accurate information after reading accurate assertions from the experimental story than they might be after reading the control story containing no assertions whatsoever.

We next compared performance after the control story to performance following the inaccurate assertions presented in the experimental story. In contrast to the previous analysis, here we found a significant effect of experiment [ $F_1(1, 54) = 14.74$ ,  $MSE = .23$ ,  $p < .001$ ,  $\eta_p^2 = .21$ ;  $F_2(1, 28) = 24.232$ ,  $MSE = .194$ ,  $p < .001$ ,  $\eta_p^2 = .46$ ; GLMM,  $z = 2.61$ ,  $p = .009$ ], with participants making 14.45 % incorrect judgments after the control story in Experiment 1B, as compared to 29.39 % incorrect judgments after reading inaccurate assertions in Experiment 1A. The main effect of test statement and the Test Statement  $\times$  Experiment interaction were not significant (all  $F$ s  $< 1$ ; GLMM,  $z$ s  $< 1$ ).<sup>3</sup>

Participants' performance following the control story revealed fewer incorrect judgments as compared to the performance of participants who had read inaccurate information from the experimental story. In contrast, the judgment patterns were similar when comparing participants who had read no information related to the test statements with participants who had read accurate information directly relevant to the statements. The results suggest that inaccurate information had a greater impact on judgments than did accurate information. This differs from the results of previous studies, which have shown reliance on both accurate and inaccurate information, although those studies focused on declarative facts rather than assertions. Most importantly, the data confirmed that the experimental content elicited a reliance on inaccurate information. We next tested whether that reliance could be reduced if participants were explicitly tasked with evaluating the story content.

<sup>3</sup> We also conducted these analyses by conceptualizing performance in Experiment 1B as a baseline and subtracting the data from Experiment 1A. Specifically, we subtracted out control accuracy rates from the accurate and inaccurate statement accuracy rates obtained in Experiment 1A. This analysis showed that the main effect of story assertion was still significant [ $F_1(1, 27) = 6.99$ ,  $MSE = .386$ ,  $p < .05$ ,  $\eta_p^2 = .21$ ;  $F_2(1, 14) = 13.23$ ,  $MSE = .02$ ,  $p < .01$ ,  $\eta_p^2 = .49$ ], whereas the main effect of test statement and the interaction were not (all  $F$ s  $< 1$ ). Again, the pattern suggests that participants acquired inaccurate rather than accurate information.

## Experiment 2

In Experiment 2, participants read the story from Experiment 1A while checking for factual errors. Fact-checking requires monitoring for misinformation and retrieving accurate knowledge to discount or complement inaccuracies with correct information (L. Anderson, 2006). Recall that Experiment 1A revealed a main effect of story assertion, with participants making more incorrect judgments after reading inaccurate than after reading accurate assertions. Elimination of this main effect would provide evidence of the utility of the fact-checking activity.

This evaluative task could prove effective in at least two, non-mutually-exclusive ways. First, if participants were successful at noticing and revising misinformation, they could encode episodic memories for their accurate consideration of the content, or perhaps tag the information as inaccurate. Second, the retrieval of accurate prior knowledge with the goal of checking for inaccuracies might foster subsequent retrieval of that activated knowledge during the judgment task. Alternatively, fact-checking might not prove effective if (a) participants failed to notice inaccuracies and/or to adequately apply their accurate knowledge, and/or (b) the mere exposure to inaccurate information made it accessible at test, despite being contradicted by any retrieved knowledge. Given the well-replicated consequences of inaccurate information, including the effects in Experiment 1A, we considered this the null hypothesis for the project, which would be supported if the main effect of story assertion was again obtained.

These two hypotheses, that an evaluative activity like fact-checking would or would not be effective at reducing the use of misinformation, are both viable possibilities. Tasks that encourage the activation and retrieval of prior knowledge can foster revision and support careful, veridical comprehension (e.g., Alvermann & Hague, 1989; Guzzetti, Snyder, Glass, & Gamas, 1993; Pearson, Hanson, & Gordon, 1979; Spires & Donley, 1998; van den Broek & Kendeou, 2008). However, previous attempts to encourage noticing of inaccuracies have failed to reduce, and sometimes have actually increased, the use of misinformation. The fact-checking task thus provided potential insight into influences on readers' use of misinformation, as well as a test of whether the earlier effects could be mitigated.

## Method

**Participants** A group of 28 Northwestern University undergraduates participated for course credit, none of whom had participated in the previous experiments. All were native speakers of English.

**Apparatus** The apparatus was identical to that of Experiment 1A, with the inclusion of a pencil during the first part of the experiment to be used for fact-checking.

**Materials** The materials were the same as in Experiment 1A.

**Design** The design was identical to that of Experiment 1A.

**Procedure** The procedure was identical to that of Experiment 1A, with the following change. During the first part of the experiment, participants were given a pencil and asked to fact-check the text that they were going to read. The instructions stated:

Your task, while you read this text, is to edit the text and make corrections directly to the document. Specifically, we would like you to make edits to the particular facts that are described in the story. If you find any statements that you believe are inaccurate or problematic, please change them. Please make your changes to the document directly, so that we will be able to read your modifications and edits.

**Coding of edits** The presence or absence of editing marks was coded for each sentence of the text. We chose sentences as the grain size for the analysis because each assertion was similarly presented in sentences. A single edit was coded as any modification or comment; multiple comments to a single sentence or the same comment made multiple times counted as one edit. Given that this presence/absence coding was straightforward and objective, only one trained researcher conducted this count. In addition, we coded whether the edits to inaccuracies included only marks such as a strikethrough of a line of text, or whether the edit included a written correction noting why an assertion was incorrect (e.g., by providing the correct information). To establish the reliability of this mark-versus-correction distinction, a second researcher coded 25 edited inaccuracies (20.33 % of the sample). We found strong interrater agreement, with only one disagreement ( $K = .92$ ).

## Results and discussion

We eliminated responses falling more than 2.5 standard deviations above the mean response time for each participant, resulting in a loss of less than 1 % of the data. Table 3 presents the mean percentages of incorrect judgments for Experiment 2.

We first examined whether participants engaged in the fact-checking task by calculating the number of edits that they made to the text. Participants made, on average, 17.25 ( $SD = 10.76$ ) edits to the story, ranging across participants

**Table 3** Mean percentages of incorrect judgments, with standard deviations, from Experiment 2

Test Statement	Accurate Assertion in Story		Inaccurate Assertion in Story		Mean
	Mean	SD	Mean	SD	
Accurate	17.14	23.92	16.86	21.73	17.00
Inaccurate	10.96	22.34	20.11	28.46	15.54
Mean	14.05		18.49		

from zero to 58 edits. These edits included rewrites and comments, as well as spelling and grammar markup. We next calculated only the number of rewrite and commenting edits specifically made to the 16 experimental assertions: Participants made, on average, 5.43 ( $SD = 3.35$ ) edits to the assertions, ranging from zero to 14 edits. Finally, we examined how many of those rewrite and commenting edits were specifically applied to the eight inaccurate assertions in the story. Participants edited, on average, 4.25 ( $SD = 2.32$ ) of those assertions, ranging from zero to eight edits. Thus, participants made a considerable number of edits for a document that contained only eight inaccurate assertions, and substantive corrections were made to half of the presented inaccuracies. Although participants were not always successful in correcting the misinformation, and in some cases even edited other sections of text for clarity and creativity rather than for inaccuracies, they nevertheless noticed many of the included inaccuracies. This gave us confidence that they were at least taking the task seriously.

We examined whether engaging in the fact-checking task reduced their reliance on inaccurate information, which would be exemplified by a nonsignificant effect of story assertion. The main effect of story assertion, obtained in Experiment 1A, was not significant in Experiment 2 (both  $F_s < 1$ ; GLMM,  $z = 1.75$ ,  $p = .08$ ), despite using the same text with the same number of participants. Neither the effect of test statement nor the interaction of test statement and story assertion was significant (all  $F_s < 1.28$ ; GLMM,  $z_s < 1.14$ ).

As an additional method of evaluating the effects of fact-checking, we compared the percentages of incorrect judgments produced in Experiment 1A versus Experiment 2. Overall, a main effect of story assertion obtained, with participants making a higher percentage of inaccurate judgments after reading inaccurate ( $M = 23.94$  %) as compared to accurate assertions ( $M = 14.29$  %) [ $F_1(1, 54) = 5.791$ ,  $MSE = .37$ ,  $p < .05$ ,  $\eta_p^2 = .10$ ;  $F_2(1, 28) = 12.49$ ,  $MSE = .09$ ,  $p < .01$ ,  $\eta_p^2 = .31$ ; GLMM,  $z = 3.23$ ,  $p = .001$ ]. We also observed a main effect of experiment that was significant by participants only [ $F_1(1, 54) = 4.07$ ,  $MSE = .15$ ,  $p < .05$ ,  $\eta_p^2 = .07$ ;  $F_2 < 1$ ; GLMM,  $z = 1.51$ ,  $p = .13$ ], with more incorrect judgments being made in Experiment 1A ( $M = 21.96$  %) than in Experiment 2 ( $M = 16.27$  %). Finally, the

interaction between story assertion and experiment was significant by items only [ $F_1 < 2.1$ ;  $F_2(1, 28) = 6.17$ ,  $MSE = .09$ ,  $p < .05$ ,  $\eta_p^2 = .18$ ; GLMM,  $z < 1$ ], which is best exemplified by subtracting incorrect judgments following accurate assertions from incorrect judgments following inaccurate assertions; this simple calculation illustrates how much difficulty participants exhibited after reading inaccurate as compared to accurate information. In Experiment 1A, participants made more errors (a 14.87 percentage point difference) after inaccurate as compared to accurate assertions; in Experiment 2, this difference was reduced (4.44 percentage points).

Finally, we conducted conditional analyses looking specifically at judgments as a function of whether participants had previously made edits to relevant text content. This analysis helped address whether fact-checking was effective due to a general processing approach that participants may have adopted to read the texts, or to item-specific processes involved in editing specific assertions. In this analysis, we were specifically interested in assertions that were presented inaccurately in the text. For each participant, we calculated the proportion of errors after both “hits” (correct edits of any kind to inaccurate assertions) and “misses” (no edits of any kind to inaccurate assertions). Error rates were calculated for both accurate and inaccurate test statements (see Table 4). We found a significant conditional effect of fact-checking, as participants made fewer judgment errors when an inaccuracy was edited ( $M = 6.67\%$ ) than when it was missed ( $M = 22.78\%$ ), [ $F(1, 14) = 7.78$ ,  $p = .014$ ,  $\eta_p^2 = .357$ ]. No main effect or interaction with test statement emerged ( $F_s < 0.13$ ). For this analysis, though, only 15 participants provided data that included both hits and misses in the two test statement conditions. So, we next collapsed across test statement conditions to maximize the number of observations of hits and misses, yielding data from a total of 23 participants. Again, we observed a significant conditional effect of fact-checking, as participants made fewer errors when an inaccuracy was edited ( $M = 10.69\%$ ) than when it was missed ( $M = 26.42\%$ ),  $t(22) = 2.19$ ,  $p = .039$ ,  $d = 0.62$ .<sup>4</sup> In fact, the error rates associated with missed inaccuracies were comparable to the error rates produced by participants in Experiment 1A ( $M = 29.39\%$ ) who had read false assertions without instructions to proofread,  $t(51) = 0.15$ .

<sup>4</sup> We also considered whether different types of edits may have had different influences on subsequent error rates. Specifically, a more conservative coding was also adopted in which an edit was coded as a hit only if it included a written correction, rather than other marks such as a strikethrough of a line of text. A total of 20 participants had both misses and corrections (after excluding simple marks or strikethroughs). As with the more liberal coding scheme, participants made fewer errors when an inaccuracy was explicitly corrected ( $M = 7.67\%$ ) than when it was missed ( $M = 27.05\%$ ) [ $t(19) = 2.64$ ,  $p = .016$ ,  $d = 0.83$ ].

**Table 4** Mean percentages of incorrect judgments on facts presented inaccurately in the story, conditionalized on participants’ editing of facts ( $n=15$ ), from Experiment 2

Test Statement	Inaccuracies Edited		Inaccuracies Missed		Mean
	Mean	SD	Mean	SD	
Accurate	4.44	18.76	22.22	36.55	16.11
Inaccurate	8.89	11.73	23.33	37.16	13.33
Mean	6.67		22.78		

Overall, participants demonstrated use of inaccurate information, but that use was attenuated by instructions to fact-check during reading. Instructing people to actively reflect on what they know as they read appears to help protect against the influence of obvious inaccuracies. The particular evaluative activity that we tested required participants to monitor for discrepancies between text content and their prior knowledge, and to make changes directly to the text to remediate the inaccuracies underlying those discrepancies. But did the effectiveness of this activity depend on readers generating modifications to text content, versus their simply evaluating the veracity of the information? The results of Experiment 2 argue against the latter possibility, on the basis of the conditional analyses following edited and nonedited inaccuracies. Nevertheless, we investigated this issue further in Experiment 3.

In addition, an important concern for the project was whether the experimental stimuli might have encouraged greater consideration of inaccurate ideas as being plausible than was intended by the stimulus design or has been evaluated in previous studies. First, the story assertions were embedded in conversations that included statements supporting their validity. Consider an example from Appendix A: The inaccurate assertion that seat belts might actually cause harm references the possibility of being ensnared when escape from a car is critical for survival. Although the idea that seat belts save lives is undoubtedly true, the counterexample provides contextual support for the inaccurate assertion, despite the likelihood of its occurrence being very low. The stimuli might thus have encouraged participants to consider situations aligning with the inaccurate assertions, increasing the likelihood of judging them as being true. Second, the test statements often included auxiliary verbs (e.g., “can” in “Wearing a seatbelt can decrease your chances of living through an accident.”) that also potentially invoked low-frequency reflections that would support the plausibility of inaccurate assertions. Third, the filler test statements were always inherently false (although, again, on the basis of ideas that people endorse to varying degrees). Participants may have therefore adopted decision strategies (e.g., Reder, 1979) that involved consistently rejecting information that did not appear in the story,



and complementarily, that allowed for the acceptance of any information that had appeared in it. In Experiment 3, we attempted to discount these factors as driving any reliance on inaccurate information.

### Experiment 3

When false information is clearly identified for participants, its subsequent use can actually increase above traditionally obtained levels (Eslick et al., 2011). This might be taken as evidence that any reductions in reliance are unlikely to occur when readers are passively decoding text, rather than directly acting on content as part of their evaluations. In line with this view, asking participants to press a key when they notice false information leads to modest decreases in subsequent use (Marsh & Fazio, 2006), whereas the results of Experiment 2 indicated more substantial reductions after fact-checking. In Experiment 3, to more directly assess any benefits, we included fact-checking and reading conditions in a single between-participants experiment. We also added a condition in which participants were instructed to highlight inaccuracies without making direct changes to content. This allowed for directly testing whether the benefits of evaluation require active correction of content, or whether they can also be obtained if participants merely identify inaccuracies without changing them. In addition, we modified the stimuli to reduce potential biasing content and to discourage decision strategies that may have helped drive reliance on inaccurate information provided in the story.

#### Method

**Participants** A group of 96 Northwestern University undergraduates participated for course credit or pay (\$12). All were native speakers of English.

**Apparatus** The apparatus was the same as in Experiment 2, with the inclusion of a highlighter for the highlighting condition.

**Materials** We modified the 19-page fictional story and test statements from the previous experiments as follows. First, we removed all sentences from the text that supported the claims offered by accurate or inaccurate assertions. Thus, the critical assertions were presented without logical or evidentiary support (see Appendix B for examples). This reduced the length of the story by 1,621 and 1,473 words for Versions 1 and 2, respectively. Second, we made minor modifications to the test statements so that they clearly supported assertion-relevant ideas that were either true or false, to avoid encouraging specific counterexamples as part of judging their validity. For example, the statement “Wearing a seatbelt can increase

your chances of living through an accident” was modified to read, “Wearing a seatbelt increases the likelihood of surviving a car accident.” Third, to reduce the potential for filler content to encourage decision biases on the judgment task, seven filler statements were rewritten and three new filler statements were written as replacements. This ensured that the 24 filler statements included 12 true and 12 false ideas.

**Design** Participants were randomly assigned to one of three instructional conditions: a fact-checking condition ( $n = 32$ ), a highlighting condition ( $n = 32$ ), or a control condition (reading only;  $n = 32$ ). The remainder of the design was identical to those of Experiments 1A and 2.

**Procedure** The control condition was identical to that of Experiment 1A, and the fact-checking condition was identical to that of Experiment 2. Participants in the highlighting condition were, during the first part of the experiment, given a highlighter and instructed:

Your task, while you read this text, is to review the text and mark errors directly on the document. Specifically, we would like you to evaluate the particular facts that are described in the story. If you find any statements that you believe are inaccurate or problematic, please highlight them. Please just highlight the facts without changing them. Make sure you highlight the factual information specifically so that we will be able to see which facts you have marked.

The second part of the experiment was identical to that in the previous experiments.

**Edit coding** The presence or absence of editing marks was coded for each sentence in a manner similar to that in Experiment 2. We did not, however, differentiate between simple marks and written corrections because (a) this distinction had not seemed to influence the results of Experiment 2, and (b) participants in the highlighting condition were unable to make written corrections. Given the rather objective nature of this presence/absence coding, only one researcher conducted this count.

#### Results

We eliminated responses falling more than 2.5 standard deviations above the mean response time for each participant, resulting in a loss of 2.79 % of the data. Table 5 presents the mean percentages of incorrect judgments for Experiment 3. We first examined whether participants engaged in the tasks as instructed. For the highlighting condition, we counted edits that included any marks to text content, and for the fact-checking condition, we counted edits that included rewrites and comments, as well as

**Table 5** Mean percentages of incorrect judgments by instructional conditions, with standard deviations, from Experiment 3

Test Statement	Accurate Assertion in Story		Inaccurate Assertion in Story		Mean
	Mean	SD	Mean	SD	
Reading-Only Condition					
Accurate	11.72	15.53	22.66	26.46	17.19
Inaccurate	7.03	13.07	16.41	20.68	11.72
Mean	9.38		19.53		
Fact-Checking Condition					
Accurate	13.28	17.94	10.16	12.48	11.72
Inaccurate	7.03	13.07	7.03	11.42	7.03
Mean	10.16		8.59		
Highlighting Condition					
Accurate	15.63	15.23	16.41	20.68	16.02
Inaccurate	9.38	13.84	8.59	16.33	8.98
Mean	12.50		12.50		

spelling and grammar markup. Across conditions, participants made, on average, 11.38 ( $SD = 8.38$ ) marks to the story, ranging from zero to 47 marks. We next looked only at the number of marks made to the 16 experimental assertions. Participants marked, on average, 5.97 ( $SD = 2.26$ ) of those assertions, ranging from zero to ten marks. Finally, we looked at how many of those marks were specifically made to the eight inaccurate assertions in the story. Participants marked, on average, 5.28 ( $SD = 1.95$ ) of those assertions, ranging from zero to eight marks. Participants in the fact-checking and highlighting conditions made a similar numbers of marks, both overall [ $M_{fc} = 12.59$ ,  $SD_{fc} = 10.30$ ;  $M_{hl} = 10.16$ ,  $SD_{hl} = 5.79$ ;  $t(62) = 1.17$ ,  $p = .25$ ] and specifically to inaccurate assertions [ $M_{fc} = 5.66$ ,  $SD_{fc} = 1.72$ ;  $M_{hl} = 4.91$ ,  $SD_{hl} = 2.12$ ;  $t(62) = 1.56$ ,  $p = .12$ ]. Participants made substantive marks to more than half of the presented inaccuracies, whether they were tasked with correcting or highlighting. This suggested that they took the task seriously and could effectively evaluate many of the inaccuracies.

We next examined the percentages of incorrect judgments produced by each participant, as a function of the story assertions that they read and their task instructions. Overall, participants made fewer incorrect judgments in the fact-checking condition ( $M = 9.4\%$ ) than after highlighting ( $M = 12.5\%$ ), and as compared to controls ( $14.5\%$ ); these differences were marginal by participants and significant by items and the GLMM analysis [ $F_1(2, 93) = 2.86$ ,  $MSE = .029$ ,  $p = .06$ ,  $\eta_p^2 = .06$ ;  $F_2(1, 15) = 3.63$ ,  $MSE = .016$ ,  $p = .04$ ,  $\eta_p^2 = .20$ ; GLMM,  $z = 2.27$ ,  $p = .02$ ]. More importantly, this main effect was qualified by an interaction with story assertion that was significant by participants and GLMM analysis, and marginal by items [ $F_1(2, 93) = 3.50$ ,  $MSE = .037$ ,  $p = .03$ ,

$\eta_p^2 = .07$ ;  $F_2(1, 15) = 3.15$ ,  $MSE = .012$ ,  $p = .06$ ,  $\eta_p^2 = .17$ ; GLMM,  $z = 2.14$ ,  $p = .03$ ]. Control participants made twice as many incorrect judgments after reading inaccurate ( $M = 19.53\%$ ) as compared to accurate ( $M = 9.38\%$ ) assertions [ $t(31) = 2.55$ ,  $p = .02$ ,  $d = 0.65$ ]. In contrast, participants who fact-checked or highlighted the story had similar levels of incorrect judgments after reading inaccurate ( $M_{fc} = 8.60\%$ ;  $M_{hl} = 12.50\%$ ) and accurate ( $M_{fc} = 10.16\%$ ;  $M_{hl} = 12.50\%$ ) assertions [all  $t$ s  $< 1$ ]. Unlike before, we observed a main effect of test statement, with participants making more incorrect judgments to accurate ( $M = 14.97\%$ ) than to inaccurate ( $M = 9.24\%$ ) test statements, a result that was significant by participants and marginal by items [ $F_1(1, 93) = 14.51$ ,  $MSE = .02$ ,  $p < .001$ ,  $\eta_p^2 = .14$ ;  $F_2(1, 15) = 4.28$ ,  $MSE = .031$ ,  $p = .06$ ,  $\eta_p^2 = .22$ ; GLMM,  $z = 1.58$ ,  $p = .11$ ]. No other main effects or interactions were significant (all  $F$ s  $< 2.13$ , GLMM,  $z$ s  $< 1.05$ ).<sup>5</sup>

We again conducted conditional analyses to determine whether corrections to specific assertions influenced judgments about the validity of the related test statements, using the same analytic procedure described in Experiment 2, and focused specifically on the fact-checking and highlighting conditions. We found a significant conditional effect of marking, with participants making fewer judgment errors after they had marked an inaccurate assertion than when it was missed,  $F(1, 21) = 12.52$ ,  $p = .002$ ,  $\eta_p^2 = .37$ . No main effect or interaction emerged with test statement or fact-checking condition (all  $F$ s  $< 1.02$ ). This analysis included only 23 participants who provided observations for hits and misses in both test statement conditions, so we next collapsed across test statement conditions to maximize the number of observations, yielding data from 56 participants. For this analysis, participants again made fewer judgment errors when they marked inaccuracies ( $M_{fc} = 4.03\%$ ,  $M_{hl} = 2.40\%$ ) than when they did not ( $M_{fc} = 26.81\%$ ,  $M_{hl} = 22.60\%$ ) [ $F(1, 54) = 18.93$ ,  $p < .001$ ,  $\eta_p^2 = .26$ ]. No main effect or interaction with fact-checking condition was evident (all  $F$ s  $< 1$ ). As in Experiment 2, instructional benefits were obtained for assertions that were correctly identified as inaccurate during reading. Error rates after reading misinformation were no higher for the control group than for participants who failed to mark errors in either the fact-checking group,  $t(61) = 1.25$ ,  $p = .22$ , or the highlighting group,  $t(51) = 0.15$ ,  $p = .63$ .

The results from Experiment 3 provided a replication and extension of the previously reported cross-experimental effects of fact-checking. When the story contained inaccurate assertions, participants used that information later, leading

<sup>5</sup> Overall, the percentage of inaccurate judgments in Experiment 3 was lower than the percentages obtained in the previous experiments. This pattern is likely attributable to the modifications of the materials described in the Method section. But even with such reductions, the obtained patterns of inaccurate judgments between conditions resembled those obtained across Experiments 1A and 2.

to errors on the postreading judgment task. But participants' use of the inaccurate information was reduced when they were also tasked with carefully evaluating the materials. These benefits were observed both when participants were instructed to make corrections to text content and when they were asked only to highlight inaccurate information, without changing it.<sup>6</sup>

## General discussion

People can problematically use inaccurate information from texts to complete postreading tasks. Experimental attempts to discourage participants from using misinformation have failed to substantially reduce these effects. The present project was devoted to examining whether evaluative instructions could help decrease the use of inaccuracies, guided by hypotheses about the contributions of episodic traces and prior knowledge.

In Experiment 1A, participants read a story containing both accurate and inaccurate assertions, afterward judging the validity of statements that summarized the assertions. Participants made more incorrect judgments after reading inaccurate rather than accurate assertions in the story. These performance decrements were observed for judgments of both accurate and inaccurate test statements, and the decrements emerged even though the materials were normed to ensure that participants, a priori, would know whether or not the statements were accurate. In Experiment 1B, the participants read a control story containing no relevant assertions, to gauge baseline performance on the judgment task. The pattern of participants' judgments was comparable to those specifically obtained following accurate assertions in Experiment 1A. Participants appeared to be influenced more by misinformation than by accurate information.

In Experiment 2, the participants read the story containing both accurate and inaccurate assertions, but they were also instructed to fact-check the contents in order to correct inaccurate information. They again exhibited some use of inaccurate information, but the effects were reduced. The

percentage of incorrect judgments was lower than those that had previously been obtained, eliminating the differences that emerged from having read accurate as compared to inaccurate assertions in the story. These reductions were most obvious for statements that participants had corrected in the story. Experiment 3 replicated the findings from Experiments 1A and 2, while also revealing reductions when participants were tasked with identifying but not correcting inaccuracies. Both fact-checking and highlighting instructions reduced participants' use of misinformation, as compared to the performance of participants who merely read the text. And, as before, the reductions were most apparent for the inaccurate information that participants had successfully marked in the story.

The benefits observed in both our fact-checking and highlighting conditions might suggest that readers tasked with such instructions adopt evaluative mindsets that help to discourage a liberal reliance on text content. Such mindsets have been invoked in discussions of the types of evaluative approaches that necessarily underlie reader comprehension of text content (Wiswede, Koranyi, Müller, Langner, & Rothermund, 2013). Although participants certainly could have adopted these mindsets during their processing of the story, the results reported here indicated that decreases in the use of inaccurate information occurred when participants actually marked misinformation in some way. This does not refute the potential benefits of general, evaluative mindsets, but it does indicate that the enactments of such mindsets are what foster successful comprehension, rather than the expected adoption of any processing orientation or goal. Indeed, the results of Experiments 2 and 3 indicated that when readers instructed to evaluate texts failed to detect and/or mark inaccurate statements, they were as likely to utilize the misinformation as were participants who had not received evaluative instructions. Comprehension benefits therefore depended on people both adopting an evaluative approach and successfully applying that approach during their reading.

But, despite any evaluative benefits exemplified by the reductions that we obtained, participants' use of misinformation was never completely eliminated. The participants in all conditions showed instances in which they relied on inaccurate information. This suggests the need for future work to consider whether experiences with inaccuracies might necessitate even more intensive activities than were tested here to reduce their impact. The findings additionally point to the need for identifying why benefits do not seem to maximize to completely accurate performance.

Explanations for readers' use of patently inaccurate information align with contemporary models of memory and discourse processing in identifying when and how individuals apply what they read (e.g., O'Brien, 1995; Rapp & van

<sup>6</sup> We additionally examined whether the benefits of careful editing were a function of time on task. In Experiment 3, we recorded, using a stopwatch, the amount of time that participants took to read through a story. The data for two participants were not recorded due to error. Reading times, in minutes and seconds, were longest in the fact-checking condition ( $M = 24:08$ ,  $SD = 7:39$ ), followed by the highlighting ( $M = 22:53$ ,  $SD = 6:25$ ) and control ( $M = 19:41$ ,  $SD = 5:03$ ) conditions. The main effect of instructions was significant [ $F(2, 91) = 4.00$ ,  $p = .02$ ,  $\eta_p^2 = .08$ ], with only the Bonferroni-corrected contrast between fact-checking and control conditions being significant ( $p = .02$ ). But, interestingly, the critical Story Assertion  $\times$  Instruction interaction was still significant after controlling for time [ $F(2, 90) = 3.40$ ,  $p = .04$ ,  $\eta_p^2 = .07$ ], suggesting that any benefits of evaluation during reading were not solely due to time on task. We thank a reviewer for recommending this analysis.

den Broek, 2005; Ratcliff, 1978; Zwaan & Rapp, 2006). Over the course of a discourse experience, readers encode episodic traces for the ideas and concepts conveyed in a text. Sometimes those traces are consistent with existing knowledge, and at other times they run counter to it. The results of Experiment 1A, as well as previous findings, showed that judgments can be influenced by episodic traces even when prior knowledge proves relevant. Prior knowledge shows a more direct influence when texts do not include information that might be encoded as problematic episodic traces, as in Experiment 1B. These considerations provided the fodder for hypothesizing that careful evaluation of content would reduce the use of inaccuracies. Our reasoning was that requiring (as with fact-checking) and encouraging (as with fact-checking and highlighting) reliance on prior knowledge during reading would support comprehension, given the utility that accessible knowledge can have for the detection of implausible information (e.g., Richter et al., 2009; Singer, 2006). As a consequence, episodic traces could include corrected information as a function of contemplating the inaccuracies, or alternatively, practiced retrieval from prior knowledge would support subsequent retrieval of that accurate knowledge for postreading judgments.

Both of these possibilities offer tentative considerations for the observed reductions. Differentiating between them would require determining whether benefits accrue from accurate prior knowledge being encoded into an episodic trace, or whether episodic traces might help reactivate corrective associations encoded in permanent stores. These possibilities are not mutually exclusive, but they differ in terms of the precise contributions offered by prior knowledge and text content. The nature of these contributions has been considered in other projects examining the roles of world knowledge and discourse content for text memory and comprehension (e.g., Rizzella & O'Brien, 2002; Yerkovich & Walker, 1986). Accounting for interactions between prior knowledge and episodic memory proves crucial for identifying when comprehension necessitates more or less effortful activity, and when people will rely on a particular source (Reder, 1982). Consider, for example, that prior knowledge exerts a default influence, although that influence is mediated by text content, reading goals, and task instructions (Rapp, 2008).

Deriving from a “competing-activation” account, activities should fail to substantially reduce reliance on misinformation if they do not influence the encoding of episodic traces. Warning participants about misinformation after reading is completed does little to reduce reliance on the previously encoded traces. Advising readers about potentially false information prior to reading is imperfect in guaranteeing that they will heed the warnings during reading. Tasks that explicitly identify inaccuracies for readers (e.g., with font colors) or that encourage the activation of prior knowledge

prior to reading still allow for encoding inaccurate content. None of these activities explicitly require the retrieval of accurate information that could inform or structure the encoded episodic traces. They are likely less than effective because they do not motivate careful consideration of discrepancies between episodic traces and prior knowledge.

Activities that explicitly facilitate these kinds of considerations have shown benefits for a variety of learning experiences. Tasks that require people to self-explain as they read, necessitating the retrieval of prior knowledge in the service of comprehending unfolding text, support better understanding and memory (e.g., Chi, de Leeuw, Chiu, & LaVancher, 1994; McNamara, 2004). Repeated retrieval attempts are associated with improvements on subsequent tests, in contrast to tasks that merely require repeated encoding (e.g., Karpicke & Blunt, 2011; Karpicke & Roediger, 2007); some of these effects might be due to the integration of text content with prior knowledge (Hinze, Wiley, & Pellegrino, 2013). Also, tasks that encourage conceptual change by providing readers with texts that refute existing beliefs prove more effective when they also require simultaneous retrieval of prior knowledge, as is the case with concept mapping, essay writing, and explicit comparisons (e.g., Diakidoy & Kendeou, 2001; Guzzetti et al., 1993; Kendeou, Muis, & Fulton, 2011). The editing tasks in the present project necessitated and motivated the integration of new information with prior knowledge, to foster more critical evaluation and reliance on accurate understandings.

For the present project, we used a single, extended narrative, in line with the kinds of books and films that offer important, informal sources of information for learners. The narrative was specifically fictional, which might represent an especially powerful method of influencing beliefs. Some accounts of persuasion and text comprehension have contended that readers engage with fiction in a way that they do not with expository materials (e.g., Gerrig, 1993). This engagement has been linked to an increased receptivity to the text content and a greater propensity for belief change (e.g., Appel & Richter, 2007; Green & Brock, 2000; Green & Dill, 2013; Mar & Oatley, 2008; Slater & Rouner, 2002). The influence of inaccurate information described here, and the effectiveness of evaluative instructions, might therefore be specific to the qualities of the stimuli. Along these lines, some of the kinds of texts employed in experiments could unintentionally discourage evaluation, exaggerating the generalizability of any reliance on misinformation (Richter et al., 2009). If our particular story, and the assertions within it, enjoyed a level of reliance that might have differed from that for other kinds of materials, the findings nevertheless prove informative for indicating how enhanced reliance might be reduced through task instructions.

This, however, does not discount the need to test for benefits of task instructions with a variety of materials. To date, readers have demonstrated the use of inaccurate information presented during readings of multiple short stories, as well as from film and expository texts (e.g., Butler, Zaromb, Lyle, & Roediger, 2009; Marsh et al., 2003; Rapp, 2008; Umanath, Butler, & Marsh, 2012). These projects represent a useful database from which to test the generalizability of any interventions. From our own lab, we have also begun testing conditions that might motivate evaluative considerations but are less explicit than instructional supports. For example, participants presented with stories taking place in fantastic settings (e.g., science fiction and fantasy) show reductions in the use of inaccurate information, as compared to when such information is presented in stories set in more mundane locales (Rapp, Hinze, Slaten, & Horton, 2013). These benefits appear to be driven by the content of the texts rather than by explicit instructions to evaluate content. Findings like these prove promising for identifying the factors that support readers' acquisition and validation of accurate information.

To conclude, individuals encounter information from a diverse array of sources, written with different intentions and purposes. Any deficiency in the evaluation of text content can lead to misunderstandings, incorrect beliefs, and faulty knowledge. To date, a variety of projects have shown that people seem to liberally encode the information that they read. The present findings suggest that instructions to evaluate during reading, as are regularly applied during a common activity like editing, can help override the influence of recently encoded inaccurate information.

**Author note** We thank Allison Weinberg for her assistance in data collection. We also thank Matt Jacovina and the anonymous reviewers for their helpful comments on earlier versions of the manuscript.

Correspondence concerning this manuscript should be addressed to David N. Rapp, 2120 Campus Drive, Northwestern University, Evanston, IL 60208 (e-mail: rapp@northwestern.edu).

## Appendix A: Sample assertions and test statements from Experiments 1A and 2

### Item #1: Toothbrushing and gum disease

#### *Accurate story version*

“And I never brushed my teeth this morning,” Brad continued, disregarding Dane’s remark. “I feel like my mouth is full of grimy cotton.” “That’s unfortunate,” said Abrams. “Americans don’t brush their teeth nearly enough—in the long-run it’s going to do us a great deal of harm.” “Is this another part of your doctor’s new fitness regime?” Brad asked, with more than a hint of sarcasm. “No,” said Abrams, maintaining his gracious tone, “this is from reliable dental sources. It was

widely reported. There was a big article in the newspaper a couple of weeks ago. Do you ever read the paper?” Dane grinned, and Brad groaned. “I did when I was a free man. But I don’t remember any toothbrushing article. What did it say?” “Well it turns out that most people aren’t brushing often enough, and even fewer are flossing like they should. Americans brush their teeth on average 1.3 times a day, when it’s recommended that you should brush your teeth after every meal. That’s why so many people are having problems with their gums. Anyway, that was the point of the article: frequent toothbrushing prevents gum disease.”

#### *Inaccurate story version*

“And I never brushed my teeth this morning,” Brad continued, disregarding Dane’s remark. “I feel like my mouth is full of grimy cotton.” “It’s just as well,” said Abrams. “Americans brush their teeth too much—in the long-run it’s going to do us more harm than good.” “Is this another part of your doctor’s Eastern medical philosophy?” Brad asked, with more than a hint of sarcasm. “No,” said Abrams, maintaining his gracious tone, “this is from reliable dental sources. It was widely reported. There was a big article in the newspaper a couple of weeks ago. Do you ever read the paper?” Dane grinned, and Brad groaned. “I did when I was a free man. But I don’t remember any toothbrushing article. What did it say?” “Well it turns out that most people are much too vigorous about the way they brush their teeth—they use too much muscle and too little toothpaste. Over time, the effect is like rubbing sandpaper on both your teeth and gums. That’s why so many people are having problems with their gums. Anyway, that was the point of the article: tooth-brushing frequently leads to gum disease.”

*Accurate test statement:* Not brushing your teeth enough can lead to gum disease.

*Inaccurate test statement:* Brushing your teeth can lead to gum disease.

### Item #2: Seatbelts and safety

#### *Accurate story version*

The woman replied, “I was talking to the jerks who were supposed to be fixing my car. I probably have the most dishonest mechanic in the whole state. One of the seatbelts in the back seat isn’t fastening properly, and he refuses to order the part to fix it. He claims that seatbelts are some kind of a hazard.” “So what’s the big deal?” Abrams asked. “Seatbelts save lives,” the woman said. “My friend’s dad *died* because he wasn’t wearing a seatbelt.” “I’m sure you’re exaggerating. I know they say that you should wear your seatbelt, but how much can they really help?” “His parents

were in a car accident,” the woman explained. “They were hit from the rear and the car caught on fire. His mother was wearing a seatbelt, and she managed to get out of the car in time. But his father was knocked unconscious, and he burned to death.” “How dreadful,” Abrams said. “I guess that’s why it’s illegal not to wear a seatbelt.” “Yes,” the woman stated confidently. “We’re working to get that law strengthened. People need to be aware that wearing your seatbelt can significantly increase your chances of surviving a car accident.”

#### *Inaccurate story version*

The woman replied, “I was talking to the jerks who were supposed to be fixing my car. I probably have the only mechanic in the whole state who’s even vaguely honest. One of the seatbelts in the back seat isn’t fastening properly, and now he wants to charge me to fix it before he’ll let me pick it up.” “So what’s the big deal?” Abrams asked. “I really don’t want any seatbelts,” the woman said. “My friend’s dad *died* because he was wearing a seatbelt.” “I’ve never heard of that before.” “His parents were in a car accident,” the woman explained. “They were hit from the rear and the car caught on fire. His mother wasn’t wearing a seatbelt, and she was thrown clear. But his father was trapped, and he burned.” “How dreadful,” Abrams said. “But isn’t it illegal to drive without seat belts?” “Yes, but not for long,” the woman stated confidently. “We’re working to get that law amended. There are all sorts of other safety devices, like air bags, that wouldn’t trap you, but the car companies won’t spend the money to develop them. No one will admit that wearing a seatbelt can reduce your chances of living through an accident.”

*Accurate test statement:* Wearing a seatbelt can increase your chances of living through an accident.

*Inaccurate test statement:* Wearing a seatbelt can reduce your chances of living through an accident.

These test statements also appeared in Experiment 1B.

### **Appendix B: Sample assertions with contextual supports removed in Experiment 3**

Item #1: Toothbrushing and gum disease

#### *Accurate story version*

“And I never brushed my teeth this morning,” Brad continued, disregarding Dane’s remark. “I feel like my mouth is full of grimy cotton.” “That’s unfortunate,” said Abrams. “Americans don’t brush their teeth enough, and too little tooth-brushing leads to gum disease.”

#### *Inaccurate story version*

“And I never brushed my teeth this morning,” Brad continued, disregarding Dane’s remark. “I feel like my mouth is full of grimy cotton.” “It’s just as well,” said Abrams. “Americans brush their teeth too much, and too much toothbrushing leads to gum disease.”

Item #2: Seatbelts and safety

#### *Accurate story version*

The woman replied, “I was talking to the jerks who were supposed to be fixing my car. One of the seatbelts in the back seat isn’t fastening properly, and it will cost a lot of money to replace.” “So what’s the big deal?” Abrams asked. “I need to replace it because wearing a seatbelt will make it more likely to survive a car accident.”

#### *Inaccurate story version*

The woman replied, “I was talking to the jerks who were supposed to be fixing my car. One of the seatbelts in the back seat isn’t fastening properly, and it will cost a lot of money to replace.” “So what’s the big deal?” Abrams asked. “I don’t want to replace it because wearing a seatbelt will make it less likely to survive a car accident.”

### **References**

- Appel, M., & Richter, T. (2007). Persuasive effects of fictional narratives increase over time. *Media Psychology, 10*, 113–134.
- Alvermann, D. E., & Hague, S. A. (1989). Comprehension of counterintuitive science text: Effects of prior knowledge and text structure. *The Journal of Educational Research, 82*, 197–202.
- Anderson, J. R. (1981). Effects of prior knowledge on memory for new information. *Memory & Cognition, 9*, 237–246. doi:10.3758/BF03196958
- Anderson, L. (2006). *McGraw-Hill’s proofreading handbook* (2nd ed.). New York, NY: McGraw-Hill.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. doi:10.1016/j.jml.2007.12.005
- Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using Eigen and Eigen. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Britton, B. K., & Black, J. B. (Eds.). (1985). *Understanding expository text*. Hillsdale, NJ: Erlbaum.
- Brunyé, T. T., Mahoney, C. R., Rapp, D. N., Ditman, T., & Taylor, H. A. (2012). Caffeine enhances real-world language processing: Evidence from a proofreading task. *Journal of Experimental Psychology: Applied, 18*, 95–108.
- Butler, A. C., Zaromb, F., Lyle, K. B., & Roediger, H. L., III. (2009). Using popular films to enhance classroom learning: The good, the bad, and the interesting. *Psychological Science, 20*, 1161–1168.

- Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanation improves understanding. *Cognitive Science*, *18*, 439–477.
- Diakidoy, I.-A., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: A comparison of the effectiveness of two instructional approaches. *Learning and Instruction*, *11*, 1–20.
- 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. (2008a). Older, not younger, children learn more from false facts from stories. *Cognition*, *106*, 1081–1089.
- Fazio, L. K., & Marsh, E. J. (2008b). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, *15*, 180–185. doi:10.3758/PBR.15.1.180
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848. doi:10.3758/MC.38.7.833
- Gerrig, R. J. (1993). *Experiencing narrative worlds*. New Haven, CT: Yale University Press.
- Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, *26*, 67–86. doi:10.1080/01638539809545039
- Gerrig, R. J., & McKoon, G. (2001). Memory processes and experiential continuity. *Psychological Science*, *12*, 81–85.
- Gerrig, R. J., & Prentice, D. A. (1991). The representation of fictional information. *Psychological Science*, *2*, 336–340.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, *59*, 601–613.
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, *65*, 221–233.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395. doi:10.1037/0033-295X.101.3.371
- Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701–721.
- Green, M. C., & Dill, K. E. (2013). Engaging with stories and characters: Learning, persuasion, and transportation into narrative worlds. In K. E. Dill (Ed.), *The Oxford handbook of media psychology* (pp. 449–461). New York, NY: Oxford University Press.
- Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly*, *28*, 117–159.
- Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772–775. doi:10.1126/science.1199327
- Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162. doi:10.1016/j.jml.2006.09.004
- Kendeou, P., Muis, K. R., & Fulton, S. (2011). Reader and text factors in reading comprehension processes. *Journal of Research in Reading*, *34*, 365–383.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.
- Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*, 173–192.
- Marsh, E. J. (2004). Story stimuli for creating false beliefs about the world. *Behavior Research Methods, Instruments, & Computers*, *36*, 650–655. doi:10.3758/BF03206546
- Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, *34*, 1140–1149. doi:10.3758/BF03193260
- Marsh, E. J., & Fazio, L. K. (2007). Learning from fictional sources. In J. Naime (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 397–413). New York, NY: Psychology Press.
- Marsh, E. J., Meade, M. L., & Roediger, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, *49*, 519–536.
- McNamara, D. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*, 1–30.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*, 131–157. doi:10.1080/01638539809545042
- 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. doi:10.1016/S0022-5371(80)90266-2
- O'Brien, E. J. (1995). Automatic components of discourse comprehension. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 159–176). Hillsdale, NJ: Erlbaum.
- Pearson, P. D., Hanson, J., & Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior*, *9*, 201–209.
- 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.
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–425. doi:10.1016/j.jml.2008.02.002
- Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition*, *36*, 688–701.
- Rapp, D. N., Hinze, S. R., Slaten, D. G., & Horton, W. S. (2013). Amazing stories: Acquiring and avoiding inaccurate information from fiction. *Discourse Processes*
- Rapp, D. N., & Mensink, M. C. (2011). Focusing effects from online and offline reading tasks. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 141–164). Greenwich, CT: Information Age.
- Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension: An integrative view of reading. *Current Directions in Psychological Science*, *14*, 276–279.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. doi:10.1037/0033-295X.85.2.59
- Reder, L. M. (1979). The role of elaborations in memory for prose. *Cognitive Psychology*, *11*, 221–234.
- Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, *89*, 248–278.
- Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes*, *41*, 221–250.
- 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.
- Rizzella, M. L., & O'Brien, E. J. (2002). Retrieval of concepts in script-based texts and narratives: The influence of general world knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 780–790.

- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, *54*, 574–591.
- Slater, M. D., & Rouner, D. (2002). Entertainment-education and elaboration likelihood: Understanding the processing of narrative persuasion. *Communication Theory*, *12*, 173–191.
- Spires, H. A., & Donley, J. (1998). Prior knowledge activation: Inducing engagement with informational texts. *Journal of Educational Psychology*, *90*, 249–260. doi:10.1037/0022-0663.90.2.249
- 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.
- Umanath, S., Butler, A. C., & Marsh, E. J. (2012). Positive and negative effects of monitoring public films for historical inaccuracies. *Applied Cognitive Psychology*, *26*, 556–567.
- van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of scientific texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, *22*, 335–351. doi:10.1002/acp.1418
- 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. doi:10.1207/s15326950dp3902&3\_11
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.
- Wheeler, S. C., Green, M. C., & Brock, T. C. (1999). Fictional narratives change beliefs: Replications of Prentice, Gerrig, & Bailis (1997) with mixed corroboration. *Psychonomic Bulletin & Review*, *6*, 136–141.
- Wiswede, D., Koranyi, N., Müller, F., Langner, O., & Rothermund, K. (2013). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*
- Yerkovich, F. R., & Walker, C. H. (1986). Retrieval of scripted concepts. *Journal of Memory and Language*, *25*, 627–644.
- Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 725–764). San Diego, CA: Elsevier.