THESIS

THE INTERACTION BETWEEN FEEDBACK TIMING, CONFIDENCE, AND ERROR CORRECTION IN EPISODIC MEMORY

Submitted by

Danielle Sitzman

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2011

Master's Committee:

Advisor: Matthew G. Rhodes

Benjamin A. Clegg
Brian L. Tracy

ABSTRACT

THE INTERACTION BETWEEN FEEDBACK TIMING, CONFIDENCE, AND ERROR CORRECTION IN EPISODIC MEMORY

Prior work has not provided clear conclusions regarding whether immediate feedback or delayed feedback leads to better retention of material. However, these theories have failed to consider how a person's confidence in their response may interact with the timing of feedback. The current experiments examined how the influence of confidence and the processing of feedback varies as a function of feedback timing. In experiment 1, participants studied a list of word pairs and were given a cued recall test. After the test, participants either received immediate feedback, delayed feedback, or no feedback. Two days later, participants returned to complete another cued recall test for the word-pairs they learned during session 1. Participants receiving feedback performed better on the final test than participants who did not receive feedback, yet the timing of the feedback did not lead to differences in accuracy or confidence ratings. The second experiment followed the same procedure as experiment 1, with the exception that participants were allowed to control the amount of time they spent processing their feedback. Consistent with experiment 1, there were no differences between immediate and delayed feedback in terms of accuracy or confidence ratings on test 2. In addition,

participants spent more time processing feedback when their answers were incorrect rather than correct. However, the timing of feedback did not impact feedback processing time. Overall, while feedback is beneficial for both memory and metamemory accuracy, the timing of feedback does not appear to differentially affect performance.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Providing feedback is not only beneficial for strengthening correct responses but also crucial for correcting errors. However, despite nearly a century of investigation, research examining the impact of feedback on memory performance has provided few definitive guidelines regarding the best methods to enhance the likelihood that information will be remembered in the future. One factor which has only rarely been examined is the impact of a person's subjective confidence and its role in how people attend to feedback. The experiments reported examined the timing of feedback and how the effectiveness of that feedback may be impacted by the errors people make and the confidence they have in their answers.

Theories of Feedback Timing

Feedback has two purposes: 1) to reinforce correct responses in order to enhance retention (Butler, Karpicke, & Roediger, 2008; Pressey, 1950; Skinner, 1954), and 2) to correct errors so that they are not strengthened and later mistakenly remembered (Kulhavy, 1977; Kulhavy & Anderson, 1972; Kulhavy, Yekovich, & Dyer, 1976). However, it is unclear whether feedback is more beneficial when given immediately or after a delay. Most early work examining the role of feedback in learning focused on immediate feedback, suggesting that feedback worked as a reinforcer, strengthening the connection between a question and answer. More recently, feedback has been regarded as

a study opportunity and, consistent with the spacing effect (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), construed as being more beneficial when given at a delay.

Within the literature on feedback and memory there is a lack of consistency regarding the definitions of immediate and delayed feedback (see also Kulik & Kulik, 1988). For the purposes of the experiments reported, immediate feedback refers to the presentation of the correct response immediately after a person provides an answer. Delayed feedback refers to item-by-item re-presentation of the correct response after a person has completed an entire test.

Immediate Feedback

Early work on feedback focused on the importance of immediate feedback in reinforcing correct responses (Pressey, 1950; Skinner, 1954). According to reinforcement theories, feedback must be presented in close temporal proximity to a person's response. Thus, when an answer is correct, immediate feedback will reinforce the response and increase the likelihood of responding the same way in the future. However, when an incorrect response is provided, errors must be corrected immediately to avoid rehearsal, and subsequent strengthening, of the incorrect answer (Metcalfe, Kornell, & Finn, 2009; Pressey, 1950).

A number of studies have shown that immediate feedback produces superior performance compared with delayed feedback or a condition given no feedback. Some of the largest effects of immediate feedback have been shown in studies conducted in the classroom (i.e., applied studies). For example, Kulik and Kulik's (1988) meta-analysis of the feedback literature reported that out of 11 classroom studies, 9 showed an advantage for immediate feedback over delayed feedback. These findings are consistent with recent

work. In a classroom experiment, Dihoff, Brosvic, and Epstein (2003) compared item-by-item feedback (immediate feedback after each response), end of test feedback, 24-hour delayed feedback, and two no feedback conditions. Participants completed a total of five quizzes, one for each feedback condition, for information from an undergraduate psychology course. Performance was measured at the end of the semester on a final test consisting of 10 questions from each of five quizzes. Overall, Dihoff et al. observed that immediate feedback led to better retention of correct responses and better correction of incorrect responses compared with all other feedback conditions (Dihoff et al. 2003). Other studies have likewise reported superior performance for immediate feedback compared with delayed feedback or no feedback at all (Brosvic, Epstein, Cook, & Dihoff, 2005). Dihoff et al. (2003) suggest that providing immediate feedback, particularly when a student answers incorrectly, creates confusion between the answer given and the correct answer, and thus prevents a student from committing to their incorrect response. To resolve this confusion, the student will then assimilate the correct response provided by feedback. Thus, proponents of immediate feedback suggest that errors need to be corrected, and correct answers need to be reinforced, as quickly as possible.

Delaying Feedback

Others argue that principles of immediate reinforcement are not sophisticated enough to address the complexities of human memory. Substantial evidence exists suggesting that delayed feedback leads to superior memory performance compared with feedback administered immediately, particularly on a delayed final test (also known as the *delayed-retention effect* or *DRE*; Butler, Karpicke, & Roediger, 2007; Brackbill, Bravos, & Starr, 1962; Kulhavy & Anderson, 1972; Pashler, Rohrer, Cepeda, &

Carpenter, 2007). Two theories have featured as the foremost explanations of the superiority of delayed feedback: the *interference-perseveration hypothesis* and the *spacing hypothesis*.

Kulhavy and Anderson (1972) proposed the *interference perseveration hypothesis* in an attempt to represent the problems with immediate feedback. According to this theory, the effect of feedback can be explained through an AB-AC paradigm. Participants first learn a stimulus and response pair (AB). On a test, they are shown the stimulus (A) and asked to provide a response (B). In an immediate feedback condition, participants would be shown the correct AB pair. If participants answer correctly (B) and are given feedback (AB), they would receive two congruent study opportunities. However, if participants answer incorrectly (AC), their answer would then interfere with the processing of the corrective feedback (AB). That is, the proactive interference generated by AC would hinder the participants' processing of the feedback (AB). Therefore, Kulhavy and Anderson suggest that delaying feedback leads to superior correction of errors because a delay allows an incorrect answer to dissipate. The dissipation of this error will alleviate interference and allow the correct response to be processed more efficiently.

More recently, researchers have invoked the spacing effect as a way to explain the effectiveness of delayed feedback (Butler et al., 2007; Pashler et al., 2007). Advocates of this idea argue that feedback provides another study opportunity; therefore, delayed feedback will be better than immediate feedback because it presents a spaced study opportunity rather than the massed opportunity available with immediate feedback. Consistent with this perspective, over 100 years of research has shown that retention is

better when learning is distributed over multiple time periods rather than massed in a single session (for a review, see Cepeda et al., 2006). If delayed feedback operates as a spaced study opportunity, then this would explain the DRE.

Butler and Roediger (2008) suggest that the spacing hypothesis and the interference-perseveration hypothesis are not mutually exclusive. For example, it may be that one better explains how to correct errors, while the other better explains how to retain correct responses. When participants are correct on an initial test, they will benefit from delayed feedback because it provides a spaced study opportunity. However, when participants are incorrect on an initial test, immediate feedback creates interference between the incorrect answer provided by the participant and the correct answer (i.e., feedback), and thus hinders the effectiveness of error correction. Delaying feedback will allow errors to dissipate and permits people to focus on encoding the correct answer. The spacing hypothesis thus explains how correct answers can be strengthened, while the interference-perseveration hypothesis explains why errors are better corrected at a delay.

Discrepancies Between Findings for Feedback Timings

Despite the abundance of research on feedback, there is still no definitive answer as to whether immediate or delayed feedback is more effective. Some of the inconsistencies present in the literature may reflect the settings in which the studies were conducted. For example, in a meta-analysis focusing on the feedback timing literature, Kulik and Kulik (1988) reported that immediate feedback led to superior performance when used in the classroom, while delayed feedback proved to be more beneficial in laboratory settings. Although these results are intriguing, Kulik and Kulik note that these discrepancies may be due to attention rather than the setting (i.e., applied versus

laboratory settings). In applied studies, there is less experimental control than in laboratory studies. Thus, when participants receive delayed feedback in the classroom, they are more likely to attend only to the questions they answered incorrectly. However, in the lab, feedback is controlled and participants typically attend to all the feedback for the same amount of time (Kulik & Kulik, 1988). Therefore, the differences seen in Kulik and Kulik's meta-analysis may be due to differences in attention to feedback rather than differences in the feedback itself.

In addition, the efficacy of feedback timing may depend upon the specific goal, whether it is to correct a wrong answer or reinforce a correct one (Butler & Roediger, 2008; see also Metcalfe et al., 2009). Butler and Roediger (2008) suggest that when a person answers a question correctly, they should perform better after delayed feedback but, when a person is incorrect, both feedback timings should be equally effective because each provides a spaced study opportunity. They had participants study 12 passages and complete a multiple-choice test. Each passage was assigned to a different feedback condition and, during the multiple-choice test, participants received either immediate feedback, delayed feedback, or no feedback on each question. One week later participants returned to complete a cued-recall test. Despite theory suggesting that final performance should be best for correct responses given delayed feedback and incorrect responses given immediate feedback, they observed that delayed feedback led to superior performance for both initially correct and incorrect responses.

To summarize, prior research has provided conflicting results regarding the timing of feedback with some studies reporting superiority of delayed feedback over immediate feedback, and other studies reporting that immediate feedback leads to better

performance than delayed feedback. More recently, some work has begun to examine how the type of answer (correct vs. incorrect) may play a role. Thus far, however, no study has provided a direct test of the interaction between feedback timing and the type of response provided (correct vs. incorrect). Butler and Roediger (2008) provide a theory to explain what their results should be, yet, do not provide data that support this theory, or explain why there are deviations. However, there may be other factors, such as a person's confidence in their response that could be contributing to these discrepancies between feedback timings. For example, the influence of response confidence may not be the same for immediate and delayed feedback. In the next section, I will discuss the link between response confidence and feedback.

Confidence and Feedback

Though the goal of feedback is to provide information about the correctness of a person's answer, other factors may impact how people attend to feedback. One factor that has rarely been examined is confidence (but see Butler et al., 2008; Butterfield & Metcalfe, 2001, 2006; Kulhavy et al., 1976; Kulhavy & Stock, 1989). Confidence refers to a participant's subjective assessment regarding the correctness of their response and can be examined in terms of absolute accuracy and relative accuracy. *Absolute accuracy* refers to the correspondence between an overall mean confidence rating and the overall mean level of performance on a test (e.g., an average confidence rating of 90% compared to a mean level of performance of 75% would mean a person was overconfident). Relative accuracy (also termed *resolution*) refers to how well a participant can distinguish between correct and incorrect responses. High levels of relative accuracy would be apparent if the correlation between response confidence and accuracy is high. This would

suggest that participants are giving high confidence ratings to correct items and low confidence ratings to incorrect items.

Prior research has examined how a person's confidence can influence their attention to feedback (Butterfield & Metcalfe, 2001; 2006; Kulhavy et al., 1976; Kulhavy & Stock, 1989). However, this research has only examined how confidence interacts with attention to immediate feedback; no prior study has examined whether this relationship changes if feedback is administered at a delay. It may be that confidence does not have the same influence if feedback is administered after a delay.

The accuracy of a response and the level of confidence in that response may mediate attention to feedback. Errors committed with low levels of confidence may be attended to differently than errors committed with high levels of confidence. Early work by Kulhavy et al. (1976) suggested that feedback only impacted answers that were accompanied by high levels of confidence; however, feedback made no impact on low confidence responses. They had participants study a 30-page booklet detailing the structure and function of the eye. Each page contained a multiple choice question relating to the information that was just presented. All participants circled the answer they deemed correct and rated their confidence in the correctness of that answer on a scale from 1 (low confidence) to 5 (high confidence). Those in the feedback condition were also given a sheet that corresponded with the questions they were answering in the booklet. After participants circled their answer in the booklet, they were also asked to erase the corresponding circle on the sheet provided. If their answer was correct, a *T* would appear; however, if they were incorrect, a different letter appeared. All participants recorded the amount of time they spent on each page, and those in the

feedback condition also recorded the amount of time they spent on the feedback procedure for each page[1]. Overall, Kulhavy et al. (1976) found that feedback was most effective when participants were highly confident in their answer for both correct and incorrect responses. Participants were more likely to maintain high confidence in correct responses (from the initial to final test), more likely to correct errors held with high confidence, and devoted more time to studying the feedback for such high confidence errors.

Butterfield and Metcalfe (2001; 2006) likewise observed that participants were more likely to correct confidently-held errors, a finding they termed the *hypercorrection effect*. They had participants answer general knowledge questions and rate their confidence in their answer. Following their confidence rating, participants were given feedback, confirming whether an answer was correct or displaying the correct response if an answer was incorrect. After a 5 minute distractor task, participants were then retested on the same questions. The gamma correlation between an initial error that was corrected on test 2 and the confidence in that initial error was positive (G = .36), suggesting that high confidence errors were more likely to be corrected than lower confidence errors (Butterfield & Metcalfe, 2001). The authors postulated that people attended to feedback more for highly confident errors. This greater attention, in turn, led to better correction for high confidence errors than low confidence errors on the final test.

In a follow-up experiment examining the allocation of attention to feedback, Butterfield and Metcalfe (2006) found that people performed more poorly at a tone detection task when detecting tones while processing feedback for high confidence errors.

[1] Times were recorded to the nearest 5 seconds from a clock in the front of a room.

Consistent with this, Fazio and Marsh (2010) showed that participants were more likely to remember the color of the font feedback was presented in (red or green) for high-confidence errors. Thus, feedback on high confidence errors may be more likely to capture attention and therefore leads to more sustained processing and correction (Butterfield & Metcalfe, 2001; 2006; Fazio & Marsh, 2010; Kulhavy & Stock, 1989).

However, Butler et al. (2008) suggest that low confidence responses can benefit from feedback as well. They presented participants with general knowledge questions and asked them to pick one of the four alternatives as an answer. Participants then rated their confidence on a scale of 1 (guess) to 4 (high confidence). After rating their confidence, those in the feedback condition were shown the correct answer, while those in the no feedback condition answered the next question. All participants were given a final cued-recall test five minutes after the initial test. Butler et al. (2008) reported that low confidence correct responses were more likely to be recalled on a later test if participants received feedback than when no feedback was given. In addition, they observed a hypercorrection effect such that high confidence errors were more likely to be corrected on a final test than low confidence errors (the hypercorrection effect was not replicated in their Experiment 2).

Based on these data, Butler et al. (2008) suggested that feedback does not work exclusively for high confidence or low confidence responses, but that it is processed differentially when there is a discrepancy between a person's actual response and their subjective assessment of that response (see also Kulhavy & Stock, 1989). Accordingly, people should allocate the most attention to feedback given for high confidence errors and low confidence correct responses. It should also be noted that feedback increased the

accuracy of participants' confidence judgments on a final test. That is, after receiving feedback on the initial test, participants were better able to give high confidence ratings to correct responses and low confidence ratings to incorrect responses on the final test (Butler et al., 2008). Thus, feedback increased both memory and metamemory accuracy.

Overall, research suggests that confidence plays a large role in how people attend to feedback. However, all prior work has examined the role of confidence using only immediate feedback and has not explored whether the relationship between confidence and feedback when feedback is delayed. For example, immediate feedback is typically provided immediately after a participant rates their confidence in their answer. This suggests that assessed confidence following immediate feedback should be fairly accessible and impact how one attends to feedback. However, when feedback is delayed, confidence may be less salient, and therefore, may have less influence on how one attends to feedback and on subsequent memory performance. To date, no prior study has addressed how correctness of a response and a person's confidence in his or her answers may interact with the timing of feedback. For example, after a delay, does confidence still impact how a person approaches feedback? Does their approach to feedback change depending on whether they answered correctly? With immediate feedback, are people more likely to attend to information when there is a discrepancy between confidence in an answer and the actual answer? Is this relationship the same with delayed feedback? I examined these questions in the experiments reported.

CHAPTER 2: EXPERIMENT 1

The goal of the first experiment was to examine whether there would be differences in learning for immediate compared with delayed feedback as a function of confidence and the likelihood of producing errors of commission. Participants in experiment 1 studied pairs of words and completed an initial test of their memory for those pairs. On the first test participants attempted to recall the target (given the cue word) and rated their confidence in that answer. Groups differed on the initial test based on the type of the feedback they received, either immediate feedback, delayed feedback, or no feedback. Prior research on the delayed-retention effect (DRE) suggests that the benefits of delayed feedback are more prominent after a delay (Brackbill et al., 1962; Kulhavy & Anderson, 1972). Thus, participants returned to complete the final test 48 hours after the first test. Of importance, the word pairs used in the current study differed in the likelihood of inducing errors held with high confidence.

Participants studied 72 word pairs, one third of which were related and two-thirds of which were unrelated. There were two types of unrelated pairs. One type, deceptive pairs (e.g., *nurse-dollar*), were used in order to facilitate high confidence errors. When a retrieval cue was presented, there was a semantically related competitor that easily came to mind (e.g., *doctor* when *nurse-do_ _ _ r* is presented). However, the actual target was always an unrelated word (e.g., *dollar*). The target word and the competitive alternative always contained the same first two and last letters. For the other type of unrelated pairs,

control pairs, there was no semantically related competitor that was easily accessible when given the retrieval cue (e.g., *officer-sh_ _ p).* Prior work has shown that these deceptive pairs lead to errors (e.g., "doctor" in the case of *nurse-do_ _ _ r*) that are held with high levels of confidence (Kelley & Sahakyan, 2003; Rhodes & Kelley, 2005). In addition, while I anticipated that participants would make errors on the control items, they were not expected to confidently endorse those errors.

Overall, I expected that participants in the delayed feedback condition would correctly recall more target words than participants in the immediate feedback condition on the final test. Based on previous research it was expected that feedback in general would lead to better performance than no feedback (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). In regard to confidence ratings on test 2, it was expected that participants receiving feedback would provide lower confidence ratings (ratings closer to zero) for incorrect responses than participants who did not receive feedback. In addition, I examined the finding that errors held with high confidence are more likely to be corrected on a subsequent test (i.e., the *hypercorrection effect*: Butterfield & Metcalfe, 2001, 2006). I anticipated that in the immediate feedback condition, participants would be more likely to correct high confidence errors. Predictions were less clear for the delayed feedback condition. Because confidence judgments are given several minutes prior to feedback, the *hypercorrection effect* may be weaker for participants given delayed feedback.

Method

Participants

One hundred and twenty students from Colorado State University participated for partial course credit. Forty participants were randomly assigned to each of the three feedback conditions.

Design

A 3 (word pair type: control, deceptive, related) X 3 (feedback type: immediate, delay, no feedback) mixed-factor design was used. Word pair type was manipulated within-subjects while feedback type was manipulated between-subjects.

Materials

Materials consisted of 72 word pairs, half of which were related filler items (e.g., *ice-water)* and half of which were unrelated pairs. Pairs were taken from Kato (1985) and Kelley and Sahakyan (2003). Additional word pairs were created following the procedures outlined by Kato and Kelley and Sahakyan. There were two versions of the unrelated pairs: deceptive pairs and control pairs. Deceptive pairs (e.g., *nurse-dollar*) were used to facilitate high confidence errors. When a retrieval cue was presented, there was a semantically related competitor that could easily come to mind (e.g., *doctor* when *nurse-do_ _ _ r* is presented). However, the actual target was always an unrelated word (e.g., *dollar*). The target word and the competitive alternative always contained the same first two and last letters. Control pairs were created by randomly pairing the target words with other cues (e.g., *nurse-sharp* and *officer-dollar).* Therefore, all target words were presented equally often in control pairs and deceptive pairs.

Procedure

In the first phase of the experiment, all of the participants studied 72 word pairs at a 4s rate with a 500 ms inter-stimulus interval (ISI). The first 3 and last 3 pairs were control pairs used as primacy and recency buffers and were excluded from all analyses reported. Study lists were randomized anew for each participant. Participants were informed that they should seek to remember the target word given the cue and three letters of the target word on a later cued recall test, and were provided with an example of a test cue (e.g. KITE-CE_ _ _R). After studying all pairs, participants completed a 5 min math filler task before moving on to the first test. Participants were informed that they would be shown the first word (cue) and asked to recall the second word (target).

For test 1, participants were tested on 74 items, each consisting of a cue word and three letters of the target word. The first 2 cue words (taken from the primacy and recency buffers) served as practice items and were not included in any analyses. All participants were shown the cue word and three letters of the target word (e.g. KITE-CE_ _ _ R) and given 10 s to write down the entire word pair (e.g. KITE-CENTER). Next, participants rated their level of confidence in the correctness of their answer on a scale from 0-100, with a rating of 0 indicating that they were not confident at all in the correctness of their answer and a rating of 100 indicating that there was a 100% probability their response was correct. Participants were given 4 s to make their confidence rating.

Feedback differed depending on whether participants were in the immediate or delayed condition. After rating their confidence, participants in the immediate feedback condition were shown the original word pair they studied (e.g. KITE-CENTER) and

asked to indicate whether they answered correctly by circling YES or NO on the answer sheet provided[2]. In the delayed feedback condition, participants proceeded to the next cue after rating their confidence. After providing an answer for all of the cues, participants in the delayed feedback condition were presented with each of the original studied pairs (feedback was presented, on average, 12 minutes after the initial response). As with the immediate feedback group, participants circled YES or NO to indicate the correctness of their response. The order of words presented at delayed feedback was the same as at test. Therefore, participants were able to examine their responses in the same order. For both the immediate and delayed feedback conditions, feedback was presented on the screen for 5 s. In the no feedback condition participants provided an answer and confidence judgment for each test item but were not shown the correct answer for any of their responses. Upon completing test 1, all participants were dismissed until session 2.

After a 48-hour delay, participants returned to complete the experiment. Test 2 was the same as test 1 with the exception that participants were not given feedback. All participants were presented with 72 cues, one at a time, and asked to write down the word pair and provide a confidence estimate. Items on the final test were presented in a fixed-random sequence that was different than the first test.

Results & Discussion

In the analyses that follow I first examined the proportion of targets correctly recalled and then the overall level of confidence participants had in correct and incorrect answers. Finally, I examined the relationship between confidence and accuracy. The alpha level was set to .05 for all analyses reported.

[2] Participants correctly circled YES or NO 99% of the time.

Percent of Targets Correctly Recalled

Test 1. All analyses of test 1 data were done using a 2 (Item type: control, deceptive) X 3 (Feedback type: no feedback, immediate feedback, delayed feedback) mixed-factor analysis of variance (ANOVA). See Figure 1 for the percentage of correctly recalled target words. Following prior work (e.g., Rhodes & Kelley, 2005), related word pairs were excluded from all analyses due to ceiling performance[3]. Overall, control items (M = 46.57, SE = 1.62) were correctly recalled reliably more often than deceptive items (M = 27.45, SE = 1.60), $F(1, 117) = 184.78$, $p < .001$, $\eta^2_p = .61$. As would be expected on the initial test, the proportion of targets recalled did not differ among the immediate, delayed, and no feedback conditions, $F < 1$. Finally, item type did not interact with feedback, $F < 1$.

The proportion of items correctly recalled on test 2 was examined using the same factors as the analysis of test 1 (Figure 2). A larger percentage of control pairs (M = 52.18, SE = 1.50) were correctly recalled than deceptive pairs (M = 32.36, SE = 1.722), $F(1, 117) = 224.00$, $p < .001$, $\eta^2_p = .66$. There was a main effect of feedback type, $F(1,117) = 8.93$, $p < .001$, $\eta^2_p = .13$. While Tukey posthoc analyses showed that there was no difference in performance between the delayed and immediate feedback conditions, p = .47, participants in both feedback conditions recalled a reliably greater proportion of

[3] One-way ANOVAs were done to examine the differences between the immediate, delayed, and no feedback conditions for related word-pairs. There was no difference among feedback conditions on test 1 in the percent of related targets correctly recalled, $F(2,117) = 1.44$, $p = .24$. However, on test 2, there was an effect of feedback type, $F(2, 117) = 3.33$, $p = .04$. There was no difference between the percentage of correct response for the immediate and no feedback conditions, $p = .44$, and for immediate compared with delayed feedback, $p = .37$. However, people receiving delayed feedback retained more correct responses than those receiving no feedback, $p = .03$.

correct responses than those in the no feedback condition, $p < .01$. The interaction between item type and feedback type was not significant, $F < 1$. Overall, these data suggest two conclusions. First, there was a strong effect of feedback as participants in the immediate and delayed feedback conditions showed superior retention on test 2 than participants in the no feedback condition (see Figure 3 for comparison between test 1 and test 2). Second, the type of feedback provided may not differentially affect memory performance as retention was similar for the immediate or delayed feedback conditions.

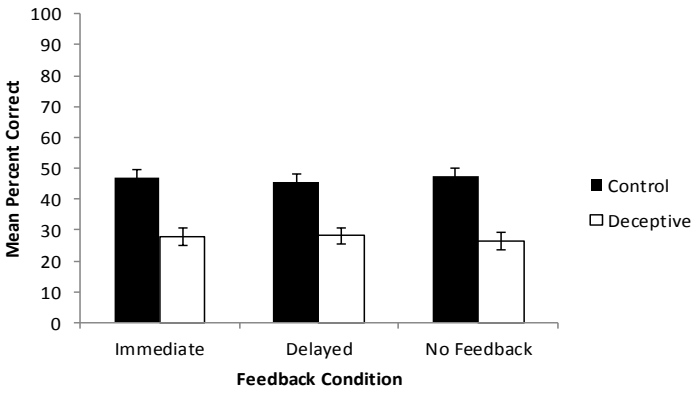

Figure 1. Mean percentage of targets correctly recalled on test 1 for experiment 1.

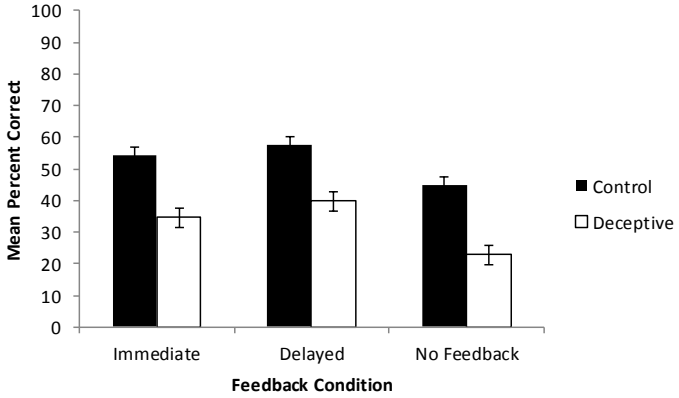

Figure 2. Mean percentage of targets correctly recalled on test 2 for experiment 1.

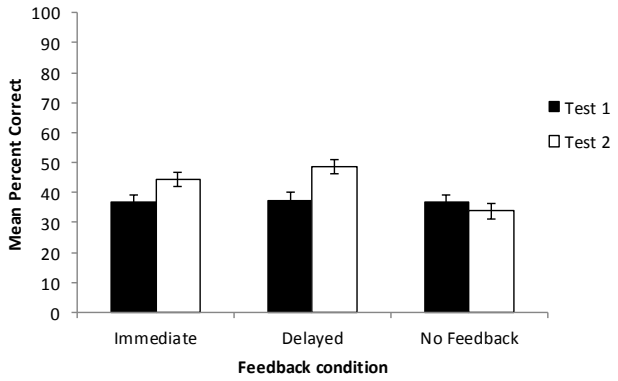

Figure 3. Mean percentage of targets correctly recalled on test 1 and test 2 for experiment 1.

I now turn to analyses conditionalized on performance on test 1 (Table 1). This was done to examine how the type of feedback may have impacted memory for items that were initially correct or incorrect. First, performance on test 2 was examined for items that were correctly recalled on test 1. Control word pairs (M = 86.86, SE = 1.30) answered correctly on test 1 were more likely to be retained on test 2 than deceptive word pairs (M = 75.53, SE = 2.55), $F(1, 110) = 18.17$, $p < .001$, $\eta^2_p = .14$. There was also a main effect of feedback type, $F(1, 110) = 5.24$, $p < .01$, $\eta^2_p = .09$. While there was no difference in the percentage of correct items retained by participants in the immediate feedback and no feedback conditions, $p = .65$, participants receiving delayed feedback retained reliably more correct responses than those in the no feedback condition, $p < .01$, and marginally more responses than those in the immediate feedback condition, $p = .07$. There was no interaction between item type and feedback type, $F < 1$.

Table 1

Percent of targets correctly recalled on test 2 conditionalized by correctness on test 1.

	Immediate Feedback		Delayed Feedback		No Feedback	
	Mean	SE	Mean	SE	Mean	SE
			Experiment 1			
Test 2						
Correct on Test 1						
Control	85.62	2.24	93.87	2.21	80.48	2.30
Deceptive	73.15	4.52	81.91	4.34	71.55	4.52
Incorrect on Test 1						
Control	31.97	2.50	29.94	2.50	13.84	2.50
Deceptive	20.42	2.11	24.31	2.11	3.81	2.11
			Experiment 2			
Test 2						
Correct on Test 1						
Control	90.91	1.98	90.15	1.98		
Deceptive	78.08	4.36	86.77	4.36		
Incorrect on Test 1						
Control	38.33	3.58	28.36	3.58		
Deceptive	27.78	4.12	27.24	4.12		

Note. SE = standard error.

Data were also conditionalized to examine the fate of items on test 2 that were initially incorrect on test 1. Control items that were incorrectly recalled on test 1 were more likely to be correctly recalled (M = 25.25, SE = 1.45) on test 2 than deceptive items (M = 16.18, SE = 1.22), $F(1, 117) = 35.31$, $p < .001$, $\eta^2_p = .23$. The type of feedback given for incorrect responses on test 1 differentially impacted the number of responses correctly recalled on test 2, $F(1, 117) = 29.39$, $p < .001$, $\eta^2_p = .33$. While there was no difference in the correction of errors between the immediate and delayed feedback conditions, $p = .94$, both feedback conditions produced better error correction than the no feedback condition, $p < .001$. There was no interaction between item type and feedback condition, $F(2, 117) = 1.36$, $p = .26$, $\eta^2_p = .02$. Thus, feedback is beneficial for correcting

errors and maintaining correct information. The results also suggest that participants may be better able to retain correct responses if they are given delayed feedback compared with immediate feedback, although this difference was only marginally reliable. Finally, although there was no difference in error correction between feedback conditions, receiving feedback lead to superior error correction relative to no feedback.

Confidence

Confidence in the correctness of responses was also analyzed using a 2 (Item Type: control, deceptive) X 3 (Feedback type: immediate, delayed, no feedback) mixed-factor ANOVA. Confidence in responses on test 1 and test 2 was examined for both correct and incorrect responses (see Table 2).

Table 2

Mean Confidence Ratings conditionalized by correctness for Experiment 1

	Immediate Feedback		Delayed Feedback		No Feedback	
	Mean	SE	Mean	SE	Mean	SE
Test 1						
Correct Items						
Control	80.35	2.49	79.17	2.45	84.54	2.55
Deceptive	88.43	2.54	86.91	2.50	91.07	2.61
Incorrect Items						
Control	18.39	2.55	20.60	2.55	22.94	2.55
Deceptive	48.32	3.37	60.07	3.37	65.04	3.37
Test 2						
Correct Items						
Control	83.12	2.78	81.90	2.67	78.59	2.78
Deceptive	86.08	2.56	86.07	2.46	87.97	2.56
Incorrect Items						
Control	18.85	2.75	23.55	2.75	27.33	2.75
Deceptive	50.44	3.47	50.54	3.47	71.65	3.47

Test 1. Participants gave reliably higher confidence ratings to correctly recalled deceptive items (M = 88.80, SE = 1.47) than control items (M = 81.35, SE = 1.44), F(1,

110) = 25.76, $p < .001$, $\eta^2_p = .19$). There was no reliable difference among feedback conditions, $F(1, 110) = 1.25$, $p = .29$, $\eta^2_p = .02$, nor did feedback interact with item type, $F < 1$. For confidence ratings on incorrect responses, participants assigned higher levels of confidence to deceptive items ($M = 57.81$, $SE = 1.94$) than control items ($M = 20.98$, *SE = 1.47*), $F(1, 117) = 381{,}80$, $p < .001$, $\eta^2_p = .77$. Confidence ratings for incorrect items also differed by feedback condition, $F(1,117) = 5.07$, $p < .01$, $\eta^2_p = .08$. Participants in the immediate feedback condition provided reliably lower confidence ratings to incorrect items than participants in the no feedback condition, $p < .01$. Confidence ratings in the delayed feedback condition did not reliably differ from either the immediate or no feedback conditions, $p > .05$. In addition, there was a reliable interaction between item type and feedback type, $F(2, 117) = 3.42$, $p = .04$, $\eta^2_p = .06$. While all participants gave similar confidence ratings for control items, $F(2,117) = 1.20$, $p = .31$, those in the immediate feedback condition gave lower confidence ratings for deceptive items than participants receiving either delayed feedback or no feedback $F(2,117) = 6.51$, $p < .01$.

Test 2. Confidence ratings for test 2 were analyzed in the same manner as confidence ratings for test 1 with confidence for correct and incorrect responses examined separately. For correct responses, deceptive items ($M = 86.37$, $SE = 1.46$) were given reliably higher confidence ratings than control items ($M = 81.20$, *SE = 1.58*), $F(1, 111) = 14.74$, $p < .001$, $\eta^2_p = .18$. There was no reliable difference among feedback conditions, $F < 1$, and item type and feedback type did not interact, $F(2, 111) = 2.41$, $p > .05$. For incorrect responses, participants provided higher confidence ratings for deceptive word pairs ($M = 57.54$, $SE = 2.00$) than control pairs ($M = 23.24$, $SE = 1.59$), $F(1,117) = 264.35$, $p < .001$, $\eta^2_p = .69$. There was also a reliable difference among feedback

conditions, $F(1, 117) = 758.72$, $p < .001$, $\eta^2_p = .14$. Although there was no difference for confidence ratings between participants in the delayed and immediate feedback conditions, both feedback conditions gave reliably lower confidence ratings to incorrect items than participants in the no feedback condition, $p < .01$. There was also a reliable interaction between item type and feedback condition, $F(2, 117) = 6.04$, $p < .01$, $\eta^2_p = .09$. For control items, participants gave similar confidence ratings regardless of feedback timing, $F(2,117) = 2.39$, $p = .10$. However, on deceptive items, participants receiving immediate or delayed feedback provided lower confidence ratings than participants in the no feedback condition, $p < .001$, $F(2,117) = 12.39$, $p < .001$. There was no difference between immediate and delayed feedback on confidence ratings for deceptive items, $p = 1.00$. Overall, these data suggest that while feedback may not directly impact confidence ratings for correct items, feedback does diminish confidence for incorrect responses. However, it appears that the timing of feedback does not play an integral role in a person's ability to adjust their confidence.

Relationship between accuracy and confidence judgments. Goodman-Kruskal gamma correlations, which have been used as a nonparametric measure of metacognitive accuracy (Nelson, 1984), were calculated to examine the relationship between accuracy and confidence ratings given for a specific answer. Correlations fall within a range of -1 to 1. Positive gamma correlations would indicate that a participant is more likely to provide higher confidence ratings to items they answered correctly and lower confidence ratings to items answered incorrectly. A negative gamma correlation would indicate that a participant is more like to give low confidence ratings to correct items and higher confidence ratings to incorrect items. Gamma correlations assess the relative accuracy

(resolution) of confidence judgments. Resolution refers to the ability to distinguish between correct and incorrect responses by assigning high levels of confidence to correct items and low levels of confidence to incorrect items. Stronger correlations suggest higher levels of resolution.

Gamma correlations (see Table 3) were analyzed using a 2 (Item Type: control, deceptive) X 3 (Feedback type: immediate, delayed, no feedback) mixed-factor ANOVA. For test 1, correlations between accuracy and confidence were stronger for control items (G = .86, SEM = .02) than deceptive items (G = .72, SEM = .04), $F(1,110) = 13.89$, $p < .001$, $\eta^2_p = .11$. As expected, there was no difference in gamma correlations between feedback conditions on test 1, $F < 1$. There was no significant interaction between item type and feedback condition, $F < 1$.

Table 3

Gamma Correlations between Confidence Judgments and Accuracy for Experiment 1

	Immediate Feedback		Delayed Feedback		No Feedback	
	G	SE	G	SE	G	SE
			Experiment 1			
Test 1						
Control Items	0.88	0.03	0.85	0.03	0.84	0.04
Deceptive Items	0.78	0.07	0.70	0.06	0.70	0.07
Test 2						
Control Items	0.89	0.32	0.87	0.03	0.77	0.03
Deceptive Items	0.68	0.07	0.71	0.07	0.50	0.07

Note. G = Gamma Correlation; SE = Standard Error.

On test 2 there was also a stronger correlation between accuracy and confidence for control items (G = .84, SEM = .12) than deceptive items (G = .63, SEM = .04), $F(1, 110) = 31.40$, $p < .001$, $\eta^2_p = .22$. A main effect of feedback type was also evident, $F(1, 110) = 3.84$, $p = .03$, $\eta^2_p = .07$. Specifically, while there was no difference between

participants in the immediate and delayed feedback conditions, $F < 1$, participants given feedback displayed a stronger correlation between accuracy and confidence ratings than those in the no feedback condition, $p < .05$. The interaction was not significant, $F < 1$. These results suggest that feedback may serve to improve resolution relative to a no feedback condition. However, the timing of feedback does not impact resolution.

In line with previous research (Butterfield & Metcalfe, 2001; 2006), the hypercorrection effect was analyzed by examining the correlation between confidence judgments given to incorrect items on test 1 that were subsequently corrected on test 2 (see Figure 4). Positive correlations indicate a hypercorrection effect (i.e., errors with higher confidence are more likely to be corrected on test 2). Due to a small cell size for the no feedback condition ($n = 12$), data were only analyzed for participants in the delayed and immediate feedback conditions. One-sample t-tests were used to test gamma correlations against chance (zero). When collapsing across item type, participants receiving immediate feedback ($G = -.01$, $SE = .08$) did not show the hypercorrection effect, $t < 1$, while participants in the delayed feedback condition ($G = .13$, $SE = .07$) showed a marginal hypercorrection effect, $t(75) = 1.76$, $p = .08$. However, the difference between these two conditions was not reliable, $F(1, 78) = 1.77$, $p = .19$. When data were calculated based on item type (i.e., collapsing across the type of feedback), the hypercorrection effect did not differ between control ($G = .03$, $SEM = .08$) and deceptive ($G = .10$, $SEM = .07$) items, $F < 1$, and neither of these item types were significantly different from zero, $t < 1$. There was a marginal interaction between item type and feedback type, $F(1, 64) = 3.15$, $p = .08$, $\eta^2_p = .05$. While there was no difference between the immediate ($G = .04$, $SEM = .13$) and delayed ($G = .03$, $SEM = .11$) feedback

conditions for control items, there was a trend in the deceptive items for a higher level of hypercorrection when participants were given delayed feedback (G = .23, SEM = .10) compared with immediate feedback (G = .08, SEM = .02), $t(69) = -1.87$, p = .07.

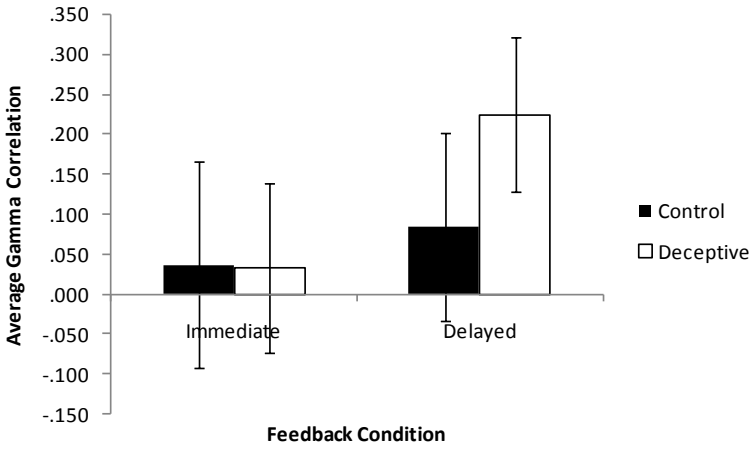

Figure 4. Mean gamma correlations between test 1 confidence and test 2 accuracy (i.e., the hypercorrection effect) for experiment 1.

Discussion

The data from experiment 1 suggest several conclusions. First, it is clear that receiving feedback leads to greater retention of correct responses and greater correction of errors on a final test than no feedback. Though not reliable, there was a trend suggesting that delayed feedback led to a superior retention of correct responses than immediate feedback. For confidence ratings there was also clear evidence that providing feedback allowed participants to more accurately differentiate between correct and incorrect responses than if they do not receive feedback. For example, on test 1, providing immediate feedback led to reliably lower confidence ratings for both correct and incorrect responses compared with delayed feedback or no feedback. Though

confidence ratings were lower in the immediate feedback condition, resolution did not differ among the three feedback conditions on the first test.

In terms of the hypercorrection effect, there is not a clear picture regarding the role of feedback timing. Previous research on the hypercorrection effect has been done using only immediate feedback (Butterfield & Metcalfe, 2001; 2006); however, the data for this experiment did not replicate the hypercorrection effect in the immediate feedback condition. The delayed feedback condition showed a slight hypercorrection effect, but the value was not significantly different from zero. Also, there was no reliable difference between the immediate and delayed feedback conditions in terms of the hypercorrection effect. These findings suggest that hypercorrection may not be robust as other research has also failed to replicate the effect (Butler et al., 2008).

CHAPTER 3: EXPERIMENT 2

The two main goals of experiment 1 were to examine the interaction between feedback timing and confidence, and to examine the differences in how feedback timing may impact sensitivity to correct and incorrect answers. Experiment 2 focused on examining the source of differences between the immediate and delayed feedback conditions. In particular, some of the differences apparent in feedback effectiveness may be due to differences in how people attend to feedback information.

To my knowledge, research examining attention to feedback has only been conducted using immediate feedback. Prior work suggests that the discrepancy between performance and confidence is related to the amount of time spent processing feedback (Butler et al., 2008; Butterfield & Metcalfe, 2001; 2006; Kulhavy et al., 1976; Kulhavy & Stock, 1989). In general, people spend more time processing feedback following incorrect than correct answers. For items answered incorrectly, as a person's confidence in their response increases, so does the amount of time they spend processing feedback. Therefore, people allocate the greatest amount of feedback processing time to errors held with high confidence (Kulhavy & Stock, 1989). However, this relationship has only been examined with immediate feedback; it is unclear whether this relationship would be as strong with delayed feedback.

In Experiment 2, my interest was in examining the differences between immediate and delayed feedback. Therefore, I only used an immediate and delayed feedback

condition (both of which led to better performance than the no feedback condition in Experiment 1). To assess the relationship between confidence and feedback processing time, participants were allowed to self-pace their examination of feedback. It was expected that, in the immediate feedback condition, the amount of time spent processing feedback would be dependent upon the answer and confidence (Kulhavy & Stock, 1989). Participants should spend the most time processing feedback for high confidence incorrect responses, followed by low confidence incorrect responses, low confidence correct answers, and lastly, the least amount of time should be spent processing feedback for high confidence correct responses. However, in the delayed feedback condition, when confidence should be less accessible, there should be less variability in feedback processing time. Participants should spend a similar amount of time processing feedback, regardless of their confidence in a response. Alternatively, given that confidence appeared to exert the same influence for delayed and immediate feedback in experiment 1, it may be that there are no differences in feedback processing times between the two conditions.

Method

Participants

Sixty students from Colorado State University participated for partial course credit. Thirty participants were randomly assigned to each one of the feedback conditions.

Design

A 3 (word pair type: control, deceptive, related) X 2 (feedback type: immediate, delay) mixed-factor design was used. Word pair type was manipulated within-subjects while feedback type was manipulated between-subjects.

Methods/Procedure

The materials were the same as Experiment 1. There were two main differences from the procedure used in Experiment 1. First, participants in both feedback conditions were allowed to self-pace the processing of feedback. When the feedback was presented, participants circled either YES or NO to indicate whether they answered correctly, and then spent as much time as they wanted studying the word pair. Inspection of the feedback was terminated by pressing the space bar. Thus, the time taken to process feedback was measured based on the latency to press the space bar. The second difference was that Experiment 2 did not include a no feedback condition as the variable of interest was time spent processing feedback.

Results & Discussion

Proportion of Targets Correctly Recalled

Test 1. Test 1 was analyzed using a 2 (Item type: control, deceptive) X 2 (Feedback type: immediate feedback, delayed feedback) mixed-factor ANOVA. Performance was assessed by examining the proportion of target words correctly recalled (see Figure 5). Overall, control items (M = 47.13, SE = 2.38) were correctly recalled reliably more often than deceptive items (M = 32.13, SE = 2.27), $F(1, 58) = 62.90$, $p < .001$, $\eta^2_p = .52$. As would be expected on the initial test, the proportion of targets recalled

did not differ between the immediate and delayed feedback conditions, $F < 1$. Finally, item type did not interact with feedback, $F(1, 58) = 2.77$, $p = .10$, $\eta^2_p = .05$.

The proportion of items correctly recalled on test 2 was examined using the same factors as the analysis of test 1. A larger proportion of control pairs ($M = 59.44$, $SE = 2.23$) were recalled than deceptive pairs ($M = 43.61$, $SE = 2.82$), $F(1, 58) = 51.88$, $p < .001$, $\eta^2_p = .47$ There was no difference in accuracy between the immediate and delayed feedback conditions, $F < 1$. The interaction between item type and feedback type was not significant, $F < 1$.

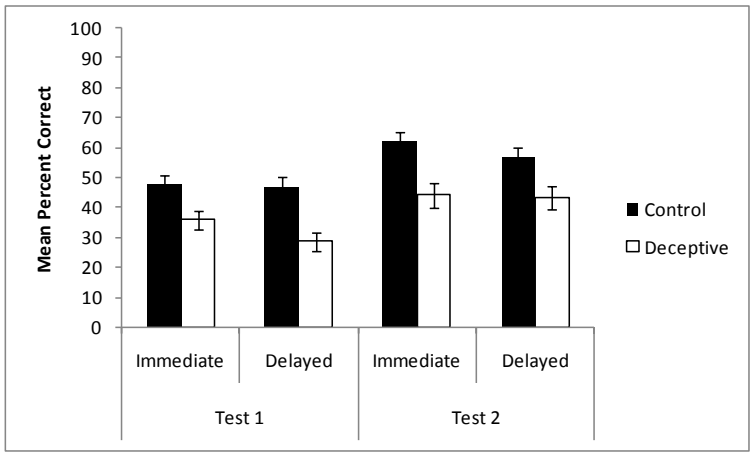

Figure 5. Mean percentage of targets correctly recalled on test 1 and test 2 for experiment 2

As in Experiment 1, data were conditionalized to examine performance based on accuracy during test 1 (see Table 1). First, performance on test 2 was examined for items that were correctly recalled on test 1. Control word pairs ($M = 90.53$, $SE = 1.40$) that were correct on test 1 were more likely to be retained on test 2 than deceptive word pairs ($M = 82.43$, $SE = 3.08$), $F(1, 58) = 8.05$, $p < .01$, $\eta^2_p = .12$. Inconsistent with experiment 1 (where participants were more likely to retain correct responses in the delayed feedback condition), there was no difference in the number of correct responses retained by

participants in the immediate feedback condition compared with participants in the delayed condition, $F(1, 58) = 1.06$, $p = .31$, $\eta^2_p = .02$. Item type did not interact with feedback type, $F(1, 58) = 2.73$, $p = .10$, $\eta^2_p = .05$.

Data were also conditionalized to examine the fate of items on test 2 that were initially incorrect on test 1. Control items that were incorrectly recalled on test 1 were marginally more likely to be correctly recalled ($M = 33.34$, $SE = 2.53$) on test 2 than deceptive items ($M = 27.51$, $SE = 2.91$), $F(1, 58) = 3.42$, $p < .07$, $\eta^2_p = .06$. There was no difference in the correction of errors between participants given immediate or delayed feedback, $F(1,58) = 1.40$, $p = .24$, $\eta^2_p = .02$, nor was there an item type x feedback interaction, $F(1, 58) = 2.24$, $p = .14$, $\eta^2_p = .04$. Overall, these data suggest that feedback timing does not differentially affect the number of correct responses retained or the number of errors corrected.

Confidence

Confidence in the correctness of responses was also analyzed using a 2 (Item Type: control, deceptive) X 2 (Feedback type: immediate, delayed) mixed-factor ANOVA. Confidence on test 1 and test 2 was examined for both correct and incorrect responses (see Table 4 and Figure 6).

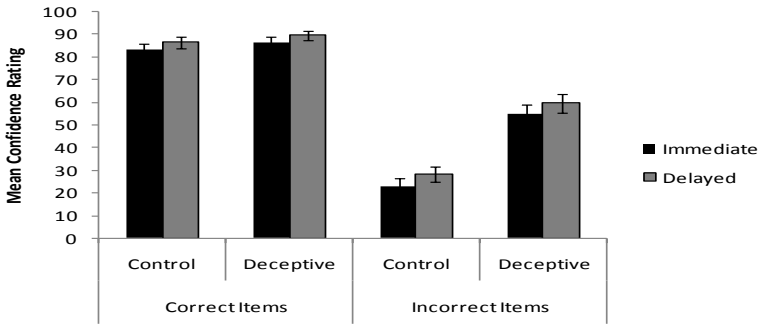

Figure 6. Overall confidence ratings on test 1 for experiment 2.

Test 1. For correctly recalled items, confidence ratings did not differ between deceptive items (M = 85.76, SE = 2.58) and control items (M = 82.86, SE = 1.85), $F(1, 58) = 1.63$, $p < .21$, $\eta^2_p = .03$. However, there was a reliable difference between feedback conditions. Participants in the immediate feedback condition gave reliably lower confidence ratings than those in the delayed feedback condition, $F(1,58) = 4.35$, $p = .04$, $\eta^2_p = .07$. Item type did not interact with feedback condition, $F < 1$.

Table 4
Mean Confidence Ratings conditionalized by correctness for Experiment 2

	Immediate Feedback		Delayed Feedback	
	Mean	SE	Mean	SE
Test 1				
Correct Items				
Control	79.86	2.61	85.85	2.61
Deceptive	80.70	3.65	90.82	3.65
Incorrect Items				
Control	23.12	2.78	26.80	2.78
Deceptive	56.96	3.14	66.57	3.14
Test 2				
Correct Items				
Control	83.43	2.37	86.43	2.37
Deceptive	86.62	2.24	89.60	2.24
Incorrect Items				
Control	23.03	3.35	28.25	3.41
Deceptive	54.85	4.07	59.68	4.14

For confidence ratings on incorrect responses, participants assigned higher levels of confidence to deceptive items (M = 61.77, SE = 2.22) than control items (M = 24.96, SE = 1.97), $F(1, 58) = 244.74$, $p < .001$, $\eta^2_p = .81$. Participants in the immediate feedback condition gave marginally lower confidence ratings to incorrect responses than participants in the delayed feedback condition, $F(1,58) = 3.67$, $p = .06$, $\eta^2_p = .06$, but feedback type did not interact with item type, $F(1,58) = 1.59$, $p = .21$, $\eta^2_p = .03$. Although differences between feedback timings were not expected on test 1, participants

in the immediate feedback conditions provided lower confidence ratings than those receiving delayed feedback. It may be that, in general, receiving immediate feedback leads to an overall decrease in confidence for responses.

Test 2. Confidence ratings for test 2 were analyzed in the same manner as confidence ratings for test 1 with confidence for correct and incorrect responses examined separately. For correct responses, there was no difference between confidence ratings given to deceptive items (M = 88.11, SE = 1.58) and control items (M = 84.93, SE = 1.67), $F(1, 58) = 2.39$, $p = .13$, $\eta^2_p = .04$. There was no reliable difference between the immediate and delayed feedback conditions, $F(1,58) = 1.39$, $p = .24$, $\eta^2_p = .02$, nor was there an interaction, $F < 1$.

For incorrect responses, participants provided higher confidence ratings for deceptive word pairs (M = 57.27, SE = 2.90) than control pairs (M = 25.64, SE = 2.39), $F(1,57) = 117.72$, $p < .001$, $\eta^2_p = .67$. However, there was no reliable difference between participants in the immediate and delayed feedback conditions, $F(1,57) = 1.28$, $p = .26$, $\eta^2_p = .02$, nor an interaction between feedback and item type, $F < 1$. Though confidence ratings on test 1 were lower for participants in the immediate feedback condition compared with those in the delayed condition, participants gave equivalent confidence ratings for correct and incorrect responses on test 2.

Relationship Between Confidence and Performance.

Gamma correlations were calculated to examine the relationship between accuracy and the confidence rating given (see Table 5). These data were analyzed using a 2 (Item Type: control, deceptive) X 2 (Feedback type: immediate, delayed) mixed-factor ANOVA. For test 1, there was a stronger correlation between accuracy and confidence

for control items (G = .87, SEM = .02) than for deceptive items (G =.60, SEM =.07), $F(1,58) = 16.19$, $p < .001$, $\eta^2_p = .22$. As expected, there was no difference between feedback conditions on test 1, $F(1, 58) = 1.34$, $p = .25$, $\eta^2_p = .02$, nor a significant interaction, $F < 1$. While correlations differed by item type, they were not influenced by the timing of feedback.

Table 5
Gamma Correlations between Confidence Judgments and Accuracy for Experiment 2.

		Immediate Feedback		Delayed Feedback	
		G	SE	G	SE
Test 1					
	Control Items	0.85	0.03	0.89	0.03
	Deceptive Items	0.53	0.10	0.67	0.10
Test 2					
	Control Items	0.87	0.03	0.87	0.03
	Deceptive Items	0.60	0.06	0.77	0.07

On test 2 the correlation between accuracy and confidence was stronger for control items (G = .87, SEM = .02) than deceptive items (G = .69, SEM = .05), $F(1, 56) = 15.26$, $p < .001$, $\eta^2_p = .21$. However, there was no difference in gamma correlations between the immediate and delayed feedback conditions, $F(1,56) = 2.88$, $p = .10$, $\eta^2_p = .05$. The interaction was not reliable, $F(1 ,56) = 3.15$, $p = .08$, $\eta^2_p = .05$.

The hypercorrection effect was analyzed by examining the correlation between confidence for incorrect items on test 1 and that item's accuracy on test 2. When collapsing across item type, neither participants in the immediate feedback condition (G = -.02, SE = .08) or participants in the delayed feedback condition (G = .04, SE = .09) showed a hypercorrection effect, $t < 1$. There was no reliable difference between the two feedback conditions, $F < 1$. There was not a hypercorrection effect present for control

items (G = .07, SEM = .09) or deceptive items (G = -.05, SEM = .09), $t < 1$ and there was no reliable difference between the two groups, $F(1,48) = 1.68$, $p = .20$, $\eta^2_p = .03$. There was no interaction between feedback condition and item type, $F < 1$. Thus, the correlations between accuracy and confidence ratings suggest that the timing of feedback does not differentially impact resolution. Also, these results do not replicate the hypercorrection effect for both the immediate and delayed feedback conditions.

Time Spent Reviewing Feedback

Feedback processing times were analyzed using a 2 (Item type: control, deceptive) X 2 (Accuracy: correct, incorrect) X 2 (Feedback Type: immediate, delayed) mixed-factor ANOVA (see Figure 7). Overall, participants spent the a similar amount of time reviewing feedback for control items (M = 3529.57 ms, SE = 155.31) and deceptive items (M = 3696.32 ms, SE = 134.23), $F(1,58) = 2.66$, $p = .11$, $\eta^2_p = .04$. Participants in the immediate feedback condition (M = 3840.38 ms, SE = 192.12) spent more time processing feedback as participants in the delayed feedback condition (M = 3385.51 ms, SE = 192.12), however, this difference was not reliable, $F(1,58) = 2.80$, $p = .10$, $\eta^2_p = .05$. However, participants did study items longer when they answered incorrectly (M = 3936.13, SE = 180.83) relative to correctly (M = 3289.77, SE = 126.48), $F(1,58) = 17.74$, $p < .001$, $\eta^2_p = .24$. There was no reliable interaction between the type of feedback and item type, $F < 1$, or between accuracy and the type of feedback, $F < 1$. The three-way interaction between item type, accuracy, and type of feedback was not significant, $F(1,58) = 2.71$, $p = .11$, $\eta^2_p = .05$. Overall, these data suggest that while people adjust their feedback processing time depending on the correctness of their response, the timing

of feedback does not seem to change the time spent processing feedback. Analysis of median reaction times showed the same results[4].

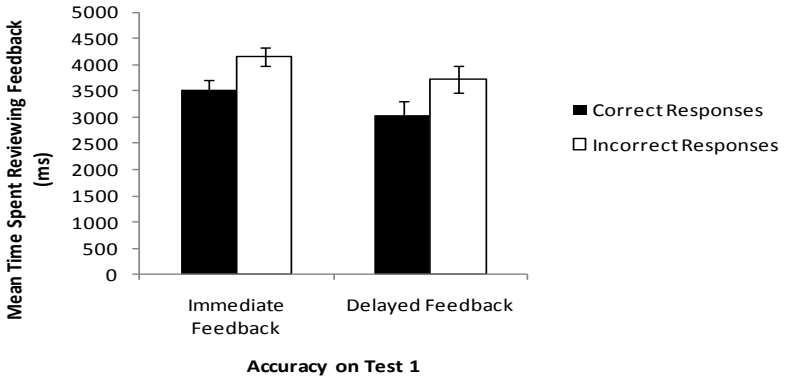

Figure 7. Mean time spent processing feedback based on feedback timing and test 1 accuracy.

Reaction time and accuracy gamma correlations. Gamma correlations were used to examine the relationship between accuracy, feedback processing time, and confidence ratings. All gamma correlations were examined using a 2 (Item Type: control, deceptive) X 2 (Feedback Type: immediate, delayed) mixed-factor ANOVA.

For the reaction time and accuracy gammas, it was assumed that participants would spend a longer amount of time processing feedback for incorrect items than correct

[4] Median feedback processing times were analyzed using a 2 (Item type: control, deceptive) X 2 (Accuracy: correct, incorrect) X 2 (Feedback Type: immediate, delayed) mixed-factor ANOVA. Overall, participants spent a similar amount of time reviewing feedback for control items (M = 3288.73 ms, SE = 149.03) and deceptive items (M = 3395.34 ms, SE = 115.65), $F(1,58) = 1.23$, $p = .27$, $\eta^2_p = .02$. Participants in the immediate feedback condition (M = 3542.53 ms, SE = 175.94) spent an equal amount of time processing feedback as participants in the delayed feedback condition (M = 3141.55 ms, SE = 175.94), $F(1, 58) = 2.60$, $p = .11$, $\eta^2_p = .04$. However, participants studied items longer when they answered incorrectly (M = 3687.75, SE = 163.35) relative to correctly (M = 2996.32, SE = 117.35), $F(1,58) = 25.16$, $p < .001$, $\eta^2_p = .30$. There was no reliable interaction between the type of feedback and item type, $F < 1$, or between accuracy and the type of feedback, $F < 1$. The three-way interaction between item type, accuracy, and type of feedback was not significant, $F(1,58) = 1.41$, $p = .24$, $\eta^2_p = .02$.

items, resulting in a negative correlation between the two variables. On test 1, there was a main effect of item type, such that there was a stronger relationship between accuracy and feedback processing time for deceptive items (*G* = -.42, *SEM* = .05) than control items, (*G* = -.31, *SEM* = .04), $F(1, 58) = 6.18$, $p = .02$, $\eta^2_p = .10$. There was no difference in feedback processing time between the immediate (G = -.38, SEM = .05) and delayed (G = -.35, SEM = .05) feedback conditions, $F < 1$, and there was no interaction, $F < 1$. Thus, there was a stronger relationship between the time spent reviewing feedback and accuracy for deceptive items than control items. This suggests that, compared with control items, participants spent a longer time reviewing feedback for deceptive items when they were incorrect and less time when they were correct.

Reaction time and confidence judgments gamma correlations. Similar to accuracy, confidence should show a negative relationship with feedback reaction time. When a participant is more confident in their response, they should spend less time processing the feedback given for correct items. However, for incorrect items, as confidence increases, so should feedback processing time. This would be consistent with previous research suggesting that people devote more attention to feedback for high confidence errors than low confidence errors (Butterfield & Metcalfe, 2006). For correct responses on test 1, participants spent marginally less time processing feedback as their confidence in their responses increased for both control (*G* = -.16, *SEM* = .08) and deceptive (*G* = -.34, *SEM* = .08) items, $F(1,31) = 3.41$, $p = .07$, $\eta^2_p = .10$. However, there was no difference in the relationship between the time spent reviewing feedback and confidence between the immediate (*G* = -.26, *SEM* = .09) and delayed (*G* = -.24, *SEM* = .09) feedback conditions, $F < 1$. There was no reliable interaction, $F < 1$ (see Figure 8).

For incorrect responses on test 1, participants spent more time processing feedback as confidence increased for both control (G = .16, SEM = .06) and deceptive (G = .15, SEM = .05) items; however, there was no reliable difference between the item types, $F < 1$. This relationship did not change between the immediate (G = .12, SEM = .06) and delayed (G = .20, SEM = .06) feedback conditions, $F(1,56) = 1.13$, $p = .29$. There was not a reliable interaction, $F(1,56) = 1.14$, $p = .21$. While it is clear that there is a relationship between confidence and time spent reviewing, this relationship was not differentially impacted by the timing of feedback.

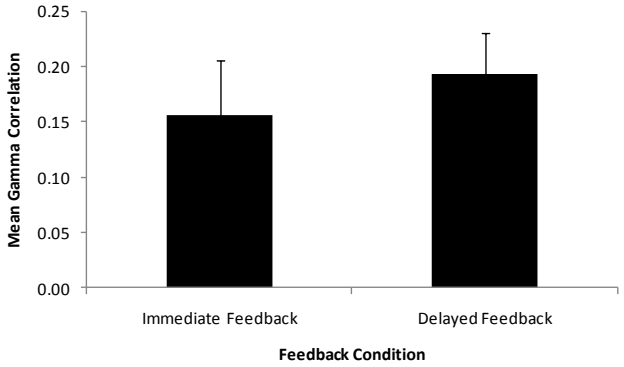

Figure 8. Mean gamma correlation between feedback processing time and confidence for incorrect responses on test 1.

Discussion

In line with experiment 1, there does not seem to be a clear advantage of one feedback timing. Participants were equally likely to correct errors and maintain correct responses regardless of whether they are receiving feedback immediately or after a delay. Feedback timing also did not impact confidence ratings on a final retention test. While participants in the immediate feedback condition had lower levels of confidence overall

on test 1, these differences did not obtain on the final retention test. Participants gave equivalent ratings regardless of feedback condition.

Of greater interest in this experiment, both accuracy and confidence impacted the time participants studied feedback. When participants were incorrect, they spent more time reviewing feedback than when they were correct. For correct responses, the more confident a person was in their response, the less time they spent studying feedback. For incorrect responses, participants spent more time processing feedback as their confidence increased. However, the timing of feedback, whether immediate or delayed, did not impact how long participants reviewed feedback.

CHAPTER 4: GENERAL DISCUSSION

In general, receiving feedback compared with no feedback led to greater retention of correct responses and greater correction of errors. Participants receiving delayed feedback in Experiment 1 were slightly more likely to retain correct responses than those who received immediate feedback. However, neither feedback condition led to greater error correction. This advantage for response retention with delayed feedback was not replicated in Experiment 2. The results regarding confidence and feedback processing times are somewhat inconsistent with both previous research and the hypotheses for this research. In the following sections, I will detail the specific findings from the current experiments and discuss these results within the context of theories of feedback timing and theories explaining the relationship between response confidence and feedback processing time.

Theories of feedback timing

Theories of feedback timing have attempted to explain the difference in performance between feedback given immediately and feedback provided after a delay (Butler et al., 2007; Butler & Roediger, 2008; Kulhavy & Anderson, 1972; Metcalfe et al., 2009; Pashler et al., 2007; Pressey, 1950; Skinner 1954). However, the results of experiments 1 and 2 do not provide support for differences between the two types of feedback. Overall, there was no difference in accuracy on test 2 between the immediate and delayed feedback conditions. However, there was a trend in experiment 1 for

participants in the delayed feedback condition to retain more correct responses than participants in the immediate feedback condition. This is consistent with theories suggesting that delaying feedback works as a spaced study opportunity increasing the likelihood participants will retain correct responses on a future test (Butler et al., 2007; Pashler et al., 2007; Smith & Kimball, 2010). Nevertheless, the trend for delayed feedback to lead to superior response retention was not replicated in experiment 2; therefore, these results should be interpreted cautiously.

Proponents of immediate feedback suggest that feedback is needed immediately in order to correct errors and to reinforce correct responses (Metcalfe et al., 2009; Pressey, 1950: Skinner, 1954). The data from the current experiments do not support these theories of reinforcement. Receiving feedback immediately led to greater accuracy than not receiving feedback; however, there was no benefit for receiving immediate feedback compared with delayed feedback. This suggests that perhaps it is not the immediacy of the feedback that is crucial, but receiving feedback compared with not receiving feedback that is critical for retention.

Several factors may explain the absence of differences between immediate and delayed feedback in these experiments and the lack of consistency in the literature overall. First, the lag to the final test has varied in much prior work. Previous research has suggested that the lag between feedback and the final test may impact the efficacy of feedback timing. For example, participants receiving delayed feedback may have less time (i.e., a shorter retention interval) between feedback and the final test than participants receiving immediate feedback. For example, Smith and Kimball (2010) showed that delayed feedback led to superior accuracy on the final test compared with

immediate feedback; however, when the lag to the final test was controlled for, this difference between feedback conditions was less pronounced. In the current set of experiments, the lag between feedback and final test was controlled, and this may explain the lack of a difference in performance between the immediate and delayed feedback conditions.

Second, prior work also does not provide a set definition of what is considered delayed feedback. Definitions of delayed feedback can vary from providing feedback 10 s after an item is presented to several days after the initial test. These variations in the interval between response and feedback make conclusions difficult. In the current experiments, delayed feedback was given directly after participants answered all the questions on the initial test, an interval of approximately 12 minutes on average. It may be that the delay of feedback was not sufficient to elicit performance differences. Smith and Kimball (2010) suggest that the likelihood of maintaining correct responses continues to increase until the delay to feedback reaches about 20% of the total retention interval. However, beyond that point, there are diminishing returns and once the lag reaches about 40% of the retention interval, the efficacy of delayed feedback starts to decrease. Future research should examine delayed feedback in more detail, especially how it compares with immediate feedback. If various delays lead to differences in performance, it would be important to explore when, and what kind of, delayed feedback will be superior to immediate feedback.

Confidence

Little work has examined the interaction between response confidence and the timing of feedback. While feedback timing did not differentially impact confidence

ratings on test 2, there were several interesting patterns on the first test. When given immediate feedback, participants' confidence ratings were reliably lower, for both correct and incorrect responses, compared with participants given delayed feedback or no feedback. This suggests that participants may be adjusting their confidence as they progress through trials on the first test. It was originally hypothesized that there should be no differences in memory or metamemory performance among the feedback conditions. Yet, differences on test 1 may be explained by the fact that immediate feedback helped participants distinguish between correct and incorrect responses.

If participants are updating their monitoring as they receive immediate feedback, then confidence ratings should change as they progress through the test. Confidence ratings for participants receiving immediate feedback on test 1 (experiment 1) were divided into quartiles (18 items in each). A one-way ANOVA demonstrated that there was a significant effect of item position on confidence ratings, $F(3,68) = 3.53$, $p = .02$. Follow up t-tests showed that confidence on the first quarter of items ($M = 74.51$, $SE = 3.85$) was significantly higher than confidence on the second quarter ($M = 59.33$, $SE = 3.02$), $t(17) = 2.63$, $p = .02$. Confidence ratings in the third quartile ($M = 63.73$, $SE = 3.54$), $t(17) = 1.95$, $p = .07$, and ratings in the fourth quartile ($M = 65.30$, $SE = 3.11$), $t(17) = 1.80$, $p = .09$, were marginally lower than ratings in the first quartile. However, there were no reliable differences between confidence ratings among the second, third, and fourth quartiles (data from experiment 2 follow the same patterns). These data suggest that participants quickly decreased their overall confidence ratings when given immediate feedback, but this pattern did not occur for the no feedback and delayed feedback conditions.

Differences in confidence during an initial test have rarely been examined in the literature. However, in a recent study, Jacoby, Wahlheim, Rhodes, Daniels, and Rogers (2010) found a similar pattern of results in an experiment examining proactive interference and immediate feedback. In the first phase, Jacoby et al. presented participants with two lists of word pairs. List 1 contained 48 word-pairs that were presented 3 times (e.g., knee-bend) that participants were to read aloud. For list 2, participants were told that they would need to remember the word pairs for a later memory test. These words involved the same cue words as list 1 (e.g., knee) but a portion of the pairs contained different target words (e.g., bone). On a final test, participants were given a cue (e.g., knee-b_n_) and specifically instructed to recall the target from list 2, although the target words from both list 1 and list 2 could complete the word stem. Half of the participants were given corrective immediate feedback and following the test all participants went through this entire procedure a second time with new word pairs.

Of interest to the current experiments are the changes in confidence reported by Jacoby et al. (2010) when participants went through the procedure a second time. The second time, participants were presented with completely new word pairs, so the feedback from session 1 could not affect accuracy for specific pairs; however, there were effects on confidence. In session 2, participants who had been given feedback during session 1 gave higher confidence ratings to correct items and lower confidence ratings to incorrect items. Participants also showed increases in relative accuracy during session 2 even though they were exposed to completely new information in that session. Jacoby et al. suggested that the feedback given during the first session changed how participants encoded items during the second session. Although the results and explanations reported

by Jacoby et al. cannot directly explain the findings in the current studies, they may help to explain why there are differences in confidence among feedback conditions during test 1. Receiving feedback may update participants' monitoring and change how they approach subsequent items. Without feedback participants may be unaware of the number of errors they are making. Thus, providing feedback immediately may make participants more sensitive to memory errors, reducing their overall level of confidence.

While feedback timing did not differentially affect confidence ratings on the final test, receiving feedback led to better relative and absolute accuracy for confidence ratings than not receiving feedback. This is consistent with previous research suggesting that feedback is beneficial for enhancing metamemory accuracy (Butler et al., 2008; Jacoby et al., 2010). For example, Butler et al. (2008) found that feedback improved both the absolute accuracy of confidence ratings and also item-by-item accuracy (i.e., relative accuracy). The results reported in the current experiments suggest that feedback timing does not differentially affect confidence ratings; however, they do indicate that feedback can be important for improving metacognitive accuracy relative to no feedback.

The interaction between feedback timing and the hypercorrection effect was of special interest in the current experiments. Previous research has shown that high-confidence errors are more likely to be corrected than errors held with low confidence (i.e., the hypercorrection effect) when participants are given immediate feedback (Butterfield & Metcalfe, 2001; 2006). It was hypothesized that the hypercorrection effect would be replicated in the immediate feedback condition, but predictions were less clear for the delayed feedback condition. Overall, the current experiments did not replicate the hypercorrection effect with immediate feedback; however, there is some evidence of a

hypercorrection effect in the delayed feedback condition. These data, and previous research (Butler et al., 2008), suggest that the hypercorrection effect may not be robust. One possibility is that the hypercorrection effect may only obtain at a short retention interval. For example, prior work on the hypercorrection effect (Butterfield & Metcalfe, 2001; 2006; Fazio & Marsh, 2010) has typically completed the entire experiment in one session without a long retention interval. Butler et al. (2008) suggest that the effect is less stable with longer retention intervals and were unable to replicate the hypercorrection effect using a 2-day retention interval. It may be that this effect is not robust, and thus, hard to replicate with longer retention intervals.

The materials used in the current study were also different than those traditionally used in studies examining the hypercorrection effect. Previous research has examined semantic memory through the use of general knowledge questions. The current experiments tried to replicate findings but with an episodic memory task that easily elicits high-confidence errors. Thus, the hypercorrection effect may be less robust in episodic memory tasks as compared with semantic memory tasks. As well, if the hypercorrection effect is less stable at longer retention intervals with semantic memory, these deficits may be even more pronounced with an episodic memory task. In the current experiments, these two variables, the type of memory task and the retention interval, are conflated. Future research may benefit by manipulating both the retention interval and type of memory task to examine the effect of each variable.

Feedback Processing Time

Butler et al. (2008) and Kulhavy and Stock (1989) have proposed theories to explain the relationship between response confidence, accuracy, and feedback processing

time. Both theories suggest that feedback processing time is based on the discrepancy between actual performance and the metacognitive assessment of that performance. When participants are highly confident in their response, but receive feedback that they are wrong, they will spend more time processing that feedback than if they had a low level of confidence in that response. Likewise, when participants have a low level of confidence in a correct response, they should spend more time processing that feedback than if they were highly confident in their response.

The results of experiment 2 supported these theories. There was a positive correlation between feedback processing time and confidence for items answered inaccurately (participants spent more time processing the feedback as their confidence increased), and the correlation was negative when items were answered accurately (participants spent less time processing feedback as their confidence increased). The main focus of experiment 2 was to assess whether the timing of feedback would impact the relationship between confidence and feedback processing time. The current results suggest that the relationship between confidence, accuracy, and feedback processing time remains stable regardless of whether feedback is administered immediately or after a delay. Thus, even after a delay, confidence information is still accessible and still plays an important role in the processing of feedback. This suggests that the influence of confidence should be equivalent for delayed and immediate feedback. However, further investigation should be done to examine how the length of delay may influence this. Perhaps, if feedback is not administered until several days later, the relationship between confidence, accuracy, and feedback timing will change from the results in the current experiments.

Interestingly, the amount of time spent processing feedback is consistent with the hypercorrection effect. People spent more time processing feedback for confidently-held errors than for errors held with little confidence in both the immediate and delayed feedback conditions. Despite this longer study time, there was still no evidence that high confidence errors are more likely to be corrected on a final test than low confidence errors. It may be that participants are demonstrating a labor-in-vain effect (Nelson & Leoneiso, 1988). The labor-in-vain effect refers to the finding that even if participants show large increases in the amount of time they spend studying items, there may be little or no gains in final performance. Even though participants increased their study time for high-confidence errors, they may not have done anything within that period of time to increase the likelihood of correcting their errors (e.g., to process the information more deeply).

Teaching participants skills that are generally used to improve memory, such as processing information in a meaningful way, may help them to better encode and retain feedback. Recently, Finn and Metcalfe (2010) showed that increasing the level of retrieval required to process feedback led to a greater memory for correct responses. When participants answered incorrectly, feedback was scaffolded such that the letters of the correct response were presented one at a time until participants could generate the correct answer. They found that increasing the level of difficulty required to process feedback led to better retention of information compared with standard feedback (i.e., the correct response presented after an answer; Finn & Metcalfe, 2010). Overall, it may be beneficial for future research to examine how encoding of feedback could improve

performance. Perhaps how feedback is encoded is more crucial to performance than the timing of feedback.

Conclusions

Overall, the results of the two experiments in this study provide several interesting patterns. While there were not many differences between immediate and delayed feedback, it is clear that providing feedback is crucial for both memory and metamemory accuracy relative to providing no feedback. More research needs to examine those variables that may impact whether or not there are differences between immediate and delayed feedback.

The null result for the hypercorrection effect was especially interesting. As noted earlier, the materials and procedures used in these experiments differ somewhat from other research on the effect (Butler et al., 2008; Butterfield & Metcalfe, 2001; 2006). If the hypercorrection effect is sensitive to variations in retention interval or type of memory task, it would be important to understand what factors are impacting the effect. The current results, in combination with findings in the literature, potentially suggest that there may be differences based up on the type of memory task (semantic vs. episodic). It is evident that providing feedback improves memory relative to a condition not provided feedback. However, it is still unclear whether feedback timing may differentially affect confidence judgments and feedback processing times. Future research should attempt to carefully define when and how different variables may influence the efficacy of feedback.

REFERENCES

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 Psychology, 61,* 213-238.

Brackbill, Y., Bravos, A., & Starr, R. H. (1962). Delayed improved retention on a difficult task. *Journal of Comparative and Physiological Psychology, 55,* 947-952.

Brosvic, G. M., Epstein, M. L., Cook, M. J., & Dihoff, R. E. (2005). Efficacy of error for the correction of initially incorrect assumptions and of feedback for the affirmation of correct responding: Learning in the classroom. *The Psychological Record, 55,* 401-418.

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.

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.

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.

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.

Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2003). The role of feedback during academic testing: The delay retention effect revisited. *The Psychological Record, 53,* 533-548.

Fazio, L. K., & Marsh, F. J. (2010). Surprising feedback improves later memory. *Psychonomic Bulletin & Review, 16,* 88-92.

Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory & Cognition, 38,* 951-961.

Jacoby, L., Wahlheim, C., Rhodes, M., Daniels, K., & Rogers, C. (2010). Learning to diminish the effects of proactive interference: Reducing false memory for young and older adults. *Memory & Cognition*, *38*, 820-829.

Kato, T. (1985). Semantic-memory sources of episodic retrieval failure. *Memory & Cognition, 13,* 442-452.

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48,* 704-721.

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

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1,* 279- 308.

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

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

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.

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.

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.

Rhodes, M. G., & Kelley, C. M. (2005). Executive processes, memory accuracy, and memory monitoring: An aging and individual difference analysis. *Journal of Memory and Language, 52,* 578-594.

Roediger, H. L., III, & Karpicke, J. D. (2005). The power of testing memory: Basic research and implications for educational practice. *Perspective in Psychological Science, 1,* 181-210.

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.