# Prior knowledge is more predictive of error correction than subjective confidence

**Danielle M. Sitzman · Matthew G. Rhodes ·
Sarah K. Tauber**

**Abstract** Previous research has demonstrated that, when given feedback, participants are more likely to correct confidently-held errors, as compared with errors held with lower levels of confidence, a finding termed the *hypercorrection effect*. Accounts of hypercorrection suggest that confidence modifies attention to feedback; alternatively, hypercorrection may reflect prior domain knowledge, with confidence ratings simply correlated with this prior knowledge. In the present experiments, we attempted to adjudicate among these explanations of the hypercorrection effect. In Experiments 1a and 1b, participants answered general knowledge questions, rated their confidence, and received feedback either immediately after rating their confidence or after a delay of several minutes. Although memory for confidence judgments should have been poorer at a delay, the hypercorrection effect was equivalent for both feedback timings. Experiment 2 showed that hypercorrection remained unchanged even when the delay to feedback was increased. In addition, measures of recall for prior confidence judgments showed that memory for confidence was indeed poorer after a delay. Experiment 3 directly compared estimates of domain knowledge with confidence ratings, showing that such prior knowledge was related to error correction, whereas the unique role of confidence was small. Overall, our results suggest that prior knowledge likely plays a primary role in error correction, while confidence may play a small role or merely serve as a proxy for prior knowledge.

**Keywords** Hypercorrection effect · Feedback · Memory · Confidence

D. M. Sitzman (✉) · M. G. Rhodes
Department of Psychology, Colorado State University,
Fort Collins, CO 80523-1876, USA
e-mail: danielle.sitzman@colostate.edu

S. K. Tauber
Department of Psychology, Kent State University, Kent, OH, USA

Feedback is essential not only for maintaining correct responses, but also for correcting errors. A host of research has suggested that a person's confidence that they have correctly answered a question influences how they process feedback (Butler, Karpicke, & Roediger, 2008; Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009). In the present experiments, we attempted to examine how confidence in a response and prior domain knowledge contribute to error correction in memory.

## Feedback and error correction

Prior work indicates that a person's confidence may influence the likelihood that they correct errors. In particular, Butterfield and Metcalfe (2001, 2006) demonstrated that, counterintuitively, participants were more likely to correct errors held with high levels of confidence than errors held with low levels of confidence, a finding termed the *hypercorrection effect*. Specifically, Butterfield and Metcalfe (2001) had participants answer general knowledge questions (e.g., "What poison did Socrates take at his execution?") and rate their confidence in the correctness of their response. Participants were then given feedback confirming whether an answer was correct or displaying the correct answer if a response was incorrect. After a short delay (5 min), participants were retested on the same questions. Hypercorrection was measured as the average of within-subject Kruskal–Goodman gamma correlations (Nelson, 1984) between confidence for errors made on the first test and the accuracy of the answers for those questions on a final test. The mean correlation across participants was positive ($G = .36$), indicating that high-confidence errors were more likely to be corrected on a subsequent test than errors held with lower levels of confidence.

Butterfield and Metcalfe (2001) posited that the hypercorrection effect occurred because attention to feedback

increased following high-confidence errors due to the discrepancy between the subjective assessment of performance (i.e., the confidence that one is correct) and the correct answer. Consistent with this, Kulhavy, Yekovich, and Dyer (1976) demonstrated that participants spent longer processing feedback after high-confidence errors than after low-confidence errors. Recently, Butterfield and Metcalfe (2006) demonstrated that participants performed more poorly on a tone detection task while processing feedback for high-confidence errors, as compared with low-confidence errors. They suggested that additional attention allocated to feedback following high-confidence errors diverted attention from the secondary task. Fazio and Marsh (2009) likewise demonstrated that participants were more likely to remember the color of the font that feedback was presented in for high-confidence errors, as compared with low-confidence errors. Thus, relative to low-confidence errors, feedback on high-confidence errors may be more likely to capture attention and, therefore, lead to more sustained processing and a greater chance of correction on a subsequent test (Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009; Kulhavy et al., 1976).

Additional research on the hypercorrection effect has indicated that prior domain knowledge also contributes to error correction (Butterfield & Mangels, 2003; Metcalfe & Finn, 2011). According to this "knew-it-all-along" explanation, people are more likely to correct errors accompanied by higher levels of prior domain knowledge. Such domain knowledge will increase the probability that participants answer questions correctly but may also elevate confidence when answering incorrectly (cf. Koriat, 2008, 2012). For example, suppose that one has high levels of knowledge of European geography but mistakenly produces "Glasgow" when asked to name the capital of Scotland. Familiarity with the domain (European geography) might elevate confidence in the erroneous response. However, given high domain knowledge, the correct response (i.e., Edinburgh) may have also been known and accessible. According to this knew-it-all-along account, prior knowledge facilitates error correction because the correct response may already be partially known (Metcalfe & Finn, 2011).

Indeed, Metcalfe and Finn (2011) observed that participants were more likely to claim that they knew the correct answer all along after receiving feedback on general knowledge questions for high-confidence errors, as compared with low-confidence errors. For example, after answering a question incorrectly, participants in one experiment generated a new response for some errors or identified the answer among multiple options. Participants were more likely to generate the correct response and were more likely to choose the correct response following high-confidence errors than following low-confidence errors. Participants were also more

likely to report that they had prior knowledge of the correct response after high-confidence errors, as compared with low-confidence errors. Thus, these knew-it-all-along judgments were reliably related to error correction, suggesting that prior knowledge plays a significant role in error correction.

Overall, the prevailing accounts of the hypercorrection effect hold that errors are more likely to be corrected either if the participant has high levels of confidence in their errors (Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009) or if he or she has high levels of domain knowledge for questions answered erroneously (Butterfield & Mangels, 2003; Metcalfe & Finn, 2011). However, it is unclear how confidence and prior knowledge uniquely or jointly influence error correction. Specifically, confidence in the correctness of an answer may reflect a variety of information, including prior knowledge of a particular domain, the rapidity with which a response was generated, and/or the amount of information coming to mind (e.g., Koriat, 2008). If confidence uniquely contributes to error correction, above and beyond factors like prior knowledge, then error correction should be influenced by how well one's initial confidence judgment is remembered. That is, a corollary of a confidence-based account of hypercorrection is that participants should be less likely to correct high-confidence errors when placed in situations where their original confidence judgments are less accessible. If errors held with high levels of confidence encourage greater attention to the correct response, hypercorrection of errors should be less prevalent if one's original confidence judgment is less accessible. However, if hypercorrection of errors is primarily driven by the level of prior knowledge, as the knew-it-all-along explanation suggests, hypercorrection effects should be evident regardless of the accessibility of prior confidence judgments. As well, if confidence judgments are largely the product of prior knowledge, confidence alone should have little or no unique contribution to error correction. We know of no prior work that has sought to manipulate the accessibility of prior confidence as a method of testing theoretical explanations of hypercorrection. While previous research provides some support for both explanations, little research has employed a design that allows one to determine the unique contribution of each to error correction.

## The present study

The two prevailing accounts of hypercorrection were examined in the experiments reported by manipulating the amount of time that elapsed between a response (and confidence judgment) and feedback. Specifically, participants answered a series of general knowledge questions and provided an

indication of their confidence that the answer produced was correct. Feedback, in the form of the correct answer, was provided either immediately after a confidence judgment or after some delay, on the assumption that confidence would be more accessible after immediate than after delayed feedback. This was followed by a second test on the same general knowledge questions. If confidence is primarily driving attention to feedback, the hypercorrection effect should be more pervasive for immediate than for delayed feedback. However, if hypercorrection is primarily a function of prior domain knowledge (which we assume would not change over a short delay), the effect should hold regardless of feedback timing. In all, the experiments reported attempted to determine the specific roles of confidence and prior knowledge during error correction.

## Experiments 1a and 1b

Participants in Experiment 1a answered general knowledge questions, rated their confidence in the correctness of their answer, and were given feedback on each answer. Feedback was provided either immediately after an answer or following a delay. Finally, participants were tested on these general knowledge questions a second time. Our primary interest was in the hypercorrection effect (i.e., the correlation between confidence in errors on test 1 and accuracy on test 2). That is, we examined whether the hypercorrection effect differed for immediate, as compared with delayed, feedback.

Experiment 1b was identical to Experiment 1a, with the exception that the final retention test occurred 48 h after the first test. Prior work (Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009) has generally used immediate retention tests (but for exceptions, see Butler, Fazio, & Marsh, 2011; Butterfield & Mangels, 2003). Thus, we examined whether hypercorrection obtained on a delayed final test when feedback was administered immediately and at a delay.

Method

### Participants

A total of 64 students from Colorado State University participated for partial course credit, divided equally between Experiment 1a and Experiment 1b.

### Materials

Seventy-two general knowledge questions were obtained from Nelson and Narens (1980). Questions were evenly divided among easy (probability of recall, $\geq$.660),

medium (probability of recall between .659 and .360), and hard (probability of recall, $\leq$.359) levels of difficulty and were chosen if they were considered reasonable for current students to answer (e.g., What is the name of Dorothy's dog in "The Wizard of Oz?" Answer: Toto). Pilot testing with a separate group of 68 participants showed that the level of difficulty reported by Nelson and Narens correlated strongly ($r = .81$) with the performance of the sample from our population of participants (see also Tauber, Dunlosky, Rawson, Rhodes, & Sitzman, in press).

### Procedure

All participants in Experiments 1a and 1b were given 72 general knowledge questions. These were randomly divided into two sets of 36 questions that served equally often in the immediate and delayed feedback conditions (the first 3 and last 3 questions in each set were buffer items and were not included in analyses). Questions were displayed on the screen one at a time, and participants were given 10 s to write down an answer on a sheet of paper provided. Participants were encouraged to answer each question to the best of their ability but were not forced to provide responses. Immediately following their answer, participants were given 5 s to rate their confidence in the correctness of their response on a scale from 0 to 100 (with 0 indicating *no confidence in the accuracy of a response* and 100 indicating *absolute confidence in the accuracy of a response*). Questions were presented in two blocks that varied on the basis of the timing of the feedback provided. Half of the participants received immediate feedback during the first block and delayed feedback in the second block, with the remaining participants receiving delayed and immediate feedback in the reverse order. In the immediate feedback condition, participants were shown the correct answer after they rated their confidence for each general knowledge question. While viewing the correct answer, they were given 5 s to circle YES or NO on an answer sheet to indicate whether their response was correct (this was done to ensure that participants were attending to feedback). For questions in the delayed feedback condition, participants first answered each question and rated their confidence. After answering all 36 questions, participants were then shown each question again with the correct response and were given 5 s to circle YES or NO on an answer sheet to indicate whether their response was correct. On average, the delay to feedback was approximately 6 min.

For Experiment 1a, after receiving feedback on the second block of questions, participants were shown the questions from the first block again. The procedure was the same, except that feedback was not provided. Participants were

then given a 5-min math distractor test. This was used to roughly equate the amount of time (lag) between feedback and the final retention test for both sets of questions. After the filler task, participants were given the second set of questions and were asked to provide an answer and rate their confidence for each question. (Questions without a response were scored as incorrect.) The entire experiment took approximately 1 h to complete. For Experiment 1b, the procedure was identical to that in Experiment 1a, with the exception that test 2 was given 48 h after the first test.

Results

In the following analyses, we first examine performance on test 1 and test 2 as a function of feedback timing. We then report our focal analyses examining the relationship between confidence judgments and error correction (i.e., hypercorrection). The alpha level was set at .05 for all analyses reported.

*Proportion of questions answered correctly*

*Experiment 1a* As would be expected, there was no difference between the immediate and delayed feedback conditions in the proportion of questions answered correctly on test 1, $t(31) = 1.04$, $p = .31$, $d = 0.18$ (see Table 1). On test 2, there was no difference in the proportion of correct responses as a function of initial feedback timing, $t < 1$. There was no difference in the proportion of correct responses retained from test 1 to test 2 regardless of whether feedback was presented immediately ($M = .99$, $SE = .003$) or after a delay ($M = .99$, $SE = .003$), $t < 1$. In addition, participants were equally likely to correctly respond to questions on test 2 that were initially incorrect on test 1 in the immediate ($M = .65$, $SE = .03$) and delayed ($M = .62$, $SE = .03$) feedback conditions, $t < 1$. Thus, there was no difference in the proportion of questions answered correctly as a function of feedback timing.[1]

*Experiment 1b* The proportion of questions correctly answered for items given immediate and delayed feedback in Experiment 1b did not reliably differ on test 1, $t(31) = 1.04$, $p = .31$, $d = 0.18$ (see Table 1). On test 2, participants were more likely to correctly answer questions

---

[1] An additional 16 participants took both tests but received no feedback on answers for questions presented on the first test. Paired tests showed that participants receiving immediate, $t(46) = 8.10$, $p < .001$, $d = 2.48$, and delayed, $t(46) = 7.34$, $p < .001$, $d = 2.25$, feedback performed significantly better than those not receiving feedback on test 2 ($M = .46$, $SE = .04$).

that had received immediate feedback, as compared with questions that had received delayed feedback, $t(31) = 2.54$, $p = .02$, $d = 0.38$. This difference in feedback was largely driven by the proportion of errors corrected. In particular, there was no difference in the proportion of correct responses retained from test 1 to test 2 between immediate ($M = .98$, $SE = .01$) and delayed ($M = .99$, $SE = .01$) feedback, $t < 1$. However, participants were reliably more likely to correct errors in the immediate ($M = .51$, $SE = .02$), as compared with the delayed ($M = .42$, $SE = .03$), feedback condition, $t(31) = 2.52$, $p = .02$, $d = 0.59$.

*Hypercorrection effect*

Consistent with previous research (e.g., Butterfield & Metcalfe, 2001, 2006), the hypercorrection effect was analyzed by examining the gamma correlation between confidence judgments given to items answered incorrectly on test 1 and the accuracy of the responses to those questions on test 2. A hypercorrection effect is indicated by a positive correlation (i.e., errors with higher confidence are more likely to be corrected on test 2) that is reliably greater than zero. Several participants reported invariant confidence judgments. These participants were excluded from analyses, reflected by variations in degrees of freedom reported for statistical tests in each of the experiments reported. (See the Appendix for the distribution of confidence judgments given to items that were incorrect on test 1 in this and subsequent experiments.)

*Experiment 1a* For both the immediate feedback condition ($G = .48$, $SE = .11$), $t(27) = 4.50$, $p < .001$, and the delayed feedback condition ($G = .52$, $SE = .08$), $t(30) = 6.38$, $p < .001$, a reliable hypercorrection effect was obtained. However, the hypercorrection effect did not differ between the two feedback timings, $t < 1$.

*Experiment 1b* A reliable hypercorrection effect was evident for questions receiving immediate feedback ($G = .38$, $SE = .11$), $t(31) = 3.54$, $p = .001$, and for questions receiving delayed feedback ($G = .35$, $SE = .07$), $t(30) = 4.61$, $p < .001$. The magnitude of the hypercorrection effect did not differ by feedback timing, $t < 1$.

Did the hypercorrection effect differ between Experiment 1a, which used an immediate retention test, and Experiment 1b, which used a delayed (48-h) retention test? We examined this by treating experiment as a between-subjects factor in a 2 (Experiment: 1a, 1b) × 2 (feedback timing: immediate, delayed) mixed-factor ANOVA. Results showed no reliable differences in the magnitude of gamma correlations between experiments,

**Table 1** Proportions of questions answered correctly

| | Test 1 | | Test 2 | |
| --- | --- | --- | --- | --- |
| | Immediate Feedback | Delayed Feedback | Immediate Feedback | Delayed Feedback |
| Experiment 1a | .44 (.02) | .46 (.02) | .81 (.02) | .80 (.02) |
| Experiment 1b | .44 (.03) | .46 (.02) | .72 (.02) | .67 (.03) |
| Experiment 2 | .40 (.02) | .41 (.02) | .65 (.17) | .62 (.17) |
| Experiment 3 | .39 (.02) | .39 (.02) | .65 (.02) | .66 (.02) |

*Note.* Standard error of the mean is presented in parentheses.

$F(1, 57) = 1.97, p = .17, \eta^2_p = .03$. As well, there was no reliable effect of feedback timing, $F < 1$, and feedback timing did not interact with Experiment, $F < 1$.

Discussion

Overall, the hypercorrection effect was evident for questions receiving immediate feedback and questions receiving delayed feedback. In addition, hypercorrection was obtained on a delayed retention test (cf. Butler et al., 2011) and did not differ from the hypercorrection effect obtained in Experiment 1a, where an immediate retention test was used. Thus, hypercorrection did not differ as a function of the delay between the confidence judgment and feedback.

**Experiment 2**

In Experiments 1a and 1b, the relationship between confidence and error correction was similar for the immediate and delayed feedback conditions. Such data would seem to minimize the unique role of confidence in moderating attention to feedback, given that confidence should be less accessible for delayed than for immediate feedback. We sought to provide more direct evidence for this in Experiment 2. In particular, participants in Experiment 2 were asked to recall their original confidence judgments after they were given feedback. As well, one might argue that the delay used in Experiment 1 between confidence judgments and the onset of feedback was minimal (approximately 6 min), providing a relatively weak test of accounts of hypercorrection contingent on access to initial confidence. Thus, in Experiment 2, the interval between a response on test 1 and feedback in the delayed feedback condition was increased from 6 to 25 min.

We anticipated that this delay would diminish the accessibility of prior confidence judgments, relative to the immediate feedback condition, rendering participants less likely to accurately recall their prior confidence judgment following a 25-min delay, as compared with immediate feedback. Accordingly, if confidence uniquely influences attention to feedback, the hypercorrection effect should be diminished for the delayed feedback condition. That is, if participants are less able to recall their level of confidence, they should be less likely to selectively attend to high-confidence errors. Alternatively, if the hypercorrection effect is mainly driven by prior knowledge, it should remain stable regardless of the accessibility of prior confidence judgments.

Method

*Participants*

Sixty-four students from Colorado State University participated for partial course credit.

*Materials and procedure*

The materials used and procedure were identical to those in Experiment 1b, with three exceptions. First, the interval between the response to a question on test 1 and feedback in the delayed feedback condition was extended to 25 min. This time was filled with unrelated tasks, such as Sudoku. Second, after receiving feedback, participants were asked to recall their initial confidence judgment on test 1 for that question as accurately as possible. Lastly, all responses were entered into the computer by the participants.

Results

*Proportion of questions answered correctly*

As was expected, there was no difference between the immediate and delayed feedback conditions in the proportion of questions answered correctly on test 1, $t < 1$ (see Table 1). However, on test 2, participants answered more questions correctly that had initially been given immediate feedback ($M = .65, SE = .02$) than questions that had initially received

delayed feedback ($M = .62$, $SE = .02$), $t(63) = 2.37$, $p = .021$, $d = 0.19$. This finding was driven by the proportion of errors corrected. Although there was no difference between immediate ($M = .97$, $SE = .01$) and delayed ($M = .95$, $SE = .02$) feedback in the proportion of correct responses retained from test 1 to test 2, $t < 1$, participants were more likely to correct errors in the immediate feedback condition ($M = .46$, $SE = .02$) than in the delayed feedback condition ($M = .40$, $SE = .02$), $t(63) = 2.76$, $p = .008$, $d = 0.35$.

*Recall of confidence judgments*

Recall accuracy for confidence judgments was examined in several ways. First, we examined the proportion of confidence judgments correctly recalled within 10 points of the initial confidence judgment (see Fig. 1; data were nearly identical when using a smaller, 5-point window). These data were analyzed in a 2 (feedback timing: immediate, delayed) × 2 (response accuracy: correct, incorrect) repeated measures ANOVA. Overall, participants correctly recalled a greater proportion of confidence judgments in the immediate feedback condition ($M = .97$, $SE = .01$) than in the delayed feedback condition ($M = .80$, $SE = .01$), $F(1, 63) = 179.44$, $p < .001$, $\eta^2_p = .74$. Participants were also more likely to correctly recall confidence judgments for items answered incorrectly ($M = .91$, $SE = .01$), as compared with items answered correctly ($M = .86$, $SE = .01$), $F(1, 63) = 16.45$, $p < .001$, $\eta^2_p = .21$. These main effects were qualified by a significant feedback timing × response accuracy interaction, $F(1, 63) = 7.28$, $p = .01$, $\eta^2_p = .10$. Participants recalled their initial confidence judgments more accurately for incorrect responses ($M = .98$, $SE = .01$) than for correct responses ($M = .96$, $SE = .01$) in the immediate feedback condition, $t(63) = 2.20$, $p = .03$. However, this discrepancy between correct



**Fig. 1** Mean proportions of initial confidence judgments correctly recalled from test 1 within 10 points of the actual confidence judgment for Experiment 2. Error bars represent one standard error of the mean

($M = .76$, $SE = .19$) and incorrect ($M = .84$, $SE = .10$) responses was greater in the delayed feedback condition, $t(63) = 3.75$, $p < .001$. Thus, overall, recall of confidence judgments in the immediate feedback condition was characterized by near ceiling levels of performance, whereas delayed feedback led to much less accurate levels of recall.

Further inspection of confidence recall data suggested that participants were highly accurate when recalling their confidence for judgments that had been 0 %. This may reflect instances where participants had no knowledge of the answer and, thus, were certain when seeing a question a second time that their initial confidence judgment was zero. By extension, instances of recalling confidence judgments of 0 % may inflate recall accuracy. Thus, we performed an additional analysis with confidence judgments of 0 % removed. These data were examined in a 2 (feedback timing: immediate, delayed) × 2 (response accuracy: correct, incorrect) repeated measures ANOVA. Overall, participants correctly recalled a greater proportion of confidence judgments in the immediate feedback condition ($M = .95$, $SE = .01$) than in the delayed feedback condition ($M = .66$, $SE = .02$), $F(1, 61) = 184.27$, $p < .001$, $\eta^2_p = .75$. Participants were also more likely to correctly recall confidence judgments for items answered correctly ($M = .87$, $SE = .01$), as compared with items answered incorrectly ($M = .75$, $SE = .02$), $F(1, 61) = 41.34$, $p < .001$, $\eta^2_p = .40$. These main effects were qualified by a significant feedback timing × response accuracy interaction, $F(1, 61) = 25.31$, $p < .001$, $\eta^2_p = .29$. When given delayed feedback, participants recalled their initial confidence judgments more accurately for correct responses ($M = .77$, $SE = .02$) than for incorrect responses ($M = .56$, $SE = .03$), $t(62) = 6.04$, $p < .001$. However, there was only a marginal difference in confidence recall accuracy for correct ($M = .97$, $SE = .01$) and incorrect ($M = .94$, $SE = .02$) responses for items in the immediate feedback condition, $t(62) = 1.92$, $p = .059$. Thus, when confidence ratings of 0 % were removed, participants were still reliably more accurate at correctly recalling their initial confidence judgments in the immediate feedback condition than in the delayed feedback condition.

In addition to examining overall recall accuracy for confidence judgments, we also examined the gamma correlation between initial confidence judgments and the recall of those judgments for all items. That is, whereas recall accuracy provides a measure of the correspondence between actual and recalled accuracy, the correlation provides an indication of whether recalled confidence judgments faithfully discriminated between high- and low-confidence judgments. A strong, positive correlation would indicate that items given low confidence yielded low recalled confidence and items given
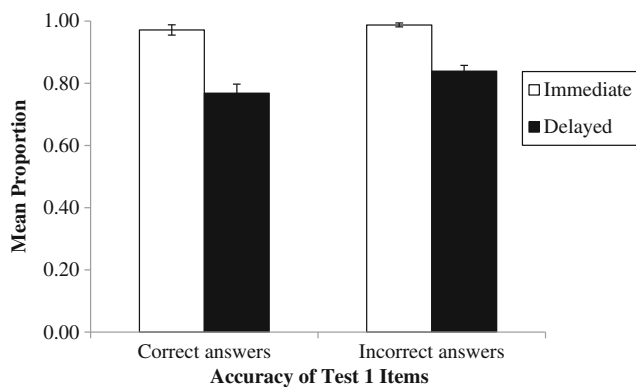
high confidence yielded high recalled confidence. These data showed that the correlation between initial confidence judgments and the recall of those judgments was stronger in the immediate feedback condition ($G = .95$, $SE = .02$) than in the delayed feedback condition ($G = .81$, $SE = .03$), $F(1, 63) = 13.47$, $p = .001$, $\eta^2_p = .18$. There was no difference between items answered correctly, as compared with errors, $F < 1$, and there was not a significant interaction, $F < 1$. Collectively, these data indicate that recall of confidence judgments was superior for immediate, as compared with delayed, feedback.

We note that we tested several control conditions to ensure that recalling initial confidence judgments did not impact overall performance and to examine whether recall accuracy for confidence judgments was influenced by whether initial confidence was recalled before or after receiving feedback.[2] Overall, recalling initial confidence judgments did not influence performance in any way. In addition, participants were less accurate in recalling initial confidence judgments after a delay, regardless of whether they attempted to recall their initial confidence immediately before receiving feedback or immediately after.

*Hypercorrection effect*

For both the immediate feedback condition ($G = .21$, $SE = .06$), $t(61) = 3.59$, $p = .001$, and the delayed feedback condition ($G = .19$, $SE = .07$), $t(61) = 2.76$, $p = .008$, a reliable hypercorrection effect was obtained. However, the hypercorrection effect did not differ between the two feedback timings, $t < 1$. Thus, hypercorrection did not differ as a function of the accessibility of confidence judgments. Further analyses suggested that hypercorrection also did not differ as a function of whether or not participants could recall their initial confidence judgments.[3]

---

[2] Two control conditions were tested to ensure that requiring participants to recall their initial confidence judgments did not alter other aspects of performance. Participants either were not asked to recall their confidence judgments or recalled their initial confidence judgments prior to receiving feedback. This was done to ensure that knowledge of the correct answer did not influence recall of confidence judgments in any way. Overall, there were no differences in performance on test 1 or test 2 or in the magnitude of hypercorrection among these conditions, Fs < 1.

[3] Specifically, the hypercorrection effect in the delayed feedback condition did not differ between items where the initial confidence was recalled ($G = .34$, SE = .11) and items where initial confidence was not correctly recalled ($G = .20$, SE = .17), t < 1. However, given that more observations contributed to data for when confidence was recalled, this result should be interpreted with some caution.

Discussion

Experiment 2 indicated that participants less accurately recalled their prior confidence judgment following delayed, as compared with immediate, feedback. If the hypercorrection effect is primarily contingent on confidence modifying attention to feedback for high-confidence errors, it should have been reduced or nonexistent when information about prior confidence was less accessible (i.e., in the delayed feedback condition). However, consistent with Experiment 1, participants in Experiment 2 showed similar levels of hypercorrection in the immediate and delayed feedback conditions.

**Experiment 3**

Collectively, the results from Experiments 1 and 2 indicated that the hypercorrection effect did not differ as a function of feedback timing. Indeed, in Experiment 2, we observed similar hypercorrection effects for immediate and delayed feedback, despite evidence of less accurate memory for confidence judgments in the delayed feedback condition, as compared with the immediate feedback condition. Such data are generally inconsistent with accounts suggesting that confidence is the primary mechanism behind attention to feedback.

In contrast, a knew-it-all-along account (Metcalfe & Finn, 2011) predicated on stable domain knowledge would predict a similar hypercorrection effect regardless of the delay between a response and feedback. That is, prior knowledge would drive confidence, such that errors made for high-knowledge domains would engender high levels of confidence. If a participant claims knowledge of the correct response all along, that individual is more likely to correct that error than errors accompanied by less prior knowledge. Although the invariance in hypercorrection across the delay prior to feedback is consistent with this hypothesis, albeit contingent on a null effect, we sought a more direct test of the knew-it-all-along account. Thus, in Experiment 3, after participants received feedback (either immediately or after a delay) they were asked to indicate whether or not they actually knew the correct answer all along (cf. Metcalfe & Finn, 2011). Prior work by Metcalfe and Finn suggested that participants are able to accurately index their knowledge of a specific domain after receiving feedback. In particular, participants claiming that they actually knew the correct response all along were more likely to correctly answer the question on a second attempt (prior to feedback), more likely to generate a correct response if wrong, and more likely to identify a correct response among several choices. Thus, the

knew-it-all-along judgment made after feedback served as a valid basis for measuring prior knowledge of the correct response to a particular general knowledge question.

Consistent with Experiments 1 and 2, we did not expect to find differences in hypercorrection in Experiment 3 based on feedback timing. More important, Experiment 3 permitted us to examine the role of both confidence and prior knowledge in error correction. If prior knowledge is the primary factor contributing to error correction, error correction should be better predicted by prior knowledge judgments than by confidence judgments. Alternatively, if confidence is the primary factor responsible for error correction, confidence judgments should be a better predictor of error correction than prior knowledge judgments. As well, if subjective confidence judgments are merely a proxy for prior knowledge, confidence judgments should not uniquely predict error correction. Experiment 3 thus allowed us to independently examine the role of each variable (confidence or prior knowledge) in error correction.

Method

*Participants*

Thirty-two students from Colorado State University participated for partial course credit.

*Materials and procedure*

The materials used and procedure were identical to those in Experiment 2, with two exceptions. First, unlike in Experiment 2, participants were not asked to recall their initial confidence judgments. Second, after receiving feedback, participants were asked to rate their prior knowledge using a measure adapted from Metcalfe and Finn (2011). Specifically, participants provided a rating from 1 to 7 for the question, "Did you actually know the answer all along?" A rating of 1 indicated *that's new to me*, while a rating of 7 indicated *I actually knew it all along*.[4] Confidence judgments and knew-it-all-along judgments were made on two different scales to ensure that participants clearly separated the two types of judgments.

Results

*Proportion of questions answered correctly*

There was no difference between the immediate and delayed feedback conditions in the proportion of questions answered

correctly on test 1, $t < 1$ (see Table 1). There was also no difference in performance on test 2 based on the timing of feedback, $t < 1$. An equal proportion of correct responses were retained from test 1 to test 2, regardless of whether the responses were initially given immediate feedback or delayed feedback, $t < 1$. In addition, errors were equally likely to be corrected regardless of initial feedback timing, $t < 1$.

*Hypercorrection effect and knew-it-all-along judgments*

*Hypercorrection* For both the immediate feedback condition ($G = .24$, $SE = .08$), $t(30) = 2.99$, $p = .005$, and the delayed feedback condition ($G = .28$, $SE = .09$), $t(31) = 3.15$, $p = .004$, a reliable hypercorrection effect was obtained. However, the hypercorrection effect did not differ between the two feedback timings, $t < 1$ (see the left panel of Fig. 2).

*Knew it all along* To examine the relationship between prior knowledge and error correction, we calculated gamma correlations between knew-it-all-along judgments given to errors on test 1 and the accuracy of responses to those same questions on test 2 (see the right panel of Fig. 2). Similar to the hypercorrection effect, positive correlations indicate that higher levels of reported prior knowledge are related to error correction on test 2. For items in both the immediate condition ($G = .68$, $SE = .04$), $t(30) = 16.43$, $p < .001$, and the delayed feedback condition ($G = .74$, $SE = .04$), $t(31) = 19.49$, $p < .001$, knew-it-all-along ratings for initial errors were strongly positively correlated with accuracy on test 2. However, the magnitude of these correlations did not differ on the basis of feedback timing, $t < 1$.

We also examined whether confidence judgments or knew-it-all-along judgments were more strongly related to error correction. These data were analyzed in a 2 (feedback timing: immediate, delayed) × 2 (judgment gamma:
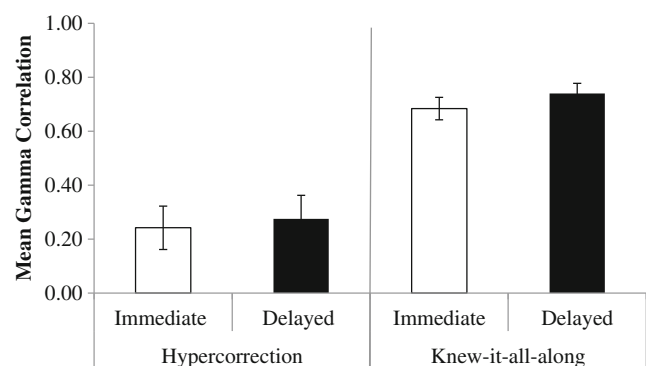


Fig. 2 Mean gamma correlations between confidence in errors on test 1 and accuracy on test 2 (i.e., the hypercorrection effect) and between knew-it-all-along judgments for errors on test 1 and accuracy on test 2, as a function of feedback timing for Experiment 3. Error bars represent one standard error of the mean

---

[4] Although Metcalfe and Finn (2011) used a dichotomous yes/no rating to assess prior knowledge, we suggest that participants may have varying degrees of prior knowledge. Thus, the Likert-type scale used permitted us to examine judgments as a continuous, rather than discrete, state.

confidence, knew-it-all-along) repeated measures ANOVA. Overall, there was no main effect of feedback timing, $F < 1$. However, the relationship between knew-it-all-along judgments and error correction ($G = .71$, $SE = .03$) was reliably greater than the relationship between confidence and error correction ($G = .27$, $SE = .06$), $F(1, 30) = 59.26$, $p < .001$. Feedback timing did not interact with judgment type, $F < 1$.

*Logistic hierarchical linear modeling analyses*

Next, we examined the simultaneous influence of confidence judgments and prior knowledge as the basis for error correction. Thus, for items that were incorrect on test 1, we explored the influence of test 1 confidence judgments and knew-it-all-along judgments on test 2 accuracy. If confidence judgments uniquely predict error correction and play a primary role in error correction, they should reliably predict greater levels of error correction than should knew-it-all-along judgments. However, if error correction is mainly driven by prior knowledge, knew-it-all-along judgments should reliably predict greater levels of error correction than should confidence judgments.

These predictions were tested in two logistic hierarchical linear models (HLMs; cf. Hines, Touron, & Hertzog, 2009; Tauber & Rhodes, 2012). The HLM allows the evaluation of multiple metacognitive judgments (i.e., confidence judgments and prior knowledge judgments) simultaneously on error correction and accounts for within-person variance. Thus, we could evaluate the relationship between each metacognitive judgment (e.g., knew-it-all-along judgments) and error correction while controlling for the other metacognitive judgment (e.g., confidence judgments). One model was used for the immediate feedback condition, and one model was used for the delayed feedback condition (see Table 2). Each model included the intercept and two predictors, which were centered on each participant's average for that variable. STATA statistical software (StataCorp, 2009) was used for all analyses.

*Immediate feedback model* As is evident in Table 2, for the immediate feedback model, the only reliable predictor of test 2 accuracy was test 1 knew-it-all-along judgments. Specifically, a 1-unit (1–7 scale) increase in knew-it-all-along judgments on test 1 was associated with a 76 % increase in the likelihood of correctly answering that question on test 2. These data suggest that participants' self-rated prior knowledge of a domain had a large influence on test 2 accuracy. However, confidence in errors on test 1 did not reliably predict accuracy on test 2.

*Delayed feedback model* For the delayed feedback model, test 1 knew-it-all-along judgments reliably predicted test 2 accuracy (Table 2). That is, a 1-unit (1–7 scale) increase in knew-it-all-along judgments on test 1 was associated with an 87 % increase in the likelihood of correctly answering that question on test 2. Test 1 confidence judgments were also a reliable predictor for test 2 accuracy such that a 1-unit increase (0 %–100 % scale) in confidence judgments on trial 1 was associated with a 0.9 % increase in the likelihood of correctly answering that question on test 2.

We note that knew-it-all-along judgments and confidence judgments were on different scales. Therefore, in order to facilitate interpretation, we transposed these judgments onto a common scale. Accordingly, a 1-unit increase in the knew-it-all-along judgments, on a 1–7 scale, would be the equivalent to a 14.29-unit increase for the confidence judgments, on a 0–100 scale (i.e., 1 out of 7 is equivalent to 14.29 out of 100). For every 1-unit increase in knew-it-all-along judgments, there is an 87 % increase in the likelihood of answering a question correctly. Thus, a 14.29-unit increase in confidence judgments led to a 12.86 % increase in the likelihood of answering the question correctly on test 2. Put differently,

**Table 2** Logistic heirarchical linear models for test 2 accuracy

| Model | Effect | Estimate (*SE*) | $e^{Estimate} - 1$ | $t$ | 95 % Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Immediate feedback | Intercept | −.25 (.16) | −.22 | 1.63 | −.56 | .05 |
| | Trial 1: Knew-it-all-along judgments | .56 (.06) | .76 | 9.51** | .45 | .68 |
| | Trial 1: Confidence judgments | .004 (.004) | .004 | 1.02 | −.004 | .01 |
| Delayed feedback | Intercept | −.20 (.14) | −.18 | 1.48 | −.47 | .07 |
| | Trial 1: Knew-it-all-along judgments | .63 (.06) | .87 | 10.46** | .51 | .75 |
| | Trial 1: Confidence judgments | .009 (.004) | .009 | 2.31* | .001 | .02 |

*Note.* *SE*s provided in parentheses.

\* $p < .05$

\*\* $p < .001$

**Table 3** Proportions of initial errors corrected, knew-it-all-along judgments for errors, and confidence judgments for errors, as a function of question difficulty for Experiment 3

| Question Difficulty | Mean Number of Errors on Test 1 | | Proportion of Errors Corrected | | Knew-It-All-Along Judgments for Errors | | Confidence Judgments for Errors | |
|---|---|---|---|---|---|---|---|---|
| | Immediate Feedback | Delayed Feedback | Immediate Feedback | Delayed Feedback | Immediate Feedback | Delayed Feedback | Immediate Feedback | Delayed Feedback |
| Easy | 3.50 (.31) | 3.56 (.34) | .80 (.04) | .84 (.03) | 4.41 (.28) | 4.66 (.28) | 15.66 (3.36) | 13.67 (2.91) |
| Medium | 5.75 (.40) | 5.72 (.36) | .57 (.05) | .63 (.05) | 3.21 (.21) | 3.50 (.23) | 16.39 (2.13) | 13.91 (1.81) |
| Hard | 8.94 (.22) | 9.09 (.18) | .26 (.04) | .25 (.03) | 1.99 (.14) | 2.22 (.14) | 12.74 (1.78) | 14.15 (1.51) |

*Note.* Standard error of the mean is presented in parentheses. Knew-it-all-along judgments were made on a scale of 1–7, with lower values denoting lower levels of prior knowledge.

if an item is answered incorrectly on test 1 but given a prior knowledge rating of 3, that item is 87 % more likely to be correct on test 2 than if the item had been given a prior knowledge judgment of 2. Similarly, if an incorrect item on test 1 was given a confidence judgment of 50 %, a participant is about 13 % more likely to get that item correct on test 2 than if the initial confidence judgment had been approximately 35 %. These data suggest that participants' prior knowledge had a substantial influence on test 2 accuracy, whereas participants' confidence on test 1 had a much smaller influence on test 2 accuracy.

## Knew-it-all-along judgments and question difficulty

Nelson and Narens's (1980) norms provide an indication of the relative level of difficulty for each question used in Experiment 3. In general, participants should have higher levels of prior knowledge for easier questions than for more difficult questions. Therefore, according to the knew-it-all-along account, participants should be more likely to correct errors for questions rated as easy, followed by medium questions and, lastly, hard questions, regardless of confidence. Thus, we analyzed knew-it-all-along judgments for incorrect responses on test 1 (see Table 3).

One-way ANOVAs of knew-it-all-along judgments for the immediate feedback condition, $F(2, 62) = 55.96$, $p < .001$, $\eta^2_p = .64$, and the delayed feedback condition, $F(2, 62) = 44.15$, $p < .001$, $\eta^2_p = .59$, indicated that judgments varied as a function of question difficulty. Errors produced on easy general knowledge questions were given reliably higher knew-it-all-along judgments than ratings after errors on medium-level questions. For the easy and medium questions, knew-it-all-along ratings were reliably greater than judgments for hard questions (all $ps < .001$; see Table 3 for descriptive statistics). Thus, participants had greater levels of prior knowledge for easier questions than for harder questions.

We also examined error correction as a function of question difficulty. In line with the prior knowledge explanation of hypercorrection, participants should be more likely to correct errors on easy questions, followed by medium questions and, finally, hard questions. In both the immediate feedback, $F(2, 62) = 64.24$, $p < .001$, $\eta^2_p = .67$, and delayed feedback, $F(2, 62) = 85.35$, $p < .001$, $\eta^2_p = .73$, conditions, error correction varied as a function of question difficulty. In both conditions, participants corrected a reliably greater proportion of errors on easy questions than on medium questions. In addition, a reliably greater proportion of errors were corrected for easy and medium questions, as compared with hard questions (all $ps < .001$).

Finally, we examined confidence judgments on errors as a function of question difficulty in a one-way

ANOVA. Confidence judgments in errors on test 1 did not differ on the basis of level of difficulty in the immediate, $F < 1$, or delayed, $F < 1$, feedback conditions. These results suggest that while participants were not more or less confident in their errors as a function of question difficulty, question difficulty impacted both knew-it-all-along judgments and error correction. Participants indicated high levels of prior knowledge and were more likely to correct errors for easier questions, as compared with more difficult questions.

Discussion

As was observed in the previous experiments, hypercorrection did not differ as a function of feedback timing. However, gamma correlations indicated that self-rated levels of prior knowledge were more closely correlated with error correction than were confidence ratings. We further examined the simultaneous influence of these variables in logistic HLM analyses. Those data showed that prior knowledge ratings for errors on test 1 were the only reliable predictor of test 2 accuracy in the immediate feedback condition; confidence was unrelated to error correction. In the delayed feedback condition, confidence played a minimal role in error correction, while prior knowledge was a much stronger predictor of test 2 accuracy for items initially answered incorrectly. We also examined the impact of question difficulty on confidence, knew-it-all-along judgments, and error correction. Participants indicated greater levels of prior knowledge for easier questions, relative to more difficult questions, and were also more likely to subsequently correct errors for easier questions, as compared with more difficult questions. However, confidence judgments for errors did not vary as a function of question difficulty. Overall, the results from Experiment 3 suggest that prior knowledge may play a more fundamental role in error correction than does subjective confidence.

**General discussion**

The three experiments reported examined the role of confidence and prior knowledge in error correction. In Experiment 1, we sought to examine the hypercorrection effect as a function of feedback timing. Regardless of whether feedback was administered immediately or after a delay, participants exhibited similar levels of hypercorrection. In Experiment 2, we increased the interval between the response and feedback in the delayed feedback condition. We also asked participants to recall their initial confidence ratings and measured the accuracy of recalling confidence judgments. Overall, participants were

less likely to accurately recall their initial confidence judgments in the delayed feedback condition, as compared with the immediate feedback condition. However, participants still exhibited similar levels of hypercorrection across feedback timings.

In Experiment 3, we directly compared the influence of confidence and prior knowledge on error correction. As with the previous experiments, hypercorrection did not differ as a function of feedback timing. However, there was a much stronger relationship between prior knowledge and error correction, as compared with confidence and error correction. Indeed, HLM analyses showed that ratings of prior knowledge were associated with a substantial increase in the likelihood of error correction in the immediate (76 %) and delayed (87 %) feedback conditions. Confidence was associated with a substantially smaller increase in error correction (13 %) in the delayed feedback condition but no reliable increase in the immediate condition. In accord with the prior knowledge hypothesis, participants were also more likely to correct errors for easier general knowledge questions, as compared with errors for harder questions.

Prior knowledge as a primary mechanism for error correction

The experiments reported suggest that prior knowledge plays a primary role in error correction. Thus, if participants have more prior knowledge in a specific domain, they are more likely to correct errors than if they have little prior knowledge. Confidence-based accounts of hypercorrection hold that when participants make high-confidence errors, the discrepancy between subjective confidence judgments and objective memory performance leads to increased attention to feedback. The present results suggest that confidence may play a small, unique role in error correction and primarily reflects the level of knowledge. Indeed, models of subjective confidence suggest that confidence consists of a conglomeration of cues from memory, not just response accuracy (see Koriat, 2008, 2012; see also Brewer & Sampaio, 2006). Information such as the ease of generating a response or the consistency of information that comes to mind influences confidence judgments. These factors are typically diagnostic of the level of prior knowledge, such that faster responses are typically generated for information that is well known.

However, confidence may at times be unrelated to the accuracy of information generated. Koriat (2008) has documented questions (termed *consensually wrong*) that consistently yield high-confidence errors across participants. For example, when asked, "what is the capital of Australia?" an erroneous response is likely to quickly come to mind (e.g., Sydney). Although participants typically answer this

deceptive question quickly and with high levels of confidence, their response is wrong because their prior (and highly familiar) knowledge is flawed (Koriat, 2012). Thus, such knowledge begets a flawed but confident response. It is this knowledge of a domain that contributes to (sometimes flawed) confidence and hypercorrection, but confidence itself plays little unique, causal role in hypercorrection. If confidence plays little unique role in hypercorrection, what accounts for the correlation between confidence and error correction? We argue that confidence judgments are largely influenced by prior knowledge. In turn, this prior knowledge is the primary mechanism driving error correction. If a participant has high levels of prior knowledge but answers a question incorrectly, that discrepancy between subjective and objective performance will enhance attention to feedback. Confidence will be related to error correction, because such judgments are largely based on prior knowledge; however, confidence likely plays a minor, unique role in error correction.

Accordingly, we suggest that the weight of the evidence favors an account of hypercorrection that is largely, but not entirely, predicated on prior knowledge. How does this conclusion fit with prior data indicating that high-confidence errors also appear to direct greater attention to the correct response? For example, secondary task performance is more likely to be disrupted for high-confidence errors (Butterfield & Metcalfe, 2006), and participants' source memory accuracy increases for high-confidence errors (Fazio & Marsh, 2009). A confidence-based account of hypercorrection suggests that the discrepancy between confidence and accuracy (e.g., high-confidence errors) increases attention to feedback, leading to higher levels of error correction. The results from the present experiments do not directly address, or argue against, an attention-based mechanism as the proximal cause of error correction. After answering a question incorrectly, participants may increase attention to feedback for items about which they have high levels of prior knowledge, thus facilitating error correction. Accordingly, participants not only may increase attention to feedback after errors accompanied by high levels of prior knowledge, but also may incorporate this new information into memory more efficiently (cf. Bransford & Johnson, 1972).

Feedback timing

Although not the primary purpose of these experiments, our data do bear on the role of feedback timing in memory performance (see Kulik & Kulik, 1988, for a review). While several theorists have held that immediate feedback may reinforce correct responses and allow errors to be corrected as quickly as possible (Pressey,

1950; Skinner, 1954), others have suggested that delayed feedback may reduce interference between feedback and an incorrect response (Kulhavy & Anderson, 1972) and serve as a spaced study opportunity (Butler, Karpicke, & Roediger, 2007; Pashler, Rohrer, Cepeda, & Carpenter, 2007). Findings regarding the superiority of one feedback timing, as compared with the other, have been mixed. In some cases, differences in performance as a function of feedback timing may be due to the setting (e.g., lab vs. classroom; Kulik & Kulik, 1988) or differences in retention interval between feedback and a final test (Metcalfe, Kornell, & Finn, 2009; Smith & Kimball, 2010).

Consistent with the broader literature, our findings regarding the timing of feedback were also mixed across the experiments reported, with immediate feedback leading to superior performance in some experiments but not others. We examined these findings formally via a fixed-effects meta-analysis of the four experiments reported. Overall, the mean weighted effect size for final recognition indicated that performance was reliably poorer for delayed, relative to immediate, feedback (Hedge's $g = -0.12$; 95 % CI: −0.23, −0.02). In addition, participants were also reliably less likely to correct errors following delayed, as compared with immediate, feedback ($g = -0.21$; 95 % CI: −0.38, −0.05). Thus, delayed feedback led to small performance decrements, relative to immediate feedback. Given the discrepancies evident in the literature, we suggest that the impact of feedback timing on retention merits additional research.

Summary and conclusions

Overall, a variety of research on the hypercorrection effect suggests that when a participant is highly confident in an error, he or she is more likely to correct that response on a later test than if he or she was not initially confident in his or her error. Our results indicate that prior knowledge of a domain may be the primary mechanism behind error correction. Although confidence may play a unique role, these data suggest that this role is minimal when compared with the role of prior knowledge. Thus, prior knowledge not only may increase attention to feedback, but also may increase the likelihood that participants can incorporate new information into memory.

**Author Note**   Danielle M. Sitzman and Matthew G. Rhodes, Department of Psychology, Colorado State University. Sarah K. Tauber, Department of Psychology, Kent State University.

Correspondence concerning this article should be addressed to Danielle M. Sitzman, Department of Psychology, Colorado State University, Fort Collins, CO, 80523-1876. E-mail: Danielle.Sitzman@Colostate.edu

# Appendix

**Table 4** The distribution of confidence judgments given to items that were incorrect on test 1

| Experiment | Feedback Condition | Level of Confidence Judgment | | | |
|---|---|---|---|---|---|
| | | 75–100 | 50–74 | 25–49 | 0–24 |
| Exp. 1a | Immediate feedback | 37 | 33 | 30 | 439 |
| | Delayed feedback | 30 | 46 | 30 | 409 |
| Exp.1b | Immediate feedback | 28 | 43 | 30 | 440 |
| | Delayed feedback | 24 | 46 | 27 | 418 |
| Exp. 2 | Immediate feedback | 80 | 88 | 74 | 897 |
| | Delayed feedback | 60 | 93 | 51 | 910 |
| Exp. 3 | Immediate feedback | 44 | 51 | 35 | 438 |
| | Delayed feedback | 38 | 54 | 26 | 461 |

*Note.* Number of errors on test 1 given confidence judgments in each quartile on a 0–100 scale.

# References

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

Brewer, W. F., & Sampaio, C. (2006). Processes leading to confidence and accuracy in sentence recognition: A metamemory approach. *Memory, 14*(5), 540–552.

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*(6), 1238–1244.

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.

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 918–928.

Butterfield, B., & Mangels, J. A. (2003). Neural correlated of error detection and correction in a semantic retrieval task. *Cognitive Brain Research, 17,* 793–817.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high-confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1491–1494.

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition Learning, 1,* 69–84.

Fazio, L. K., & Marsh, F. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review, 16,* 88–92.

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging, 24,* 462–475. doi:10.1037/a0014417

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119*(1), 80–113.

Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 945–959.

Kulhavy, R. W., & Anderson, R. C. (1972). Delayed-retention effect with multiple-choice test. *Journal of Educational Psychology, 63,* 505–512.

Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68,* 522–528.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58,* 79–97.

Metcalfe, J., & Finn, B. (2011). People's hypercorrection of high-confidence errors: Did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 437–448.

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adult's vocabulary learning. *Memory & Cognition, 37,* 1077–1087.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109–133.

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.

Pressey, S. L. (1950). Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *Journal of Psychology, 29,* 417–447.

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

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24,* 86–97.

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.

StataCorp. (2009). *Stata Statistical Software: Release 11*. College Station, TX: StataCorp, LP.

Tauber, S. K., & Rhodes, M. G. (2012). Multiple bases for young and older adults' Judgments of learning (JOLs) in multitrial learning. *Psychology and Aging, 27,* 474–483.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (in press). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods*.