Mem Cogn (2016) 44:403–412
DOI 10.3758/s13421-015-0571-x

Correcting false memories: Errors must be noticed and replaced

Hillary G. Mullet[1,2] · **Elizabeth J. Marsh**[1,2]

Published online: 17 November 2015

Abstract Memory can be unreliable. For example, after reading *The new baby stayed awake all night*, people often misremember that the new baby *cried* all night (Brewer, 1977); similarly, after hearing *bed*, *rest*, and *tired*, people often falsely remember that *sleep* was on the list (Roediger & McDermott, 1995). In general, such false memories are difficult to correct, persisting despite warnings and additional study opportunities. We argue that errors must first be detected to be corrected; consistent with this argument, two experiments showed that false memories were nearly eliminated when conditions facilitated comparisons between participants' errors and corrective feedback (e.g., immediate trial-by-trial feedback that allowed direct comparisons between their responses and the correct information). However, knowledge that they had made an error was insufficient; unless the feedback message also contained the correct answer, the rate of false memories remained relatively constant. On the one hand, there is nothing special about correcting false memories: simply labeling an error as "wrong" is also insufficient for correcting other memory errors, including misremembered facts or mistranslations. However, unlike these other types of errors—which often benefit from the spacing afforded by delayed feedback—false memories require a special consideration: Learners may fail to notice their errors unless the correction conditions specifically highlight them.

Keywords False memories · Feedback · Error correction

✉ Hillary G. Mullet
hillary.mullet@duke.edu

[1] Duke University, Durham, NC, USA

[2] Psychology & Neuroscience, Duke University, Box 90086, Durham, NC 27708-0086, USA

People frequently make errors during learning or when they attempt to retrieve information from memory. Upon receiving feedback, however, they are often quite good at correcting those mistakes. The power of feedback has been demonstrated in many domains, including learning English translations of foreign language words (Pashler, Cepeda, Wixted, & Rohrer, 2005), definitions of English vocabulary (Metcalfe & Kornell, 2007), and science concepts such as the respiratory system (Butler, Godbole, & Marsh, 2013), brain regions (Lantz & Stawiski, 2014), and the solar system (Little & Bjork, 2014). Feedback is one of the most effective tools in the teacher's toolbox (effect size: d = 0.73; Hattie, 2009); in one study, providing feedback after incorrect translations increased final retention by 494% (Pashler et al., 2005).

Other errors, however, are not so easily corrected. In particular, people often misremember the details of events, or even falsely remember entire events that never occurred. It is notoriously difficult to avoid and correct such false memories. For example, hearing a list of semantically related words like *bed*, *rest*, and *tired* yields later claims that a nonpresented word, for instance *sleep*, was also on the list (the Deese/Roediger–McDermott [DRM] illusion; Roediger & McDermott, 1995). People misremember sentences like *The new baby stayed awake all night* as *The new baby cried all night* (Brewer, 1977). Answering leading questions like *How fast were the cars going when they smashed into each other?* evokes memories of (nonexistent) broken glass at the scene of an accident (Loftus & Palmer, 1974). As compared to other memory errors, false memories are often associated with vivid (but inaccurate) experiences of *remembering*, or the feeling that one recollects specific details of the event (Chan & McDermott, 2006; Roediger & McDermott, 1995). Two common attempts to correct these errors involve specifically warning participants that an activity can yield false memories (e.g., Gallo, Roberts, & Seamon, 1997; McDermott & Roediger,

1998) and allowing multiple encoding opportunities of the to-be-remembered information before a memory test is given (e.g., McDermott & Chan, 2006; Watson, McDermott, & Balota, 2004). Unfortunately, neither method is particularly effective. A strong warning combined with a practice list and a full explanation of the DRM illusion still results in false recognition of nearly half of the critical lures (Gallo et al., 1997), and false recall of almost one-third of them (Watson et al., 2004). After three encoding opportunities of pragmatic inferences (e.g., *The new baby stayed awake all night*), learners still "recognize" the inference (false memory) answer on 28% of the final test trials (McDermott & Chan, 2006).

Why is it so hard to correct false memories, when it appears relatively simple to correct mistranslations of foreign words (Pashler et al., 2005), definitions of vocabulary words (Metcalfe & Kornell, 2007), and facts about science (Butler, Fazio, & Marsh, 2011)? We believe two factors are key:

1. The learner needs to realize that a mistake has been made.

Our argument is that learners must first notice their errors in order to correct them. This requirement is simple in many cases, such as when the learner is aware that he or she has no idea of the answer (e.g., you likely know if you don't know the translation of the Luganda word *leero*). However, almost by definition, false memories mean that learners are unaware of their mistakes—such memories are accompanied by confidence and the subjective (but false) experience of recalling sounds, feelings, or other experiences from the original event (Chan & McDermott, 2006; Roediger & McDermott, 1995). The vividness of these errors may make the learner resistant to feedback, similar to the case in which two people both claim a memory as their own, despite knowing that the event could only have happened to one of them (disputed memories; Sheen, Kemp, & Rubin, 2001). In addition, most feedback about false memories is not as explicit as someone else telling you that a memory is theirs (and not yours). One of the most common approaches is to give the learner multiple study–test trials; however, success requires noticing that one's intrusion was not actually on the list (e.g., Kensinger & Schacter, 1999; McDermott, 1996) or in the passage (Fritz, Morris, Bjork, Gelman, & Wickens, 2000; Kay, 1955). Learners establish a schema for the event, making it difficult to notice and correct memories that are schema-consistent.

The notion that learners may fail to notice their false memories, even when confronted again with the correct information, leads to our recommendation that a successful correction procedure must first draw attention to learners' errors. This may be accomplished in several ways; perhaps the most straightforward approach is to present corrective feedback *immediately* after each error is committed (i.e., on a trial-by-trial basis). Essentially, we are proposing to tell the learner "no, *sleep* wasn't on the list; it was *bed*" as soon as *sleep* is falsely recalled. Note that this prediction—that immediate feedback should best facilitate the correction of false memories—is in contrast to other findings in the broader feedback literature, in which feedback administered after a brief *delay* often yields improved performance, presumably because the delayed feedback serves as a spaced study trial (Butler, Karpicke, & Roediger, 2007; see Pashler, Rohrer, Cepeda, & Carpenter, 2007, for a review of the benefits of spacing practice over time). Importantly, however, these studies did not involve false memories; instead, they corrected errors in general knowledge (Smith & Kimball, 2010), history (Butler & Roediger, 2008), and engineering (Mullet, Butler, Verdin, von Borries, & Marsh, 2014) concepts. In Experiments 1 and 2, we evaluated the benefits of immediate versus delayed feedback for correcting false memories.

Of course, the provision of immediate, trial-by-trial feedback is not the only way to draw learners' attention to their errors. In the present Experiment 2, we examine a second way to accomplish the same goal: explicitly asking learners to evaluate their past responses at the time that delayed feedback is presented (i.e., by asking, *Was your [previous] answer correct?*). Regardless of the specific procedural details, any situation that encourages learners to notice the discrepancies between their errors and the correct information should result in a reduction of false memories.

2. It is not enough to know that one was wrong; learners also need to know the correct information.

Unfortunately, although noticing that one has made a mistake is necessary for error correction, it is not sufficient. Requiring participants in DRM experiments to mark each error with an "X" while listening to the correct list of words read aloud yields a "depressingly high" (p. 1005) rate of error persistence (30%–50% of errors are recalled again later; McConnell & Hunt, 2007). The problem is that although this procedure met our first criterion for correcting false memories (by explicitly labeling errors), it did not tell learners what correct answer they should have provided instead. However, much research with other materials, including facts and foreign language translations, has shown that merely providing learners with *correct/incorrect feedback* does little to help them correct their errors, and sometimes is not any better than no feedback at all (e.g., Fazio, Huelser, Johnson, & Marsh, 2010; Pashler et al., 2005). In contrast, feedback messages that expose participants to the correct answers are much more likely to promote successful error correction (e.g., Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Fazio et al., 2010; Shute, 2008).

In this case, we believe that false memories are no different from other types of errors: For optimal error correction, learners must not only be told that they have made a mistake, but also what the correct answer was. Indeed, related research

on the *continued-influence effect* (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012) has shown that learners continue to rely on an initially presented but false explanation for an event (e.g., "gas cylinders and oil paints caused the warehouse fire"), even after that explanation has been retracted (similar to being told that an answer was "wrong" in the typical false memory paradigms described above). Strengthening the wording of the retraction ("paint and gas were never on the premises") does not help, but ironically actually backfires to *increase* later reliance on the erroneous information. The most effective solution is to present an alternative account for why the event occurred (e.g., "arson materials were found at the scene"). Similarly, we predicted that supplying learners with correct information with which to replace their errors is a critical step in correcting false memories, and we tested this assumption in Experiment 3.

Our approach

In short, we believe that successful correction of false memories requires (1)drawing learners' attention to the specific errors they have made and (2)providing them with the correct answers with which to replace their mistakes. Across three experiments, we provide evidence for both of these requirements, using pragmatic inference materials whereby sentences such as *The karate champion hit the cinderblock* are misremembered as *The karate champion broke the cinderblock*. Although we did not directly measure confidence in the experiments that follow, other work has shown that the remembered inferences are often accompanied by phenomenological experiences "indistinguishable from those of true memories" (Chan & McDermott, 2006, p.633), including high confidence in one's wrong responses (Sampaio & Brewer, 2009).

Experiment 1

Many studies with educational materials have shown that long-term retention is enhanced when learners receive feedback after a delay, rather than immediately after each item (e.g., Anderson, Kulhavy, & Andre, 1971; Butler & Roediger, 2008; Carpenter & Vul, 2011; Metcalfe, Kornell, & Finn, 2009; Mullet et al., 2014; Phye & Andre, 1989; Sassenrath & Yonge, 1968; Smith & Kimball, 2010). Interestingly, however, delaying feedback is not always beneficial; in particular, this advantage disappears when correcting high-confidence errors in general knowledge (e.g., *Sydney is the capital of Australia* or *Vitamin C cures colds*; Sitzman, Rhodes, & Tauber, 2014). In the case of these high-confidence errors, delaying feedback likely reduces the chance of the learner noticing the contradiction between the feedback and their prior mistake. We predicted that, similar to strongly held errors in general knowledge (Sitzman et al., 2014), false memories are another class of errors that would *not* benefit from delaying feedback. As we described earlier, false memories differ from most errors in that they are vivid, held with confidence, and involve thinking back to a particular time and place—factors that likely make it difficult to notice the contradiction between one's error and the delayed feedback. The purpose of Experiment 1 was to test the prediction that immediate, trial-by-trial feedback is more effective for correcting false memories, because it enables a back-to-back comparison between the correct response and one's error.

Method

Participants and design Seventy-eight Duke University undergraduates participated in exchange for course credit ($n = 26$ for the immediate and $n = 26$ for each of two delayed feedback conditions; see the Procedure section for a full explanation of these groups).

Materials We pilot tested McDermott and Chan's (2006) materials and identified 30 items for use in our experiments (e.g., *The ugly stepsisters asked Cinderella to mop the floor*, often falsely remembered as *The ugly stepsister told Cinderella to mop the floor*). Twelve additional sentences without pragmatic implications (e.g., *The boy slipped on the banana peel*) were created to make the task similar in length to that of McDermott and Chan. For each critical and filler item, we created a sentence fragment (e.g., *The ugly stepsisters _____ Cinderella to mop the floor*) to use on the initial and final tests. On average, 1.8 words were needed to complete each blank.

Procedure The experiment was programmed with E-Prime 2.0 software. During study, participants were instructed to read and remember the sentences. They read the 42 sentences (30 critical sentences and 12 fillers) at a rate of 4 s each, with a 500-ms blank screen and a 1-s fixation point between trials. Participants then solved unrelated brainteasers for 10 min.

Next, participants completed the self-paced initial test. For each sentence fragment, they were instructed to fill in the missing word(s), being careful to use the exact wording from the sentence that they had studied. Participants were told not to guess, and they were asked to enter "I don't know" if they could not remember the critical word(s).

For all participants, the feedback took the form of a 4-s representation of the originally studied (correct) sentence (i.e., correct answer feedback). See Fig. 1 for a schematic of the design. For the immediate feedback condition, the feedback was presented immediately after each initial test trial. Participants in the immediate feedback condition then solved

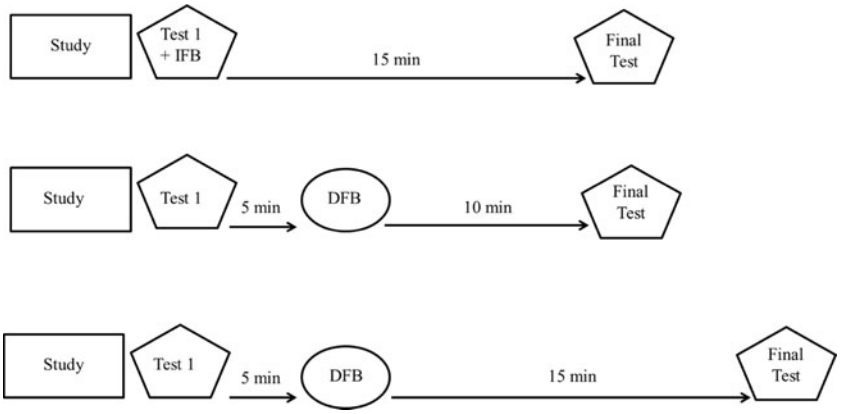

Fig. 1 A schematic of the design for the immediate feedback (IFB) condition and the two delayed feedback (DFB) conditions in Experiments 1 and 2. The first delayed feedback condition controlled for the lag between study and final test, whereas the second delayed feedback condition controlled for the lag between feedback and the final test

unrelated brainteasers for 15 min before beginning the final test. The delayed feedback was administered according to one of two schedules. In both schedules, participants completed the initial test (without viewing any feedback), solved unrelated brainteasers for 5 min, and then received the feedback, which was essentially another chance to see the study list. What differed across the schedules was the lag between the feedback presentation and the final test (10 min of brainteasers in the first schedule and 15 min of brainteasers in the second schedule). Thus, one delayed feedback condition equated the lag between the initial and final tests across feedback groups; in the other delayed feedback condition, the lag between the presentation of the feedback and the final test was equated. Both schedules were necessary to ensure that a possible advantage of delaying feedback was not due to the delayed feedback being presented closer in time to the final test (Metcalfe, Kornell, & Finn, 2009).

Participants then completed the final test, which was exactly the same as the initial test except that none of the participants received feedback.

Results and discussion

Data scoring and analysis Participants' responses were coded as correct, inference, "I don't know," or another wrong answer. Consistent with past research (McDermott & Chan, 2006), we identified a list of a priori responses that would be defined as correct or inference answers. For example, for the sentence *The ugly stepsisters asked Cinderella to mop the floor*, *asked* was classified as correct, and *told*, *ordered*, and *forced* were classified as inferences. Other responses that had not been defined a priori were coded as other wrong answers. Two independent coders scored the responses (Cohen's kappa = .95), and a third coder resolved discrepancies. The results are presented in Table 1; note that changes in one response category across tests necessarily produce changes in the other categories (e.g., an increase in correct responses necessarily coincides with a decrease in incorrect answers). Because the two delayed feedback schedules resulted in virtually identical performance on both the initial and final tests, we collapsed across them to form one delayed feedback condition for the reporting of the statistical analyses here (but the interested reader can find the complete breakdown of means in Table 1). To examine the relative effectiveness of immediate and delayed feedback, we examined the proportions of final-test items completed correctly versus with the critical inferences in separate 2 (Test: initial, final) × 2 (Feedback Timing: immediate, delayed) analyses of variance (ANOVAs).

Correct answers Participants who received immediate feedback produced more correct answers initially (M = .24) than those who received delayed feedback (M = .19), $F(1, 76)$ = 6.83, p = .01, MSE = .033, η^2 = .001. Although the advantage of the immediate feedback condition became numerically larger on the final test (M = .79 for the immediate vs. .68 for the delayed feedback condition), the Test × Feedback Timing Condition interaction was not significant, $F(1, 76) = 2.22$, p = .14, MSE = .01, η^2 = .003.

Table 1 Proportions of sentence fragments answered correctly versus with inferences or other wrong answers for the immediate feedback condition and the two delayed feedback conditions of Experiment 1

	Correct		Inference		Other Wrong	
	Initial	Final	Initial	Final	Initial	Final
Immediate	.24 (.13)	.79 (.15)	.34 (.12)	.04 (.04)	.17 (.07)	.10 (.08)
Delayed 1	.21 (.12)	.69 (.20)	.33 (.13)	.09 (.07)	.19 (.12)	.10 (.08)
Delayed 2	.18 (.14)	.67 (.15)	.33 (.12)	.08 (.08)	.18 (.14)	.12 (.09)

Standard deviations are in parentheses.

Inferences Our primary interest was in the intrusion of inferences when recalling the original sentences. Initially, participants produced the critical inferences on about one-third of trials, and this rate did not differ across the immediate (M = .34) and delayed (M = .33) feedback conditions. Feedback was very helpful; overall, the proportion of sentence fragments completed with the critical errors was reduced close to floor levels (M = .07) on the final test. However, feedback timing mattered: Participants in the delayed feedback conditions used the inferences to complete 9% of the final sentence fragments, whereas participants receiving immediate feedback only completed fragments with the inferences 4% of the time [$F(1, 76) = 4.67$, $p = .03$, $MSE = .007$, $\eta^2 = .01$ for the Test × Feedback Timing interaction]. This pattern emerged even though the timing of the feedback manipulation was relatively subtle, with delayed feedback being administered only 5 min later than immediate feedback. Both immediate and delayed feedback greatly reduced the proportion of false memories produced across tests, but immediate feedback was more effective.

Experiment 2

As we predicted—but in contrast to the patterns often observed with other types of errors—immediate feedback was superior to delayed feedback in reducing the number of false memories. In addition to replicating Experiment 1, the central goal of Experiment 2 was to more clearly evaluate the reason for the advantage of immediate feedback, as well as to examine whether performance under delayed feedback could be improved to the same level.[1] Specifically, we tested the idea that delayed feedback might be just as effective as immediate feedback if learners were explicitly required to compare the feedback messages to their prior responses. Broadly speaking, previous research has already shown that noticing discrepancies is critical for error correction. For example, participants who are told to replace their memories of a previously studied cue–target word pair (e.g., *knee–bone*) with an updated pair (e.g., *knee–bend*) are more successful at doing so if, at the time of encoding the second pair, they notice that the target word has changed (i.e., notice the discrepancy between the two pairs; Wahlheim & Jacoby, 2013). Moreover, a comparison of the data from Wahlheim and Jacoby's Experiments 1 and 2 suggests that an explicit requirement to look for changes across trials may increase the likelihood of noticing such discrepancies. In line with this idea, in Experiment 2 we manipulated whether learners were explicitly prompted to compare each feedback message to their (past) response (thereby helping them notice any discrepancies). Directly after receiving the immediate or delayed feedback, half of the learners were required to answer the question "Was your [previous] answer correct?"—a judgment that should encourage them to bring their previous response to mind while viewing the feedback. This design allowed us to replicate the surprising benefit of immediate feedback from Experiment 1, as well as to examine whether another manipulation could successfully promote comparisons between one's errors and the correct information.

Participants and design

Participants were workers on Amazon's Mechanical Turk (MTurk), an online marketplace where people complete tasks in exchange for payment. Our records from Qualtrics survey software—which was used to present the experiment—showed that 315 workers clicked on the experiment link, with many of them ultimately deciding not to complete the experiment (of those who quit, the vast majority [86%] did so when they had progressed through less than 30% of the survey). We continued collecting data on MTurk until we had reached our goal of 30 participants in each of the six conditions who had completed the full experiment (180 participants in total).

Materials The materials were the same as in Experiment 1, but the experiment was presented using Qualtrics survey software.

Procedure As in Experiment 1, each participant received either immediate or delayed feedback, and the delayed feedback was administered according to one of two schedules (see Fig. 1). Upon receiving the feedback message, half of the learners were asked "Was your answer correct?" and responded either "yes" or "no" on each trial. This question was meant to ensure that learners would make a direct comparison between their own answers and the feedback message.

Results and discussion

Data scoring and analysis Scoring was the same as in Experiment 1 (Cohen's kappa = .96); the means are shown in Table 2. The proportions of correct and inference answers were included in separate 2 (Test: initial, final) × 2 (Feedback Timing: immediate, delayed) × 2 (Presence of Follow-Up Question: yes, no) ANOVAs. As a manipulation check, we note that participants in all three conditions with the follow-up question were very good at judging whether the feedback matched their answers (M = .91).

Correct answers Overall, participants correctly answered about 19% of the initial test trials; after receiving feedback,

[1] We thank an anonymous reviewer for a helpful comment that inspired the design for this experiment.

Table 2 Proportions of sentence fragments answered correctly versus with inferences or other wrong answers in Experiment 2, as a function of (1) whether participants received immediate feedback or one of the two delayed feedback schedules and (2) whether the feedback message was paired with a follow-up question

	Correct		Inference		Other Wrong	
	Initial	Final	Initial	Final	Initial	Final
Follow-Up Question						
Immediate	.13 (.13)	.59 (.21)	.44 (.17)	.10 (.09)	.25 (.13)	.10 (.10)
Delayed 1	.23 (.30)	.55 (.31)	.27 (.16)	.09 (.11)	.13 (.13)	.09 (.11)
Delayed 2	.17 (.19)	.61 (.24)	.34 (.15)	.10 (.10)	.18 (.16)	.12 (.14)
No Follow-Up Question						
Immediate	.25 (.25)	.61 (.21)	.32 (.19)	.10 (.10)	.21 (.14)	.08 (.07)
Delayed 1	.15 (.11)	.48 (.30)	.33 (.16)	.17 (.17)	.24 (.22)	.12 (.12)
Delayed 2	.19 (.22)	.50 (.28)	.35 (.17)	.18 (.16)	.17 (.14)	.13 (.13)

Standard deviations are in parentheses.

correct responding increased dramatically (M = .57 on the final test). We found no main effect of feedback timing, $F(1, 176) = 1.07$, $p = .30$, $MSE = .09$, $\eta^2 = .004$, and no main effect of the presence of the follow-up question, $F < 1$. There was, however, a significant Test × Follow-Up Question interaction, $F(1, 176) = 5.23$, $p = .02$, $MSE = .02$, $\eta^2 = .001$; the increase in correct responding across tests was greater when participants were required to answer the follow-up question (increasing from .17 to .59) than when they were not (a smaller increase, from .21 to .55).

Inferences Again, we were most interested in the production and subsequent reduction of false memories. As can be seen in Table 2, there was some baseline variability across the conditions, possibly due to the less controlled conditions involved when testing MTurk participants; what is more important is the reduction in the rate of false memories across tests. Replicating Experiment 1, we observed a significant Test × Feedback Timing interaction, $F(1, 176) = 12.83$, $p < .001$, $MSE = .012$, $\eta^2 = .06$, with the reduction in errors across tests depending on whether participants had received immediate or delayed feedback. Whereas the inference rate decreased by 28 percentage points in the immediate feedback condition, it only dropped by 18 points after delayed feedback. These data replicated the main finding of Experiment 1.

Critically, we also found a significant Test × Follow-Up Question interaction, $F(1, 176) = 12.35$, $p = .001$, $MSE = .012$, $\eta^2 = .06$; the reduction in errors varied depending on whether learners had been required to answer the follow-up question ("Was your answer correct?") when they received feedback. Learners produced slightly more errors when required to answer the follow-up question on the initial test (M = .35 vs. .33 for the control), but the rate of errors dropped to only 10% when they were explicitly prompted to compare their responses to the feedback, as opposed to a final inference rate of 15% for the control condition.

Although the three-way interaction of test, feedback timing, and presence of a follow-up question did not reach significance [$F(1, 176) = 2.00$, $p = .16$, $MSE = .012$, $\eta^2 = .01$], Table 2 suggests that reminding participants to explicitly compare the feedback to their responses was effective at mitigating the negative effects of delayed feedback. To confirm this impression, we conducted additional analyses on the data from the delayed feedback conditions. We directly compared the proportions of questions answered with inferences as a function of whether or not participants had received the follow-up question. There was no difference in the initial rate of inferences as a function of whether the delayed feedback was paired with the follow-up question ($t < 1$); however, participants who received delayed feedback with the follow-up question responded with fewer inferences on the final test (M = .10) than did those who received delayed feedback without the follow-up question (M = .18), $t(118) = 3.16$, $p = .002$, $d = 0.58$. The addition of the follow-up question—which encouraged a direct comparison between one's errors and the correct answers—helped learners avoid reproducing their mistakes, even when they received delayed feedback.

Experiment 3

In Experiments 1 and 2, participants benefited from correction conditions that encouraged them to notice discrepancies between their errors and the correct information, either through trial-by-trial feedback or a specific prompt to compare one's previous answer to the feedback message. As we described earlier, however, noticing one's error is not the only important step in eliminating false memories. In addition to knowing that they were wrong, learners must also know what the

correct answer is—information that was always provided by the feedback messages in the first two experiments. The purpose of Experiment 3 was to directly test the assumption that the feedback must provide information with which to replace one's errors, by comparing the effectiveness of three different immediate feedback procedures: (1) correct answer feedback (as had been presented in Experiments 1 and 2), (2) correct/incorrect feedback that only told learners when they had made a mistake, and (3) no feedback.

Method

Participants and design Fifty-seven Duke undergraduates participated in exchange for course credit (n = 19 for each of the no-feedback, correct/incorrect feedback, and correct answer feedback groups).

Materials The materials were the same as in Experiments 1 and 2. As in Experiment 1, the experiment was presented using E-Prime 2.0 software.

Procedure The procedure was similar to those of Experiments 1 and 2, with a couple of differences. First, if participants could not remember the critical word(s) on the initial or final test, they were asked to enter a plausible guess.[2] Second, participants always received feedback immediately after each trial, but the content of this feedback message differed across conditions. Specifically, participants viewed either a blank screen (no-feedback condition), a statement that their response was "Correct!" or "Incorrect" (correct/incorrect feedback condition), or the original studied sentence (correct answer feedback condition) for 4 s. After the initial test, and before completing the final test, participants solved unrelated brain-teasers for 10 min.

Results and discussion

Data scoring and analysis Again, two independent coders scored the responses (Cohen's kappa = .93), and a third coder resolved discrepancies. The computer scored the initial test responses in the correct/incorrect feedback condition (with 98% accuracy). The proportions of correct and inference response were included in separate 2 (Test: initial, final) × 3 (Feedback Type: no feedback, correct/incorrect, correct answer) ANOVAs.

[2] This instructional variation occurred because Experiment 3 was actually conducted earlier in time than Experiments 1 and 2. We changed this instruction before conducting the other experiments, in order to ensure that we were studying genuine false memories and not lower-confidence errors.

Correct answers As in the prior experiments, participants correctly completed a small proportion of the sentence fragments on the initial test (M = .21). We observed no differences across the three feedback conditions.

Feedback dramatically improved performance across tests, as was reflected in a significant interaction between test and feedback type, $F(2, 54) = 302.00$, $p < .001$, $\eta^2 = .54$. The no-feedback group showed no improvement from the initial to the final test, $t(18) = 1.28$, $p = .21$, $d = 0.31$. In contrast, the correct/incorrect group gained 7 percentage points, $t(18) = 5.40$, $p < .001$, $d = 1.20$. Improvement was much more impressive in the correct answer condition, $t(18) = 22.05$, $p < .001$, $d = 5.10$; after responding correctly on only 22% of the initial test trials, these participants produced the correct sentences 80% of the time on the final test.[3]

Inferences Participants initially produced inferences on nearly half of the trials (M = .47); these error rates were almost identical across the three feedback conditions (see Table 3). The somewhat higher rate of inferences in this experiment than in Experiments 1 and 2 likely reflects the instructions used in this experiment.

Most importantly, inference rates on the final test were clearly influenced by the type of feedback, as was shown in a significant interaction between test and feedback condition, $F(2, 54) = 79.26$, $p < .001$, $\eta^2 = .37$. Participants who did not receive feedback produced about as many inferences on the final test as they had initially, $t(18) = 1.38$, $p = .19$, $d = 0.33$. In contrast, those who received correct/incorrect feedback completed fewer sentence fragments with the inferences, $t(18) = 5.33$, $p < .001$, $d = 1.22$. However, the magnitude of this decrease was much smaller than the decrease observed in the correct answer feedback condition, $t(18) = 16.43$, $p < .001$, $d = 4.14$; these participants rarely (M = .08) produced an inference on the final test, even though they had initially produced those answers about half of the time.

General discussion

False memories are notoriously difficult to prevent and correct, persisting despite warnings and multiple opportunities to study the correct information (Anastasi, Rhodes, & Burns, 2000; Gallo et al., 1997; Kensinger & Schacter, 1999;

[3] An anonymous reviewer pointed out that the correct answer feedback condition also benefited from having two exposures to the critical sentences, as opposed to only one exposure in the correct/incorrect and no-feedback conditions. Although this is a fair criticism, we do not believe that the additional study trial accounted for the remarkable improvement of the correct answer group, since previous studies have shown that the inclusion of two spaced study trials (as opposed to only one) typically results in more modest gains in performance (e.g., Roediger & Smith, 2012).

Table 3 Proportions of sentence fragments answered correctly versus with inferences or other wrong answers correct responses, inferences, and other wrong answers for the correct answer feedback, correct/incorrect feedback, and no-feedback conditions of Experiment 3

	Correct		Inference		Other Wrong	
	Initial	Final	Initial	Final	Initial	Final
Correct answer	.22 (.12)	.80 (.14)	.48 (.12)	.08 (.07)	.30 (.08)	.13 (.09)
Correct/incorrect	.21 (.14)	.28 (.15)	.46 (.12)	.32 (.12)	.33 (.11)	.41 (.10)
No feedback	.19 (.11)	.21 (.12)	.49 (.12)	.48 (.13)	.31 (.09)	.31 (.11)

Standard deviations are in parentheses.

McDermott & Roediger, 1998; Neuschatz, Payne, Lampinen, & Toglia, 2001; Watson et al., 2004). Nonetheless, we found that the vast majority of these errors were easily eliminated when the correction procedure abided by two general principles.

First, learners needed to notice that they had made an error; errors were greatly reduced when learners received trial-by-trial feedback and/or a specific prompt to compare their own answers to the feedback message. In the absence of these sorts of conditions, we suspect that many false memories may go unnoticed, in part because these errors are held with high confidence and vividness. Moreover, the close semantic relationship between a false memory and the correct information likely also contributes to learners' failures to notice their mistakes. That is, the false memory for *sleep* occurs precisely because this word shares a strong semantic relationship with the presented words; similarly, the false memory that *The new baby cried all night* is produced precisely because the correct version of the sentence was "designed to lead the listener to make schema-based inferences" (Sampaio & Brewer, 2009, p. 159). By definition, learners' errors share close semantic overlap with the correct information, which likely makes these errors particularly difficult to notice. The literature on semantic illusions supports the same point: Learners are surprisingly willing to provide meaningful answers to nonsensical questions such as *How many animals of each kind did Moses take on the ark?*, as long as the errorful term (*Moses*) shares a close semantic overlap with the correct answer (*Noah*; Erickson & Mattson, 1981; van Oostendorp & de Mul, 1990).

It is likely that students are often more aware of their errors in educational contexts, both because the erroneous responses may be made with lower confidence and because the contrast between one's error and the correct information is more obvious (e.g., it is easy to see that one has made an error when the correct answer to a math problem is 14, but one has produced the answer 6). Of course, however, there are exceptions to this general statement. For example, low-performing students are often thought to be "doubly cursed," in that they lack both the knowledge to perform well on academic tasks and the awareness that they are poor performers (Kruger & Dunning, 1999). Thus far, this "unskilled but unaware" phenomenon has been primarily demonstrated in terms of global predictions of performance (e.g., low-performing students tend to overpredict their overall performance on an upcoming test); however, it is also possible that it would apply, on an item-by-item basis, to evaluations of previous correctness. In other words, low performers may struggle to detect discrepancies between their own (wrong) answers and corrective feedback, thus perpetuating their poor performance.

As we argued previously, however, noticing one's error is only the first step toward error correction: Learners must also be provided with correct information in order to replace their mistakes. This conclusion has already been demonstrated many times in the broader literature on error correction (Bangert-Drowns et al., 1991; Fazio et al., 2010; Shute, 2008); here, we showed that the same rule applies for decreasing the production of false memories (see also Lewandowsky et al., 2012). Previous attempts to correct these strongly held errors often failed to abide by this principle, likely explaining the high rates of error persistence (e.g., McConnell & Hunt, 2007). The conclusion that effective feedback must go beyond identifying an error as "wrong" is also consistent with other findings that "forgetting" an old memory is more difficult than replacing it. For example, in the directed-forgetting literature, participants first encode some items (e.g., a list of words) and are then asked to intentionally forget that material and to learn some new information (e.g., a different list of words). Importantly, the encoding of List 2 is critical in facilitating the forgetting of List 1. In other words, after receiving the "forget" cue, participants must receive competing material to encode; otherwise, they are unable to forget the first list (Gelfand & Bjork, 1985; Pastötter & Bäuml, 2007, 2010).

The present results demonstrate that feedback can be effectively used to correct false memories, so long as learners receive correct information with which to replace their error. Moreover, such feedback is most useful when received directly after committing an error, or when the correction conditions otherwise facilitate a direct comparison between learners' errors and the correct answers. Despite the surprising robustness of false memories in the face of other correction attempts,

abiding by both of these recommendations enables learners to eliminate nearly all of their mistakes. However, one issue for future research involves the translation of this work from the laboratory to the real world. That is, it is less clear that it will always be possible to abide by both recommendations in correcting many types of real-world errors. For example, in contrast to translations of foreign words or definitions or obscure vocabulary, it is much less likely that an objective "truth" could be compared to one's personal memory. That is, barring photographs or a video-recording, no dictionary or other reference volume can provide corrective feedback for false memories of one's personal experiences. Moreover, one's personal memories may engender a sense of reliving or may elicit high confidence beyond the level simulated in the present experiments, lessening people's willingness to accept corrections.

Author note A National Science Foundation Graduate Research Fellowship supported the first author. We thank Taylor Walls, Michael Liou, Andrew Ball, and Rajan Khanna for their assistance with the data collection and coding.

References

Anastasi, J. S., Rhodes, M. G., & Burns, M. C. (2000). Distinguishing between memory illusions and actual memories using phenomenological measurements and explicit warnings. *American Journal of Psychology, 113,* 1–26.

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology, 62,* 148–156.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213–238.

Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition, 5,* 673–678.

Butler, A. C., Fazio, L. K., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high confidence errors return. *Psychonomic Bulletin & Review, 18,* 1238–1244. doi:10.3758/s13423-011-0173-y

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105,* 290–298.

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13,* 273–281. doi:10.1037/1076-898X.13.4.273

Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36,* 604–616. doi:10.3758/MC.36.3.604

Carpenter, S. K., & Vul, E. (2011). Delaying feedback by three seconds benefits retention of face–name pairs: The role of active anticipatory processing. *Memory & Cognition, 39,* 1211–1221. doi:10.3758/s13421-011-0092-1

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380. doi:10.1037/0033-2909.132.3.354

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20,* 540–551.

Fazio, L. K., Huelser, B. J., Johnson, A., & Marsh, E. J. (2010). Receiving right/wrong feedback: Consequences for learning. *Memory, 18,* 335–350.

Fritz, C. O., Morris, P. E., Bjork, R. A., Gelman, R., & Wickens, T. D. (2000). When further learning fails: Stability and change following repeated presentation of text. *British Journal of Psychology, 91,* 493–511.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review, 4,* 271–276. doi:10.3758/BF03209405

Gelfand, H., & Bjork, R. A. (1985). *On the locus of retrieval inhibition in directed forgetting*. Boston: Paper presented at the meeting of the Psychonomic Society.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Kay, H. (1955). Learning and retaining verbal material. *British Journal of Psychology, 46,* 81–100.

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology, 16,* 399–415.

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

Lantz, M. E., & Stawiski, A. (2014). Effectiveness of clickers: Effect of feedback and timing of questions on learning. *Computers in Human Behavior, 31,* 280–286. doi:10.1016/j.chb.2013.10.009

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13,* 106–131. doi:10.1177/1529100612451018

Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition, 43,* 14–26. doi:10.3758/s13421-014-0452-8

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of auto-mobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589.

McConnell, M. D., & Hunt, R. R. (2007). Can false memories be corrected by feedback in the DRM paradigm? *Memory & Cognition, 35,* 999–1006.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35,* 212–230. doi:10.1006/jmla.1996.0012

McDermott, K. B., & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition, 34,* 1273–1284.

McDermott, K. B., & Roediger, H. L., III. (1998). Attempting to avoid illusory memory: Robust false recognition of associates persists under conditions of explicit warnings and immediate tests. *Journal of Memory and Language, 39,* 508–520. doi:10.1006/jmla.1998.2582

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

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition, 37,* 1077–1087. doi:10.3758/MC.37.8.1077

Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes student knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition, 3,* 222–229. doi:10.1016/j.jarmac.2014.05.001

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory, 9,* 53–71.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3–8. doi:10.1037/0278-7393.31.1.3

Pashler, H., Rohrer, D., Cepeda, N., & Carpenter, S. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14,* 187–193.

Pastötter, B., & Bäuml, K.-H. (2007). The crucial role of postcue encoding in directed forgetting and context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 977–982. doi:10.1037/0278-7393.33.5.977

Pastötter, B., & Bäuml, K.-H. (2010). Amount of postcue encoding predicts amount of directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 54–65.

Phye, G. D., & Andre, T. (1989). Delayed retention effect: Attention, perseveration, or both? *Contemporary Educational Psychology, 14,* 173–185.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814. doi:10.1037/0278-7393.21.4.803

Roediger, H. L., III, & Smith, M. A. (2012). The "pure-study" learning curve: The learning curve without cumulative testing. *Memory & Cognition, 40,* 989–1002. doi:10.3758/s13421-012-0213-5

Sampaio, C., & Brewer, W. F. (2009). The role of unconscious memory errors in judgments of confidence for sentence recognition. *Memory & Cognition, 37,* 158–163. doi:10.3758/MC.37.2.158

Sassenrath, J. M., & Yonge, G. D. (1968). Delayed information feedback, feedback cues, retention set, and delayed retention. *Journal of Educational Psychology, 59,* 69–73.

Sheen, M., Kemp, S., & Rubin, D. (2001). Twins dispute memory ownership: A new false memory phenomenon. *Memory & Cognition, 29,* 779–788.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153–189.

Sitzman, D. M., Rhodes, M. G., & Tauber, S. K. (2014). Prior knowledge is more predictive of error correction than subjective confidence. *Memory & Cognition, 42,* 84–96. doi:10.3758/s13421-013-0344-3

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 80–95.

van Oostendorp, H., & de Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica, 74,* 35–46.

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive reminding in proactive effects of memory. *Memory & Cognition, 41,* 1–15. doi:10.3758/s13421-012-0246-9

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger–McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32,* 135–141. doi:10.3758/BF03195826