# IOWA STATE UNIVERSITY
**Digital Repository**

2013

# Delaying feedback is beneficial, but only when curious about the answer

Kellie Marie Mullaney
*Iowa State University*

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Cognitive Psychology Commons

**Delaying feedback is beneficial, but only when curious about the answer**


by


**Kellie Mullaney**



A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE



Major: Psychology

Program of Study Committee:
Shana K. Carpenter, Major Professor
Alison L. Morris
Max Guyll



Iowa State University

Ames, Iowa

2013

# TABLE OF CONTENTS

## LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT

When students are tested, delaying feedback of the correct answer by a few seconds is more beneficial than providing it immediately (e.g., Brackbill & Kappy, 1962). The current study refers to this effect as the *delay-of-feedback benefit*. Little is known about the theoretical nature of the delay-of-feedback benefit. The current study investigated the hypothesis that a delay of feedback is beneficial because it allows a learner to anticipate—i.e., look forward to, or devote attentional resources to—the feedback when it arrives. Participants were asked a list of obscure trivia questions. Immediately after answering each question they were asked to indicate how curious they were to know the answer (on a scale of 1-6), which was then presented to them immediately or after a delay of a few seconds. A final test given several minutes later revealed better memory for items that had previously received delayed feedback compared to immediate feedback, but only under conditions in which participants expressed high curiosity to know the answer. No delay-of-feedback benefit ever occurred for items that received low curiosity ratings. This interaction between feedback timing and curiosity was obtained in Experiment 1 in which immediate vs. delayed feedback was manipulated within-subjects. Experiment 2 manipulated feedback timing between-subjects and obtained this same interaction only when the duration of the delay was varied (i.e., either 2, 4, or 8 seconds) rather than constant (i.e., always 4 seconds). These results demonstrate that participants are most likely to benefit from delayed feedback when they are curious about the answer, and when they are uncertain about exactly when the answer will appear.

# CHAPTER 1. INTRODUCTION

When participants attempt to answer a question on a test and answer it incorrectly, it is very unlikely that they will correct this error unless they receive knowledge of the correct answer. Research has shown that in order to correct errors, participants must be provided with feedback (i.e., the correct answer) after attempting to answer a question. For example, Pashler, Cepeda, Wixted, and Rohrer (2005) presented participants with Luganda-English word pairs and then gave them a cued recall test in which they had to supply the English translation of the Luganda word. Immediately after responding, one group of participants was presented with the correct English translation, whereas another group was informed of whether or not their answer was accurate but was not actually shown the answer. A third group received no feedback at all. On a final test over the word pairs given one week later, the only group of participants to show improvement from the initial test to the final test was the group that received feedback of the correct answer during learning. The group that received right/wrong feedback exhibited a decrease in memory accuracy from the initial test to the final test. When the answer on the initial test was correct, final memory accuracy did not differ between the groups. However, when the initial answer was incorrect, the group that received knowledge of the correct response showed more improvement than all other groups. This led to the conclusion that correct answer feedback is particularly beneficial for correcting initial errors.

Fazio, Huelser, Johnson, and Marsh (2010) confirmed these results using a within-subjects design in which participants were asked to learn information presented in a science passage. After reading the passage and taking a cued recall test over facts from

the passage, participants received correct answer feedback for some questions, right/wrong feedback for other questions, and no feedback for the rest of the questions. On a final test 20 minutes later, items for which correct answer feedback was provided were retained the best. Lhyle and Kulhavy (1987) found that similar results applied to learning of anatomical knowledge. After viewing and being tested over the structure and function of the human eye, participants who received feedback in the form of the correct answer retained more than participants who did not receive feedback.

## Timing of Feedback

While these studies have shown that feedback of the correct answer is essential for learning, it is also important to identify the optimal time during which to provide that feedback. Researchers who studied operant learning using animal paradigms concluded that rewards must be administered immediately after the desired behavior in order to reinforce that behavior—i.e., increase the chances that it will occur again.  The rewards given in these animal paradigms are comparable to feedback in studies on human learning, in that they provide knowledge of the desired response. Pubols (1958) showed that while learning a maze, rats given a reward (i.e., a food pellet) immediately after each successful trial required fewer attempts to meet a pre-specified level of learning (i.e., 19 successful trials out of 20), compared to rats that received the reward after a brief delay (e.g., 4 seconds).  In addition, as the delay interval between the correct behavior and the reward increased (e.g., from 4 seconds to 8 seconds), so did the number of trials needed for the rats to reach the pre-specified level of learning. In other words, learning suffered when the rat had to wait for knowledge of the correct response. In addition to resulting in a slower rate of learning, delaying a reward during learning also resulted in slower

response times associated with learned bar-pressing behaviors (e.g., Perin, 1943), suggesting that a learned behavior is weaker or more fragile when acquired under conditions of delayed vs. immediate knowledge of the correct response.

These studies suggested that the best way to increase a desired response was to provide knowledge of the correct response immediately after it occurred. Several decades ago, some researchers applied this logic to human learning (Pressey, 1926; Skinner, 1954, 1958; Vargas, 1986). These researchers' advice to classroom instructors was that after answering a question, students should be provided with feedback (i.e., the correct answer) immediately in order to increase the likelihood of learning that response. Early studies on human learning appeared to confirm this. For example, Saltzman (1951) presented participants with pairs of numbers and asked them to guess which number in the pair was correct. Participants who received feedback immediately after each response required fewer trials to correctly answer all of the number pairs on two consecutive trials, compared to those who received feedback 6 seconds after each response. Similarly, Bourne (1957) found that the longer participants were required to wait for feedback while learning to classify shapes into categories, the more errors they committed before reaching a pre-specified level of learning. Furthermore, Hockman and Lipsitt (1961) found that while learning to associate colored lights with buttons (e.g., when an orange light appears press the left button), participants who received feedback immediately after providing a response made fewer errors compared to those who received feedback after a 10-second delay. These studies using human learners appeared to confirm the results of earlier animal studies (i.e., that delaying feedback is detrimental to learning).

The studies discussed so far all have in common that memory was assessed immediately after exposure to the material. However, in real word learning instances it is more likely that a learner would acquire information with the intent of remembering it for a much longer period of time. For example, when students learn information for a test, it will have to be retained for a minimum of several hours, or possibly even for several weeks if the test is a cumulative final exam. Therefore, it is important to examine memory not only immediately after exposure to the information, but also after some time has elapsed.

While the provision of feedback after a delay may be unfavorable to immediate memory, it can actually be beneficial for memory that is measured after some time has passed. Brackbill and colleagues (e.g., Brackbill & Kappy, 1962; Brackbill, Bravos, & Starr, 1962; Brackbill, Isaacs, & Smelkinson, 1962; Brackbill, 1964) demonstrated this by measuring both immediate and delayed memory for information presented in a discrimination task. For instance, Brackbill and Kappy (1962) presented children with pairs of line drawings (e.g., a boat and a star) and asked them to guess which of the drawings was correct. After each response, feedback was given either immediately or after a 10-second delay. If the correct response was selected, a marble was provided that could later be redeemed for a prize. Similar to previous research, immediate provision of feedback facilitated acquisition—that is, the children required fewer trials to reach a pre-specified criterion of three correct responses in a row. However, one day later participants were required to relearn the pairs to the same criterion, and on this test the children who had received feedback after a 10-second delay actually outperformed those who received immediate feedback. Using a similar design, Lintz and Brackbill (1966)

observed the same pattern of results on a retention test that was delayed by seven days. It appears, therefore, that delaying presentation of feedback during learning may have adverse effects on acquisition, but beneficial effects on later retention. This facilitation of delayed feedback on later retention of information has been termed the delay-of-feedback benefit.

## Theoretical Explanations for the Delay-of-Feedback Benefit

*Rewards*

Why does memory retention benefit when feedback is delayed for a brief period of time during learning? In the studies by Brackbill and colleagues, feedback was always presented in conjunction with a token reward (i.e., a marble to be redeemed for a prize). It is not clear, therefore, whether the benefits observed were due to the delay of feedback per se, or due to the delay of the reward. If a delayed presentation of feedback is only beneficial when accompanied by a reward, it might indicate that the presence of important motivational factors underlie this effect. To what degree is the delay-of-feedback benefit driven by delayed presentation of the feedback itself in the absence of a tangible reward?

Several studies have addressed this question by replicating the basic design of the studies by Brackbill and colleagues but without presentation of the marble reward. First, Sturges, Sarafino, and Donaldson (1968) presented participants with a U. S. state and two cities, and asked them to guess which city was the capital. Feedback of the correct answer, but not a reward, was given either immediately or 10 seconds after each response. Seven days later, those who had received delayed feedback required fewer trials to relearn the capitals to a pre-determined level of learning than participants who

had received immediate feedback. Second, Sassenrath and Yonge (1969) measured

retention for introductory psychology material that was initially recalled via a multiple-

choice test in which feedback occurred immediately or 10 seconds after each response.

Participants who had received feedback after a delay answered more questions correctly

on a final test given five days later than those who had received the feedback

immediately. Finally, Rankin & Trepper (1978) asked participants a series of questions

about the biology of human sexuality and found that participants presented with feedback

delayed by 15 seconds performed better on a final test given over the same questions one

day later compared to participants who were provided with feedback immediately after

each question. These studies demonstrated that delaying presentation of feedback per se,

even when a reward is not provided, is beneficial for delayed retention.

*Time on Task*

What mechanism(s) might underlie the retention benefit associated with delaying

feedback? A careful inspection of the studies discussed so far reveals that participants

spent more time on trials that involved delayed feedback compared to immediate

feedback (e.g., Brackbill & Kappy, 1962; Brackbill et al., 1962; Brackbill et al., 1962;

Brackbill, 1962; Lintz & Brackbill, 1966; Sturges et al., 1968; Sassenrath & Yonge,

1969; Rankin & Trepper, 1978). For example, in Sassenrath and Yonge (1969),

participants were presented with a question pertaining to introductory psychology

material and attempted to guess the answer. Feedback of the correct answer was then

shown to them for 5 seconds. For participants who received immediate feedback, the total

time spent on each trial was therefore equivalent to the amount of time that it took them

to respond, plus 5 seconds for presentation of the answer.  On the other hand, for

participants who received 10-second delayed feedback, the total time spent on each trial

was equivalent to the amount of time that it took them to respond, plus a 10-second delay,

plus a 5-second presentation of the answer. After making their responses, therefore, the

group that received delayed feedback spent a total of 15 seconds on each trial, whereas

the group that received immediate feedback spent only 5 seconds on each trial.

Spending more time on each trial could have resulted in an advantage for the

group that received delayed feedback relative to immediate feedback, for at least two

reasons. First, if participants initially guessed the correct answer and then had to wait for

10 seconds before being presented with the answer, they could have spent that 10-second

period rehearsing or thinking about the answer. Those in the immediate feedback group,

on the other hand, did not receive the potential benefit of this extra rehearsal time. Given

that rehearsal time has known benefits on memory (e.g., Waugh, 1967), it is possible that

the benefits of delaying feedback could be attributed simply to additional time spent

thinking about the correct answer.

Second, even if participants initially guessed the wrong answer, being required to

wait 10 seconds before seeing the feedback may have provided additional opportunities

to try to guess the answer. Research on the testing effect (i.e., *retrieval practice*) has

consistently demonstrated that the act of trying to recall information—even if those recall

attempts are unsuccessful (e.g., Kornell, Hays, & Bjork, 2009)—produces significant

benefits on later memory for the material being learned (e.g., Roediger & Karpicke,

2006; Carpenter, Pashler, & Cepeda, 2009; Agarwal, Karpicke, Kang, Roediger, &

McDermott, 2008). The additional time afforded by delaying presentation of the correct

answer, therefore, could simply be a product of the well-known benefits of retrieval practice.

There is only one known study that ruled out the potential confounding effects of rehearsal and retrieval practice on the delay-of-feedback benefit. Carpenter and Vul (2011) first showed participants a series of faces paired with names and then gave them a cued-recall test in which they had to supply the name when shown each face. In one condition, participants were given 8 seconds to try to recall the name, then as soon as they entered a response, they were shown a blank screen for 3 seconds, followed by presentation of the correct face and name for 2 seconds (the Delayed Feedback Condition), resulting in a total of 13 seconds per trial. In another condition, they were simply provided with the correct face and name for 5 seconds immediately after making a response (the Immediate Feedback Condition), again resulting in 13 seconds spent on each trial. A third condition allowed participants 11 seconds to make a response, and then immediately provided the correct answer for 2 seconds (the Prolonged Test Immediate Feedback Condition), resulting again in 13 seconds spent on each trial. If the total time spent on each trial influences the delay of feedback benefit, then holding total time constant should produce no differences in final retention of face-name pairs across the three conditions. If having more time to rehearse the correct answer accounts for the benefits of delayed feedback, then the Immediate Feedback Condition should produce the best retention of face-name pairs. If having extra time to engage in retrieval practice accounts for the delay of feedback benefit, then the Prolonged Test Immediate Feedback Condition should produce the best retention. Finally, if there is something beneficial about the delay of the correct answer per se, even after controlling for rehearsal time and

time spent trying to recall the correct answer, then the best retention should occur for the condition in which the presentation of the correct answer was delayed. Indeed, this benefit of delayed feedback was found by Carpenter and Vul (2011). This was the first known study to rule out the effects of rehearsal time and retrieval practice, to find that delaying presentation of feedback by a few seconds produces significant benefits on retention, over and above any influences of rehearsal time and retrieval practice. This suggests that the delay-of-feedback benefit is more likely to be driven by the processing that occurs during the delay, rather than time spent on each trial per se.

*Processing During the Delay*

There have been two known hypotheses that have proposed what type of processing could be occurring during a feedback delay that contributes to later retention. First, Kulhavy and Anderson (1972) proposed the interference-perseveration hypothesis, suggesting that delayed feedback is beneficial due to forgetting that occurs during the delay. It was proposed that if an incorrect response is given to a question, a delay serves as time for that incorrect response to be forgotten before feedback is received. If the incorrect response is forgotten, it will not interfere with the processing of feedback when it is received. On the other hand, if feedback is received immediately after answering a question, the incorrect response will still be active when feedback is received, and is therefore more likely to interfere with processing of the feedback. The greater potential for interference experienced after immediate relative to delayed feedback would presumably result in lower accuracy when recalling the answers at a later time. According to this hypothesis, the delay-of-feedback benefit depends primarily on the process of forgetting that occurs during a delay period.

An alternative hypothesis, proposed by Carpenter and Vul (2011), suggests that during a brief delay prior to presentation of the answer, participants may be actively anticipating what the answer is. Being forced to wait for the answer might instill within participants an increased sense of curiosity, focus, or preparedness to encode the forthcoming answer. Based on the evidence that the delay-of-feedback benefit is eliminated under conditions in which participants must perform a distractor task during the 3-second delay (Carpenter & Vul, Experiment 3), the authors concluded that the delay-of-feedback benefit depends on the active engagement of attentional resources during the delay period, rather than a passive process such as forgetting which was suggested by Kulhavy and Anderson (1972).

### Current Study

The current study is aimed to better understand the processes that might be driving the benefits of short delays of feedback on memory retention. Two experiments utilized the same basic design of Carpenter and Vul (2011), but expanded upon this work in important ways. In both experiments, participants answered a series of trivia questions (e.g., *What does a deltiologist collect?*) and immediately afterward, rated their curiosity to know the answer on a scale of 1 (not at all curious) to 6 (highly curious). Feedback of the correct answer (e.g., *Postcards*) was then provided immediately after the curiosity rating or after a 4-second delay period during which the question remained on the screen.

In line with the reasoning put forth by Carpenter and Vul (2011), delaying presentation of the answer by four seconds forces participants to wait for the answer and thus creates an opportunity for them to anticipate what is coming, whereas providing the answer immediately does not allow time for this process to take place. The current study

explored the delay-of-feedback benefit under conditions in which this anticipatory processing would seem more likely to occur—i.e., when participants are curious to know the answer. Participants' tendencies to anticipate the answer to a question would presumably occur to a greater degree for items that arouse high levels of curiosity than for items that arouse lower levels of curiosity. Therefore, it was hypothesized, that the items that are most likely to be anticipated (i.e., those that arouse high curiosity) would benefit more from delayed feedback than the items that are less likely to be anticipated (i.e., those that arouse less curiosity). The key result that would support this hypothesis is an interaction between feedback timing and level of curiosity, such that the benefits of delayed feedback over immediate feedback are stronger under conditions in which participants are more curious to know the answer, relative to conditions under which participants are less curious to know the answer. This interaction would be a novel finding that would suggest that certain conditions, such as a high level of curiosity, are required in order for delayed feedback to be beneficial for later memory.

Experiment 1 utilized a within-subjects design in which half of the items were randomly assigned to receive delayed feedback, and half to receive immediate feedback. As such, participants were not aware ahead of time whether the feedback on any given trial would occur immediately or after a delay. Any effects associated with feedback delay could therefore be driven by the delay of feedback per se, or by the uncertainty of the delay. Based on evidence showing that delay-of-feedback benefits are stronger under conditions in which the length of the delay is varied rather than constant (e.g., Schroth, 1995), it is possible that the uncertainty of the delay in Experiment 1 could encourage

anticipatory processing that benefits memory retention for high-curiosity items, over and beyond the effects of feedback delay per se.

Experiment 2 was designed to explore this question. Participants answered the same trivia questions from Experiment 1 and then immediately rated their curiosity to know the answer. Unlike Experiment 1, however, the feedback delay was manipulated between-subjects and a new condition was added. After making the curiosity rating, one group of participants received feedback immediately (the Immediate Group) , another group  received 4-second delayed feedback on every trial (the Constant Delay Group), and the last group received feedback delayed by an unpredictable duration—2, 4, or 8 seconds— (the Varied Delay Group).  Participants in the Immediate and Constant Delay Conditions were informed ahead of time of the exact timing of feedback they would experience, and received the same feedback timing throughout the experiment. However, those in the Varied Delay Condition, similar to those in Experiment 1, never knew ahead of time exactly when the feedback would occur on any given trial.

Results of Experiment 2 were analyzed according to planned comparisons that explored the same 2-way interaction from Experiment 1 (Level of Curiosity: high vs. low x Feedback Timing: immediate vs. delayed). If the results of Experiment 1 were driven by the delay of feedback per se and not the uncertainty about when that feedback would occur, then Experiment 2 should replicate the same Level of Curiosity x Feedback Timing interaction with respect to both the Constant and Varied Delay Conditions. On the other hand, if uncertainty about when the answer would occur contributed significantly to this interaction in Experiment 1, then this interaction would be expected

to occur in Experiment 2 only for the Varied Delay Condition but not for the Constant Delay Condition.

In both experiments, it is also expected that curiosity would be generally beneficial to memory, such that questions that arouse higher levels of curiosity will be recalled with greater accuracy than questions that arouse lower levels of curiosity. This result would be consistent with past studies that have found that higher levels of curiosity are associated with greater retention (Berlyne, 1954; Berlyne, 1966). For example, in Berlyne's (1954) study, participants took an initial test over a series of questions about animals and were told to mark the questions about which they were most curious. They then read facts about each of the animals which provided answers to the previously-asked questions, and were then given a final test over the questions. Accuracy on the final test was better for the questions that participants indicated they were curious about compared to questions that they were not curious about, suggesting that memory is better for items that arouse higher levels of curiosity.

# CHAPTER 2. EXPERIMENT 1

Experiment 1 was designed to explore whether the size of the delay-of-feedback benefit differs as a function of how curious participants are to know the answer to a question. Participants were asked to guess the answers to a list of general knowledge questions (e.g., *What does a deltiologist collect?*), and immediately afterward were asked to rate how curious they were to know the answer on a scale from 1 (not at all curious) to 6 (very curious). Participants were then shown feedback of the correct answer (e.g., *Postcards*) either immediately (Immediate Feedback) or after a 4-second delay (Delayed Feedback). After 5 minutes, participants were given a final test over all of the questions.

## Method

*Participants*

Thirty-two undergraduate students at Iowa State University volunteered to participate in this experiment in order to fulfill partial requirements for an introductory psychology course.

*Materials*

Seventy general knowledge questions and answers were obtained from a book of trivia facts (Botham, 2006). A complete list of the questions and answers is provided in Appendix A.

*Procedure*

Participants were first informed that they would be learning a list of trivia facts and would be tested over them at a later time. They were then presented with each of the 70 questions one at a time in a random order, and were asked to provide their best guess of the answer. Each question was presented in the center of the computer screen, and

participants were allowed unlimited time to answer. After typing a response and pressing

ENTER, the response that participants had typed disappeared from the screen, but the

question remained and participants were asked "How curious are you to know the answer

to this question?" A scale ranging from 1 (not at all curious) to 6 (very curious) was

displayed at the bottom of the screen, and participants were asked to enter a number

between 1 and 6 that corresponded to their curiosity about the answer to that particular

question[1].

For half of the questions, as soon as participants entered a curiosity rating, they

received immediate feedback in which the curiosity rating scale disappeared, and the

answer was displayed directly below the question for 10 seconds. For the other half of the

questions, participants received delayed feedback in which the curiosity rating scale

disappeared from the screen and the question alone remained for 4 seconds, followed by

the answer appearing below the question for 6 seconds. In both conditions, a total of 10

seconds was spent on each trial after a curiosity rating was entered to ensure that

participants did not spend more time on questions in one condition compared to the other.

The order of presentation and assignment of questions to receive either immediate or

delayed feedback was randomly determined for each participant.

After participants answered all 70 questions, entered a curiosity rating, and

received feedback, they engaged in a distractor task in which they were asked to type in

the names of as many of the 50 U.S. states as they could think of within a 5-minute time

period. Participants typed the name of each state onto the computer screen, and were

instructed to press ENTER after they typed in the name of each state, which caused that

state to appear in a list on the bottom half of the computer screen. Checking the list of

already-entered states could help participants to avoid naming the same state multiple

times.

After the 5-minute distractor task, participants took a final test over the 70

questions in a new random order. On the final test, participants were presented with each

question that they had seen before, one at a time, and they were given unlimited time to

type in an answer below the question. Feedback was not provided, and participants did

not rate their curiosity about the answers to the questions on the final test. Upon

completion of the final test, participants were debriefed and thanked.

## Results and Discussion

*Scoring*

All of the responses given by participants were hand scored and were given full

credit (1 point) if they were an exact match or contained minor spelling errors (e.g., *Dr.*

*Suess instead of Dr. Seuss*). Half credit (1/2 point) was given to responses that were

considered to be only partially correct (e.g., *Ford instead of Gerald Ford*). Two raters

who were blind to the conditions scored 20% of the responses, and interrater correlations

were significant for both the Immediate and Delayed Conditions ($r$s > .98, $p$s < .05). The

remainder of the scoring was finished by one rater who was blind to the conditions.

*Initial Test*

The mean accuracies, curiosity ratings, and response times (RTs) on the initial test

are displayed in Table 1 below. Accuracy on the initial test was generally very low, and a

paired-samples t-test showed that it did not differ significantly for items that received

either immediate or delayed feedback [$t(31) = 0.08$, $p > .05$]. The average curiosity

ratings given also did not differ significantly for items that received either immediate or

delayed feedback [$t(31) = 0.87$, $p > .05$]. Finally, there was not a significant difference

between immediate and delayed feedback for time spent guessing the answers to the

questions [$t(31) = 0.65$, $p > .05$] or time spent providing curiosity ratings to the questions

[$t(31) = 1.65$, $p > .05$].

Table 1

*Mean curiosity rating, accuracy, RT for guessing the answer (Answer RT), and RT for rating curiosity*
*(Curiosity RT) on the initial test for Immediate and Delayed Feedback Conditions in Experiment 1*

| Feedback Timing | Measure | | | |
|---|---|---|---|---|
| | Curiosity (1-6) | Accuracy | Answer RT | Curiosity RT |
| Immediate | 3.175 (1.136) | .061 (.056) | 10850 (3079) | 1408 (409) |
| Delayed | 3.229 (1.152) | .062 (.054) | 11065 (3249) | 1457 (437) |

*Note.* Times are reported in ms. Standard deviations are given in parentheses.

*Final Test*

To be sure that prior knowledge did not influence the results, analyses of final test

accuracy included only questions that participants answered incorrectly on the initial test.

On average, only 6.34% of the items (4 items, on average) were eliminated per

participant as a result of being answered correctly on the initial test, which includes both

those that received full credit and those that received partial credit. Thus, final test

accuracy was based on 66 items per participant, on average.

The proportion of questions answered correctly on the final test was analyzed as a

function of both curiosity level and feedback timing. The curiosity rating given to each

question during the initial test was dichotomized as either high or low according to

whether that item received a rating that fell within the upper half of the scale (i.e., 4

through 6), or the lower half (i.e., 1 through 3), respectively. Therefore, items could fall into one of four possible categories: questions given high curiosity ratings that received delayed feedback (Delayed High), questions given low curiosity ratings that received delayed feedback (Delayed Low), questions given high curiosity ratings that received immediate feedback (Immediate High), or questions given low curiosity ratings that received immediate feedback (Immediate Low). Seven participants who did not contribute at least two items to each of these four possible categories were excluded from the following analyses[2]. Figure 1 displays performance on the final test for the 25 remaining participants as a function of feedback timing and curiosity.
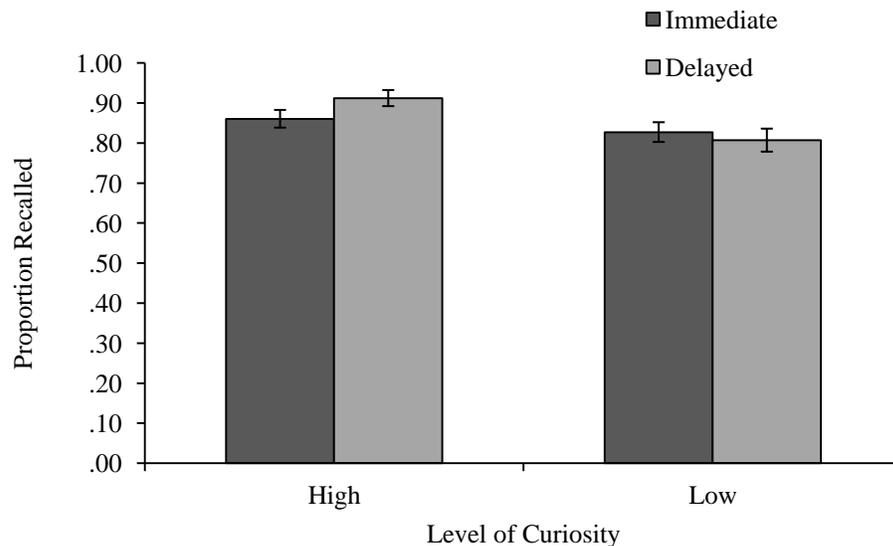


*Figure 1.* Proportion of answers correctly recalled on final test in Experiment 1 for the Immediate and Delayed Feedback Conditions as a function of level of curiosity.

A 2 x 2 Repeated Measures ANOVA revealed a significant effect of level of curiosity [$F(1, 24) = 13.79$, $p < .05$], in that items given high curiosity ratings were recalled with greater accuracy than items given low curiosity ratings. While there was no

significant main effect of feedback timing [$F(1, 24) = 0.66$, $p > .05$], there was a significant interaction between feedback timing and level of curiosity [$F(1, 24) = 4.73$, $p < .05$]. There was no difference in retention between immediate and delayed feedback for low curiosity items [$t(24) = 0.67$, $p > .05$], but there was a significant advantage of delayed over immediate feedback for high-curiosity items [$t(24) = 2.64$, $p < .05$][3].

Since participants were given unlimited time to guess the answer and rate their curiosity for each question, an analysis of RTs was conducted to explore the possibility that participants spent more time on questions in a certain condition (e.g., Delayed High) compared to the other conditions. The means and standard deviations for the time spent answering the questions and rating curiosity are shown in Table 2.

Table 2

*Mean RT for guessing the answer (Answer RT) and RT for rating curiosity (Curiosity RT) on the initial test as a function of level of curiosity for the Immediate and Delayed Conditions in Experiment 1*

| Feedback Timing | Answer RT | | Curiosity RT | |
|---|---|---|---|---|
| | High Curiosity | Low Curiosity | High Curiosity | Low Curiosity |
| Immediate | 11004 (3171) | 10789 (3640) | 1630 (378) | 1533 (462) |
| Delayed | 11213 (3022) | 10740 (3419) | 1740 (393) | 1544 (466) |

*Note.* Times are reported in ms. Standard deviations are given in parentheses.

For the time required to answer the questions, a 2 x 2 (Level of Curiosity: high vs. low x Feedback Timing: immediate vs. delayed) Repeated Measures ANOVA revealed no significant main effects or interaction (all $F$s < 0.80). The same 2 x 2 ANOVA on time spent providing curiosity ratings revealed no significant main effect of feedback timing and no significant Level of Curiosity x Feedback Timing interaction (both $F$s < 2.95).

There was, however, a marginally significant main effect of level of curiosity [$F(1, 24) =$ 3.99, $p = .06$], in that participants spent more time providing curiosity ratings for high-curiosity items compared to low-curiosity items ($t(49) = 2.39, p < .05$). Thus, participants spent longer assigning ratings to items for which they were more curious, but no differences in RTs ever emerged between items that received immediate vs. delayed feedback. It seems that the time spent on each trial during the learning phase did not coincide directly with the pattern of memory accuracy that was observed on the final test. Therefore, the significant Feedback Timing x Level of Curiosity interaction observed for memory accuracy cannot be attributed to any differences in the amount of time spent trying to guess the answer or rate curiosity during the learning phase.

As a supplemental analysis, gamma correlations (Goodman & Kruskal, 1954) were used to analyze the relationship between level of curiosity and feedback timing. Unlike a Pearson correlation that assesses the strength of a relationship between two continuous variables, a gamma correlation assesses the strength of a relationship between two variables of an ordinal nature, such as the relationship between curiosity rating (1 through 6) and later accuracy (which was classified as 0, 0.5, or 1 to denote an incorrect, partially correct, or correct response, respectively). Gamma correlations represent the probability of like order (i.e., accuracy on the final test increases as initial curiosity increases, resulting in a positive gamma value) and the probability of unlike order (i.e., accuracy on the final test decreases as initial curiosity increases, resulting in a negative gamma value). The higher the absolute value of gamma, the stronger the relationship between curiosity and final test accuracy. Due to the within-subjects design of Experiment 1, it was possible to calculate gamma correlations between curiosity and

accuracy for both the Immediate and Delayed Feedback Conditions for each individual participant. The two gamma values for each participant could then be directly compared to determine whether curiosity was more strongly linked with later memory accuracy for the Delayed or Immediate Feedback Condition.

Computation of a gamma correlation is not possible if a participant does not produce more than one value on one of the measures (e.g., if a participant exclusively rates all items at a curiosity level of 6, or does not recall any items correctly on the final test). The current results exclude three participants for whom this was the case. One participant gave all of the questions a rating of 6 and two participants gave all of the questions a rating of 1. For the remaining 29 participants, a paired-samples t-test showed that curiosity ratings were positively associated with final test accuracy to a greater degree for items that received delayed feedback compared to those that received immediate feedback, $t(28) = 2.09$, $p = .054$ (two-tailed), $p = .03$ (one-tailed). These results (see Figure 2) are consistent with the interaction between curiosity level and feedback timing that demonstrated stronger delay-of-feedback benefits on memory accuracy for items that aroused more curiosity compared to those that aroused less curiosity.
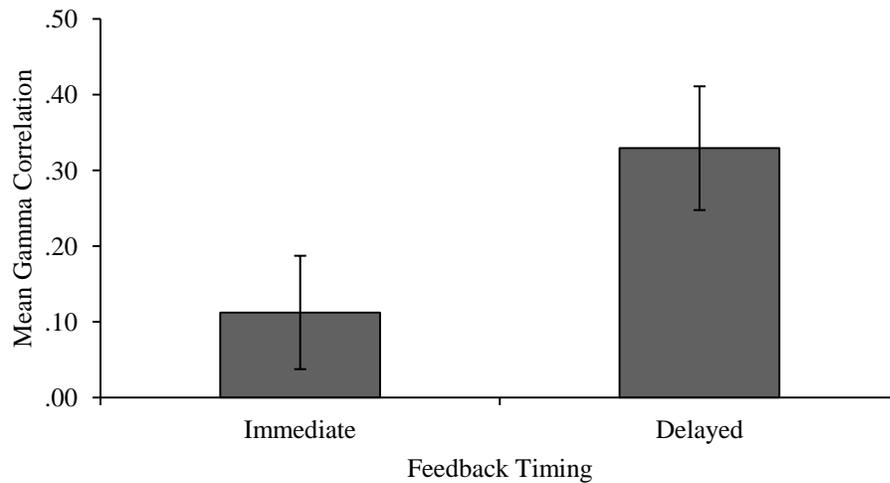
*Figure 2.* Mean gamma correlations between curiosity rating and accuracy on the final test for Experiment 1 for the Immediate and Delayed Conditions.

The results of Experiment 1 confirmed the hypothesis that the delay-of-feedback benefit is stronger under conditions in which participants are more curious to know the answer to a question compared to conditions in which the same participants are less curious to know the answer. These findings add some support to the hypothesis that items that are more likely to be anticipated (i.e., those that arouse greater curiosity) benefit more from delayed feedback than items that are less likely to be anticipated (i.e., those that arouse less curiosity).

In Experiment 1, feedback delay was manipulated within-subjects, so participants were not aware ahead of time whether the feedback would be immediate or delayed for any individual item. Feedback delay therefore varied systematically with the uncertainty of the delay. This makes it difficult to know whether the effects observed were being driven by the delay of feedback per se, or by the uncertainty about when the feedback would be provided. Experiment 2 was designed to further examine these possibilities.

# CHAPTER 3. EXPERIMENT 2

Experiment 2 was designed to explore whether a sense of uncertainty about the timing of feedback contributed to the results of Experiment 1. In a between-subjects design, participants learned the same trivia facts from Experiment 1, with one group of participants receiving immediate feedback (the Immediate Group), a second group receiving feedback after a 4-second delay (the Constant Delay Group), and a third group receiving feedback after a 2-, 4-, or, 8-second delay (the Varied Delay Group). The Immediate and Constant Delay Groups represent a between-subjects replication of the design of Experiment 1, in which immediate vs. 4-second delayed feedback was manipulated within-subjects.

In Experiment 1, feedback delay was manipulated within-subjects and so it is possible that the delay-of-feedback benefit observed for high-curiosity items was driven to some degree by the delay itself, as well as uncertainty about whether or not there would be a delay. Experiment 2 separately examined these two factors by comparing memory retention for facts learned through immediate feedback to those learned through a condition in which only the delay was manipulated (i.e., the Constant Delay Group), vs. a condition in which both the delay and uncertainty about the length of the delay were manipulated (i.e., the Varied Delay Group).

If feedback timing alone was responsible for the effects observed in Experiment 1, then the Curiosity Level x Feedback Timing interaction should be observed for both the Constant and Varied Delay Groups. However, if uncertainty about the timing of the delay contributes to this effect beyond the influence of the delay itself, then this

interaction should occur only for the Varied Delay Group but not for the Constant Delay Group.

## Method

*Participants*

A total of 208 participants were recruited from the same participant pool as in Experiment 1. Each participant was randomly assigned to one of three feedback groups: 70 were assigned to the Immediate Group, 68 to the Constant Delay Group, and 70 to the Varied Delay Group.

*Materials*

A list of 69 questions from Experiment 1 was used. One question from Experiment 1 was randomly chosen to be excluded (i.e., *What was the biggest-selling toy in 1957?*), in order to ensure that 23 questions were  assigned to each delay duration (i.e., 2, 4, and 8 seconds) in the Varied Delay Group.

*Procedure*

As in Experiment 1, participants were shown each of the 69 questions, one at a time in random order, and were given unlimited time to type in their answers. Immediately after typing in their response, they rated their curiosity to know the answer using the same 6-point scale from Experiment 1. Immediately after entering a curiosity rating, feedback was administered according to one of the following between-subjects groups:  (1) *Immediate*, in which feedback was provided immediately after a response was entered, (2) *Constant Delay*, in which a 4-second delay was experienced, and then feedback was presented, or (3) *Varied Delay*, in which one-third of the items were

randomly assigned on an individual participant basis to have their answers displayed after a 2-second delay, one-third after a 4-second delay, and one-third after an 8-second delay.

Questions were presented in random order for each individual participant, so those in the Varied Delay Group never knew ahead of time on any given trial whether feedback would be presented after 2, 4, or 8 seconds. Feedback for all three of the groups always consisted of the correct answer displayed below the question for 6 seconds, which ensured that equal time was spent viewing the correct answer for all participants. For the Constant Delay and Varied Delay Groups, the question alone remained on the screen during the delay period, just like in Experiment 1.

After all 69 items were tested in this fashion, participants performed an unrelated distractor task consisting of two parts. The first part was identical to the task used in Experiment 1, in that participants were asked to type the names of as many U.S. states as they could think of in a 5-minute period. The second part consisted of a 25-letter x 25-letter word search puzzle, below which a list of 35 words were displayed (see Appendix B). Participants were instructed to circle each of the words that they found in the puzzle and then cross the word out in the list below the puzzle. They were given 10 minutes to find as many words as possible. The words that participants were asked to find were not answers to any of the trivia questions. The total time spent on the distractor task was increased in Experiment 2 (i.e., from 5 to 15 minutes) in order to decrease the likelihood of ceiling effects occurring, as accuracy on the final test in Experiment 1 was generally high ($M = .85$, $SD = .11$).

After completing the distractor task, participants took a final test over the same 69 questions in a new random order. Participants were asked to enter their responses as each

question was presented on the screen. Feedback was not provided, and participants were

not asked to rate their curiosity on the final test. After completing the final test, the

following question appeared on the screen: "Have you participated in another study (this

semester or in a previous semester) in which you have been asked the same trivia

questions you were asked today?"[4]. After entering a response of yes or no to this

question, the following question appeared on the computer screen: "Did you always have

to wait the same amount of time before receiving feedback of the correct answer?

(EXAMPLE: you always had to wait 3 seconds before the correct answer appeared on the

screen)." This question was used as a manipulation check to determine whether

participants in the Varied Delay Group experienced uncertainty about the timing of

feedback while participants in the Immediate and Constant Delay Groups did not. If

participants answered "no" to the question (i.e., indicating that they did *not* always have

to wait the same amount of time for feedback), it is more likely that they experienced

some uncertainty about the timing of the feedback compared to participants that answered

"yes" to the question (i.e., indicating that they thought they did always wait for the same

amount of time)[5]. After responding yes or no to this question, participants were debriefed

and thanked for their participation.

### Results and Discussion

*Scoring*

As in Experiment 1, 20% of the responses were hand-scored by two raters, blind

to the experimental conditions, who awarded full credit (1 point), partial credit (1/2

point), or no credit (0 points) for each response. Inter-rater correlations were significantly

high (all $r$s > .98, $p$s < .05), so the remainder of the responses were scored by a single

rater who was blind to the conditions.

*Initial Test*

 Mean accuracies, curiosity ratings, and RTs were analyzed in the same fashion as

in Experiment 1 and are displayed in Table 3 below. Mean accuracy on the initial test

was generally low. A series of One-Way Between-Subjects ANOVAs revealed no

significant differences among the three feedback groups with respect to accuracy [$F(2,$

$206) = 1.03, p > .05$], curiosity ratings, [$F(2, 206) = 0.40, p > .05$], or time spent guessing

the answers to the questions [$F(2, 206) = 0.41, p > .05$]. There were, however, differences

in the amount of time spent providing curiosity ratings to the questions among the three

groups [$F(2, 206) = 7.95, p < .05$]. Post-hoc comparisons using Tukey's HSD test

indicated that participants in the Varied Delay Group spent longer entering curiosity

ratings than those in the Immediate Group. There were no significant differences between

the Constant Delay and Immediate Groups or between the Constant Delay and Varied

Delay Groups.

Table 3

*Mean curiosity rating, accuracy, RT for guessing the answer (Answer RT), and RT for rating curiosity (Curiosity RT) on the initial test for Immediate, Constant Delay, and Varied Delay Groups in Experiment 2*

| Feedback Timing | Measure | | | |
| --- | --- | --- | --- | --- |
| | Curiosity | Accuracy | Answer RT | Curiosity RT |
| Immediate | 3.269 (1.039) | .078 (.051) | 11898 (4169) | 1477 (370) |
| Constant Delay | 3.362 (1.204) | .074 (.045) | 11576 (4003) | 1612 (422) |
| Varied Delay | 3.194 (1.090) | .067 (.042) | 12298 (5810) | 1781 (549) |

*Note.* Times are reported in ms. Standard deviations are given in parentheses.

*Final Test*

As in Experiment 1, only questions that were answered incorrectly on the initial test were included in the final test analyses in order to ensure that prior knowledge would not influence the results. On average, only 7.33% of the items (five items on average) was eliminated per participant as a result of being answered correctly (i.e., given full or partial credit) during the initial test. Therefore, final test accuracy was based on an average of 64 items per participant.

As in Experiment 1, final test accuracy was examined as a function of level of curiosity (high vs. low) and feedback timing (immediate vs. delayed). Each item was dichotomized as either a high-curiosity item (i.e., given a rating between 4 and 6) or a low-curiosity item (i.e., given a rating between 1 and 3). Eleven participants who did not contribute at least two items to both the high and low curiosity categories were excluded from this analysis[6].

Final test performance was analyzed using a series of planned comparisons that permitted analysis of the same 2 x 2 (Level of Curiosity: high vs. low x Feedback Timing: immediate vs. delayed) interaction from Experiment 1. Two separate 2 x 2 Mixed ANOVAs were conducted, with curiosity level as the within-subjects factor and feedback timing as the between-subjects factor. The first of these ANOVAs analyzed the effects of curiosity level (high vs. low) for participants in the Immediate vs. Constant Delay Groups (see Figure 3, upper panel), and the second of these ANOVAs analyzed the effects of curiosity level (high vs. low) for participants in the Immediate vs. Varied Delay Groups (see Figure 3, lower panel).
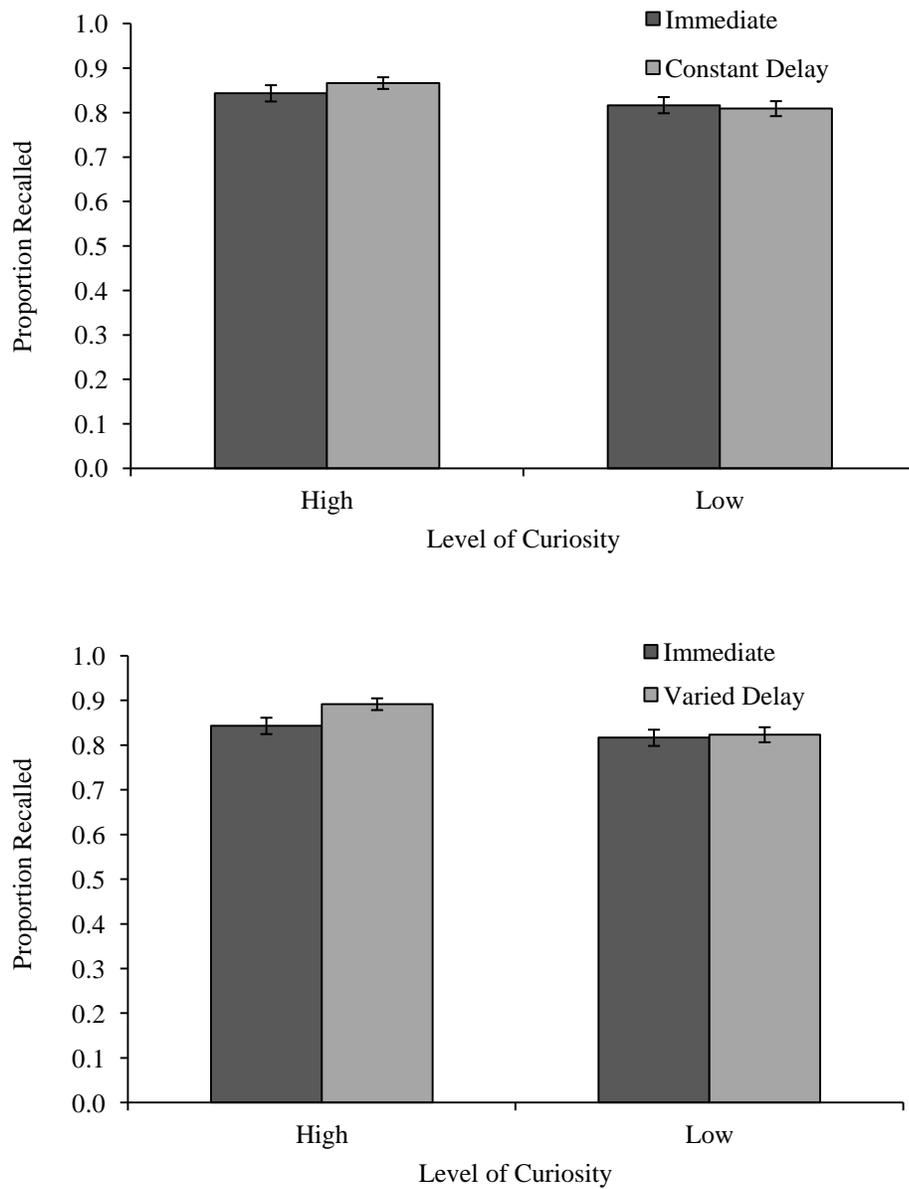
*Figure 3*. Results on the final test for Experiment 2. The upper panel depicts the accuracy for the Immediate and Constant Delay Groups. The lower panel shows the accuracy for the Immediate and Varied Delay Groups. A significant Level of Curiosity x Feedback Timing interaction was observed between the Varied Delay and Immediate Groups (lower panel) but not between the Constant Delay and Immediate Groups (upper panel).

While the main effect of feedback timing was not significant for either of the

analyses (both $F$s < 1.75), a significant main effect of level of curiosity was found for

both analyses (both $F$s > 12.25, both $p$s < .05). Participants achieved higher accuracy on

the final test for questions that were given high curiosity ratings compared to low

curiosity ratings, which is similar to the pattern that was seen in Experiment 1.

When the feedback delay was constant (Figure 3, upper panel), the Feedback

Timing x Curiosity Level interaction was not significant [$F(1, 129) = 1.65$, $p > .05$].

Delayed feedback produced no benefit over immediate feedback for either high-curiosity

or low-curiosity items (both $t$s $< 0.90$).  However, when the feedback delay was varied

(Figure 3, lower panel), the Feedback Timing x Curiosity Level interaction approached

significance [$F(1, 129) = 3.62$, $p = .06$]. While delayed feedback produced no benefit

over immediate feedback for low-curiosity items [$t(129) = 0.28$, $p > .05$], a significant

benefit emerged for delayed feedback over immediate feedback for high-curiosity items

[$t(129) = 2.16$, $p < .05$].These results replicate the same interaction that was found in

Experiment 1[7].

Since participants were given unlimited time to guess the answer and rate their

curiosity for each question, a set of analyses was conducted to explore the possibility that

more time was spent answering the questions or rating curiosity in certain condition (e.g.,

Varied Delay High) compared to others . The means and standard deviations for each of

the groups are reported in Table 4.

Table 4

*Mean RT for guessing the answer (Answer RT) and rating curiosity (Curiosity RT) on the initial test as a function of level of curiosity for the Immediate, Constant Delay, and Varied Delay Groups in Experiment 2*

| Feedback Timing | Answer RT | | Curiosity RT | |
|---|---|---|---|---|
| | High Curiosity | Low Curiosity | High Curiosity | Low Curiosity |
| Immediate | 13059 (4873) | 11080 (3979) | 1824 (756) | 1455 (415) |
| Constant Delay | 12101 (4078) | 11034 (4489) | 1878 (541) | 1658 (499) |
| Varied Delay | 12982 (5864) | 12178 (6187) | 1981 (589) | 1870 (686) |

*Note.* Times are reported in ms. Standard deviations are given in parentheses.

The time that participants spent guessing the answers to the questions was analyzed using a series of 2 x 2 (Level of Curiosity: high vs. low x Feedback Timing: immediate vs. delayed) Mixed ANOVAs. The first analyzed potential differences in RTs between high-curiosity and low-curiosity items for participants in the Immediate vs. Constant Delay Groups, and the second analyzed potential differences in RTs between high-curiosity and low-curiosity items for participants in the Immediate vs. Varied Delay Groups. In both of these 2 x 2 analyses, there was no main effect of feedback timing and no significant interaction between level of curiosity and feedback timing (all $F$s < 3.96). For both analyses, the only significant effect was a main effect of level of curiosity [both $F$s > 13.47, both $p$s < .05], in that the time required to answer the initial question was longer for high-curiosity compared to low-curiosity items, just like in Experiment 1.

Next, the time that participants spent providing curiosity ratings to the questions was analyzed using the same 2 x 2 (Level of Curiosity: high vs. low x Feedback Timing: immediate vs. delayed) Mixed ANOVAs. For both analyses, a main effect of level of curiosity was found [both $F$s > 22.27, both $p$s < .05], in that the time required to provide

a curiosity rating was longer for high-curiosity compared to low-curiosity items. While there was no main effect of feedback timing when comparing the Immediate and Constant Delay Groups, [$F(1, 129) = 2.58$, $p > .05$], there was a significant main effect when comparing the Immediate and Varied Delay Groups [$F(1, 129) = 10.35$, $p < .05$] in that participants in the Varied Delay Group spent longer giving curiosity ratings than participants in the Immediate Group. No significant interaction was found between level of curiosity and feedback timing for either of the analyses (both $F$s < 4.02).

As in Experiment 1, it appears that the time spent on each trial during the learning phase did not coincide directly with the pattern of memory accuracy that was observed on the final test. Thus, the effects of curiosity and feedback timing—particularly the interaction between these two variables that was observed in the Varied Delay Group but not in the Constant Delay Group—cannot be generally attributed to any systematic differences in the time spent trying to guess the answer or rate curiosity during learning.

*Effects of Different Feedback Delay Durations*

Given the significant advantage in final test accuracy of the Varied Delay Group over the Immediate Group for high-curiosity items, an additional set of analyses was conducted to separately explore the effects of the three different varied delay durations (i.e., 2-second, 4-second, and 8-second delays) on final test accuracy. Memory accuracy for the Immediate Group and each of the three Varied Delay Conditions was compared for both high-curiosity and low-curiosity items using a series of planned comparisons. Paired-samples t-tests were used to compare the 2-Second Delay, 4-Second Delay, and 8-Second Delay Conditions to one another, and independent-samples t-tests were conducted to compare the Immediate Group to the each of the three Varied Delay

Conditions. Figure 4 presents the mean proportion of answers recalled correctly on the

final test for the Immediate Group and the three delay durations of the Varied Delay

Group as a function of whether items were given high curiosity ratings (upper panel) or

low curiosity ratings (lower panel).
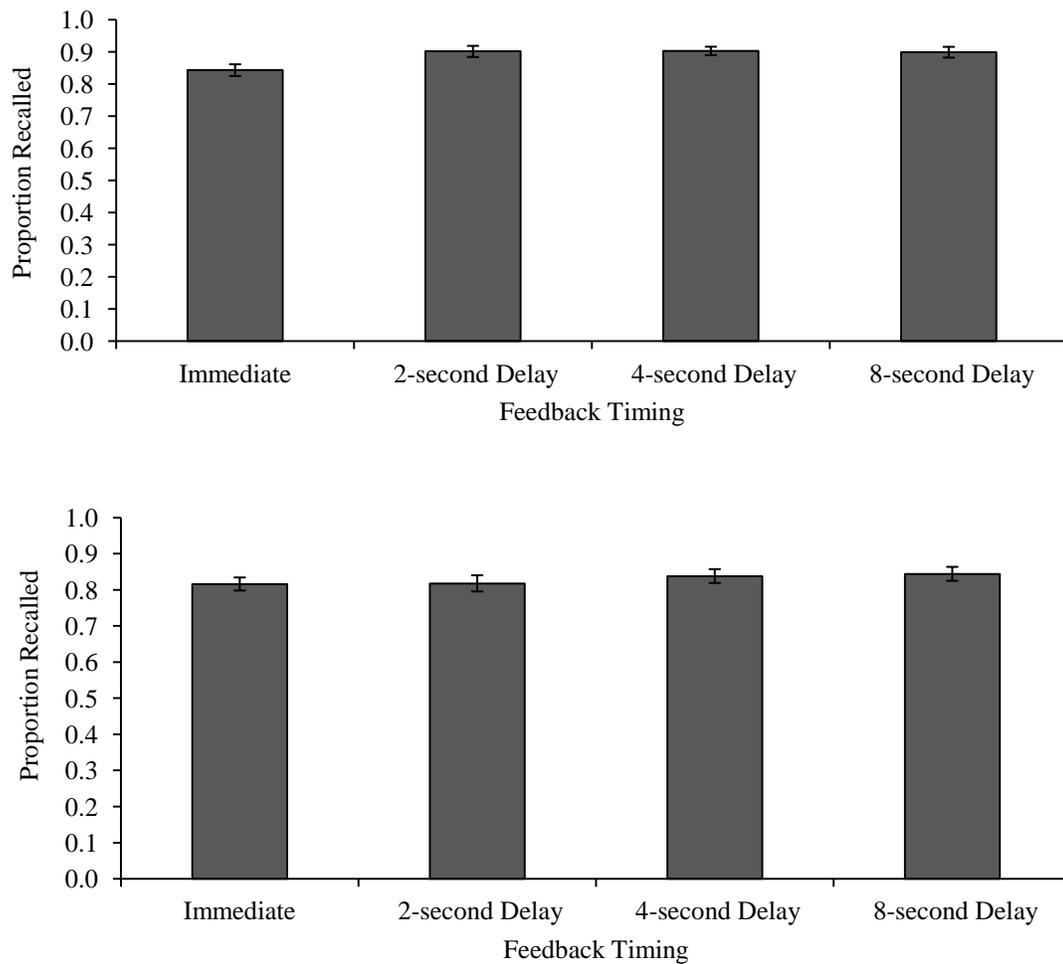




*Figure 4*. Results on the final test for Experiment 2. The upper panel shows the accuracy for the Immediate Group, and 2-Second Delay, 4-Second Delay, and 8-Second Delay Conditions for items rated high in curiosity. The low panel depicts the accuracy for the Immediate Group, and 2-Second Delay, 4-Second Delay, and 8-Second Delay Conditions for items rated low in curiosity. No significant differences were found among the conditions for the low-curiosity items (lower panel). A significant advantage of each of the delay conditions was found over the Immediate Group for the high-curiosity items (upper panel).

There were no significant differences when comparing the four conditions for the low-curiosity items (all $t$s < 1.04, all $p$s > .05), When comparing the four conditions for the high-curiosity items, however, it was found that that accuracy in the 2-, 4-, and 8-Second Delay Conditions was higher than accuracy for the Immediate Group (all $t$s > 2.22, all $p$s < .05), but no significant differences were observed among items that received 2-second, 4-second, or 8-second delayed feedback (all $t$s < 0.27, all $p$s > .05).

*Gamma Correlations*

A gamma correlation between curiosity ratings given during the initial test and accuracy on the final test was also computed for each participant. In order to compute a gamma correlation, participants had to produce more than one value for both curiosity rating and accuracy (e.g., they could not rate all of the questions at a curiosity level of 1 or recall all of the items correctly on the final test). Four participants were eliminated due to not meeting these requirements. In the Immediate Group, one participant was eliminated due to giving a rating of 1 to all of the questions, in the Constant Delay Group two participants were eliminated because they answered all of the questions correctly, and in the Varied Delay Group one participant gave a rating of 6 to all of the questions.

A One-Way Between-Subjects ANOVA indicated that there were no differences in mean gamma correlations among the Immediate, Constant Delay, and Varied Delay Groups [$F(2, 201) = 0.75$, $p > .05$]. While no significant differences were found, the overall pattern of results (see Figure 5) supports the findings from the 2 x 2 ANOVAs reported above (i.e., a Level of Curiosity x Feedback Timing interaction between the Varied Delay and Immediate Groups, but not between the Constant Delay and Immediate Groups) in that there was a stronger positive association between retention and level of

curiosity for the Varied Delay Group compared to the Constant Delay and Immediate Groups.
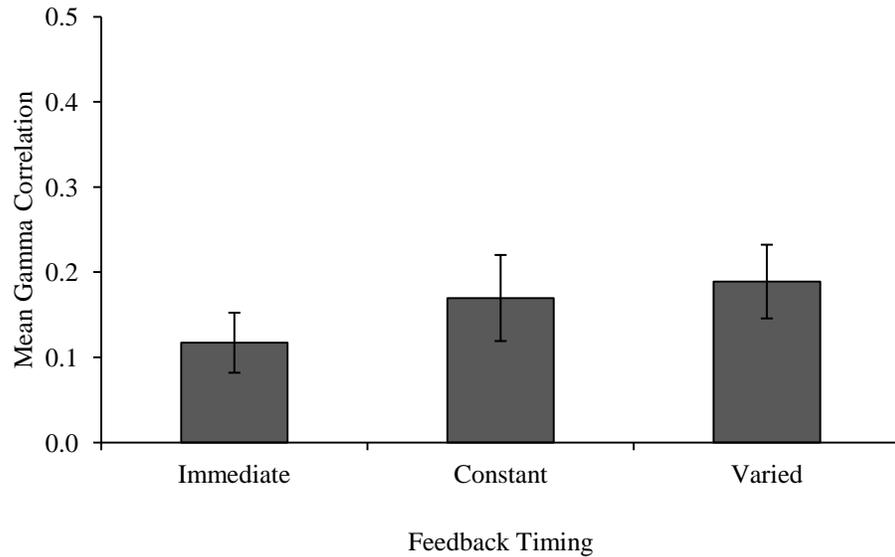


*Figure 5.* Mean gamma correlations between curiosity rating and accuracy on the final test for Experiment 2 for the Immediate, Constant Delay, and Varied Delay Groups.

Separate gamma correlations between curiosity ratings and accuracy on the final test were then computed for the 2-Second, 4-Second, and 8-Second Delay Conditions in the Varied Delay Group. A total of 32 participants were eliminated due to not producing more than one value for both curiosity and accuracy—one from the Immediate Group, two from the Constant Delay Group, and 29 from the Varied Delay Group (9 of whom did not produce at least two values for the 2-Second Delay Condition, 9 for the 4-Second Delay Condition, and 11 for the 8-Second Delay Condition).

A series of planned comparisons were conducted to compare the gamma correlations for participants in the Immediate Group, Constant Delay Group, and each of the Varied Delay Conditions (i.e., 2-second, 4-second, and 8-second delays). Paired-

samples t-tests were conducted to compare each of the Varied Delay Conditions to one another, and independent-samples t-tests were used to compare both the Immediate and Constant Delay Groups to each of the Varied Delay Conditions. These planned comparisons indicated that there were no differences in mean gamma correlations among the Immediate Group, Constant Delay Group, 2-Second, 4-Second, and 8-Second Delay Conditions (all $ts < 1.47$, all $ps > .05$). While no significant differences were found, the overall pattern of results (see Figure 6) supports the idea that the strongest relationship between level of curiosity and accuracy on the final test was seen for the 2-Second Delay Condition.
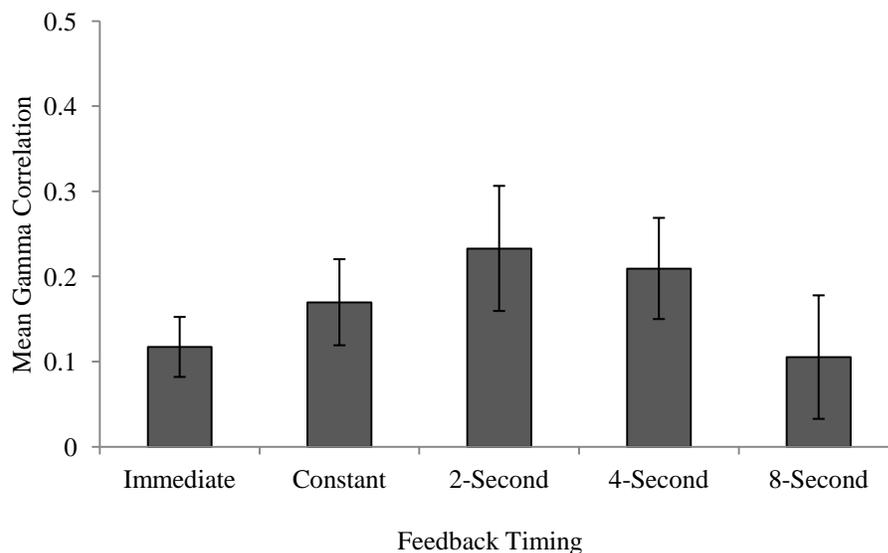


*Figure 6.* Mean gamma correlations between curiosity rating and accuracy on the final test for Experiment 2 for the Immediate Group, Constant Delay Group, and the 2-Second, 4-Second, and 8-Second Delay Conditions in the Varied Delay Group.

The results of Experiment 2 confirm those of Experiment 1 in that delaying feedback by a few seconds was more beneficial than providing it immediately, but only for items that aroused high curiosity. Both experiments demonstrated a significant

Feedback Timing x Level of Curiosity interaction, whereby the benefits of delayed

feedback over immediate feedback were stronger for items that received high curiosity

ratings (i.e., 4 and above) than for those that received low curiosity ratings (i.e., 3 and

below). Experiment 2 further demonstrated that this interaction held only when the

duration of the feedback delay was varied rather than constant, suggesting that the

observed benefits of feedback delay are obtained only under conditions in which

participants are not aware in advance of how long the feedback delay will be.

**CHAPTER 4. GENERAL DISCUSSION**

The current study sheds new light on the delay-of-feedback benefit, and the conditions under which brief delays of feedback are most likely to benefit memory retention. In both experiments, participants attempted to answer a general knowledge question, and then rated their curiosity to know the answer to that question before receiving feedback that was either provided immediately or delayed by several seconds. In Experiment 1, a final test that was given after five minutes revealed superior memory retention for items that were previously given feedback that was delayed by four seconds rather than provided immediately. Importantly, however, this delay-of-feedback benefit was only observed for those items that received high curiosity ratings from participants (i.e., a 4 or higher on a 1-6 point scale). For items that received low curiosity ratings (i.e., a 3 or lower), no delay-of-feedback benefit emerged, such that items that had received immediate feedback were retained just as well on the final test as those that had received delayed feedback.

These results are consistent with a number of other studies showing that brief delays of feedback can be beneficial for learning information such as U.S. city-state pairs (Sturges et al., 1968), course material (e.g., Rankin & Trepper, 1978; Sassenrath & Yonge, 1969), and face-name pairs (Carpenter & Vul, 2011). However, the current study is the first known study to identify the conditions under which this effect is most likely to emerge. Specifically, items that arouse higher levels of curiosity are more likely to benefit from feedback delays than are items that arouse lower levels of curiosity.

This finding is consistent with the reasoning put forth by Carpenter and Vul (2011), who proposed that a short feedback delay may be beneficial for memory because

it encourages anticipatory processing that results in a heightened degree of arousal or attention that is applied to the feedback when it occurs. Such processing would be less likely to occur in the case of immediate feedback, because participants are not given a chance to anticipate anything about the answer. In support of this hypothesis, Carpenter and Vul (2011) found that manipulations designed to reduce this anticipatory processing (e.g., performing a distracting task during the delay) eliminated the benefit of delayed feedback over immediate feedback.

If anticipatory processing drives the delay-of-feedback benefit in this type of paradigm, then any circumstance in which anticipatory processing is reduced or eliminated should eliminate the delay-of-feedback benefit. The current study adds important new data that support this hypothesis by demonstrating that items that are more likely to be anticipated (i.e., those that arouse higher levels of curiosity) benefit significantly from delayed feedback, whereas those that are less likely to be anticipated (i.e., those that arouse lower levels of curiosity) do not benefit from delayed feedback. Within the same experiment, differences in the potential anticipation of the correct answer consistently determined whether or not the delay-of-feedback benefit would occur.

Experiment 2 demonstrated that the benefit of delayed feedback for high-curiosity items only occurs under conditions in which feedback is delivered after a variable time delay of either 2, 4, or 8 seconds, rather than a constant delay of 4 seconds. A feedback delay that is always the same duration may become predictable and cease to engage participants' attention as much as a delay that is ever-changing and less predictable. An ever-changing delay may be more likely to encourage a sense of ongoing anticipation, as

participants themselves cannot regulate their internal response before they receive the feedback, because they do not know exactly when it is coming. In support of this notion, there is at least one study that has shown a link between uncertainty about the timing of an upcoming event and physiological anticipation (Monat, Averill, & Lazarus, 1972). In this study, half of the participants were told that they would receive a shock at the end of a 3-minute period (*Time Known Group)* and half were told that they would receive a shock at some point during a 3-minute period *(Time Unknown Group).* The autonomic arousal (i.e., heart rate and skin conductance response) of participants in the Time Unknown Group was higher than that of the participants in the Time Known Group. These findings indicate that anticipation of an upcoming event is greater under conditions in which the timing of that event is uncertain rather than certain. To the extent that anticipation of feedback involves a similar type of physiological response as that measured in Monat et al.'s (1972) study, feedback delay durations that are unpredictable may produce a greater sense of anticipation and result in superior memory accuracy, as was found in Experiment 2.

One way in which this possibility could be explored is by more directly measuring participants' levels of anticipation while waiting for feedback. In the current study, anticipation was measured through self-ratings of curiosity to know the correct answer. A more precise way to measure anticipation of an upcoming event (i.e., feedback) that has been used in the past is to record skin conductance responses (SCRs) throughout the experiment (Behcara, Damasio, Tranel, & Damasio, 1997; Dawson, Schell, & Courtney, 2011; Monat et al., 1972). If participants benefit from a delay of feedback more when their SCRs are higher (i.e., are more likely to be anticipating the

answer) compared to when they are lower (i.e., are less likely to be anticipating the answer), this would further support the hypothesis that a delay is beneficial because it serves as a time for participants to anticipate or prepare for the upcoming answer.

The idea that anticipatory processing is a key ingredient for the effectiveness of delayed feedback could explain the tendency for the delay-of-feedback benefit to be rather small in some previous studies (e.g., Carpenter & Vul, 2011), and non-existent in other studies. For example, in Brosvic, Epstein, Dihoff, and Cook's (2006) study, participants learned Esperanto (i.e., a false, constructed language) vocabulary words (e.g., peza: heavy) and were then given a 50-item multiple-choice test over the definitions of the Esperanto words (e.g., peza: _____). One group of participants was provided with the correct answer immediately after guessing the answer, whereas another group was provided with the correct answer after a delay of 5 seconds. On a final test given one week later, no difference in accuracy was observed between delayed and immediate feedback. It is possible that participants in this study did not find the materials particularly interesting (e.g., perhaps because Esperanto is a false language that they would likely never have to use again), and therefore did not anticipate the feedback during the delay period. Future research is encouraged that can more systematically explore whether differences in the perceived interest of the materials reliably relate to differences in the size of the feedback-delay-benefit.

## Curiosity and Memory

Results from the current study support the findings of past research demonstrating a positive relationship between curiosity and later memory. One study that examined this relationship was conducted by Berlyne (1954), who gave participants an initial test over a

series of questions about animals, and were told to mark the questions that they were curious to know the answers to. They then read facts about each of the animals, some of which provided answers to the questions that were previously asked.  Immediately after reading the facts, participants were given a final test over the same set of questions. Accuracy on the final test was better for the questions that participants indicated they were curious about compared to questions that they were not curious about, which suggests that experiencing a sense of curiosity to know an answer can increase one's memory for that answer. A similar association between curiosity and memory has also been found when measuring retention of general knowledge questions (Kang Hsu, Krajbich, Loewenstein, McClure, Wang, & Camerer, 2009), authors of quotations (Berlyne,1966), short stories (Maw & Maw, 1961), and products presented in advertisements (Bull & Dizney, 1973). Results from the current study support the findings of these past studies, in that questions given high curiosity ratings were remembered better than those that were given lower curiosity ratings.

## Directions for Future Research

Another way that the results of the current study could be expanded is to use a longer retention interval, as fairly short retention intervals (i.e., 5 minutes in Experiment 1 and 15 minutes in Experiment 2) were used in the current study. It would be beneficial to examine if the same Level of Curiosity x Feedback Timing interaction occurs with longer retention intervals (e.g., after several days rather than minutes), because in real-world settings people typically want to remember the information they are learning for more than a few minutes (e.g., when studying for a test that will occur several days later).

Furthermore, future studies which seek to examine the effects of delayed feedback on retention of information should take into account the length of a delay implemented before providing feedback, as it might impact how beneficial the delay will be for later memory. Are longer delays always beneficial, or is there an optimal delay duration that produces the best memory retention for a given set of materials? One possible reason for the potential benefit of shorter delay intervals is that participants may grow bored, or become less interested (i.e., experience a decrease in curiosity), if they have to wait too long to receive the correct answer to a question that they were just asked. If participants do become bored during the delay, therefore, it is unlikely that they would be anticipating the upcoming feedback, which would then cause the delay to be less beneficial compared to if they were anticipating the feedback.

## Educational Implications

The results of the current study have important implications for both instructors and students. One goal that most, if not all instructors have is for their students to remember as much information from their classes as possible. One way to increase the chances of meeting this goal would be for instructors to incorporate delayed feedback into their lesson plans. A simple way that this could be done is by implementing a short (e.g., 1 to 4 seconds) variable delay before providing feedback during in-class or online quizzes. It is important to note, however, that delayed feedback is not beneficial under all conditions (i.e., when students are not likely to anticipate the feedback during the delay). Regardless of this limitation, it is probable that students would benefit from delayed feedback as they are likely to be very interested in, or curious to know, the answers to any questions that they are being asked because their grades depend on if they get the

answer correct or not. Therefore, it is likely that students will anticipate the feedback they are about to receive during the delay the majority of the time.

Not only can instructors implement delayed feedback into students' routines, students can also incorporate delayed feedback into their study routines themselves in order to improve their memories of the material they are trying to learn. For instance, many students make flashcards to prepare for upcoming exams (e.g., to learn definitions of vocabulary words). Students could study these flashcards with a partner who would wait several seconds after the answer was guessed before flipping the flashcard over to reveal the correct answer. Studying with a partner would ensure that the student would not know when the feedback would be given for each question (i.e., would experience uncertainty about the timing of feedback).

This flashcard method is a very simple way for students to implement delayed feedback during their test preparation that would add very little additional study time. As noted above, however, delayed feedback only appears to be beneficial when one is likely to experience as sense of anticipation during the delay. This limitation is not likely to be a problem for students using delayed feedback to study information for class however, because it is likely that the students will anticipate the correct answer, as it is essential for them to learn the correct answers in order to do well on the upcoming exam.

# FOOTNOTES

[1] This method allows participants themselves to determine the curiosity value for each item. An alternative method might involve pre-classifying items as either high or low curiosity based on ratings collected from a separate group of participants. However, given potential individual differences in curiosity, the method based on participants' own judgments on an item-by-item basis seemed to be the more sensitive and precise way to measure these effects.

[2] Of these seven eliminated participants, one did not contribute at least two items to the Delayed Low Condition, three did not contribute at least two items to the Delayed High Condition, two did not contribute at least two items to the Immediate High Condition, and one did not contribute at least two items to either the Immediate Low or Delayed Low Conditions.

[3] The same results emerged when final test accuracy was based on participants who contributed at least one response to each of the four conditions. In this case, the same seven participants who were eliminated from the main analysis (for failing to produce at least two responses to each of the four conditions) also failed to produce at least one response to each of the four conditions. Thus, the final test analysis was based on the same group of 25 participants, whether it was limited to participants who produced just one response per condition, or two responses per condition.

[4] Only six participants responded "yes" to this question. When the same set of 2 x 2 Mixed ANOVAs that are reported in the "Final Test" section were conducted again excluding these six participants (two from the Immediate Group, one from the Constant Delay Group, and three from the Varied Delay Group), the same pattern of significant

results emerged. For both analyses, a significant main effect was found for level of curiosity [both $F$s > 10.26, both $p$s < .05] (i.e., accuracy was higher for high-curiosity compared to low-curiosity items), but not for feedback timing [both $F$s < 1.73]. Also, the Feedback Timing x Curiosity Level interaction was not significant when they delay of feedback was constant [$F(1, 126) = 3.45, p > .05$]. The Feedback Timing x Curiosity Level interaction was significant, however, when the feedback delay was varied $F(1, 124) = 6.24, p < .05$] (i.e., delayed feedback produced no benefit over immediate feedback for low-curiosity items [$t(124) = 0.10, p > .05$], but there was a benefit of delayed feedback over immediate feedback for high-curiosity items [$t(124) = 2.36, p < .05$]).

[5] In the Immediate Group 61% of participants answered the question correctly (i.e., answered "Yes"), 71% of participants in the Constant Delay Group answered correctly (i.e., answered "Yes"), and 73% of participants in the Varied Delay Group answered correctly (i.e., answered "No"). The results of a Pearson's Chi-Square test indicated that the accuracy of the response did not depend significantly on the group to which the participant was assigned [$\chi^2(2) = 2.36, p > .05$], indicating that participants in any given group were no more likely to have answered the question correctly compared to those in the other two groups. Therefore, the majority of participants in the Immediate and Constant Delay Groups appeared to have been aware that the feedback was presented after the same amount of time on every trial, while those in the Varied Delay Group appeared to have been aware that the feedback was *not* presented after the same amount time on every trial.

[6]Three participants in the Immediate Group did not contribute at least two items to the Low-Curiosity Condition and two did not contribute at least two items to the High-Curiosity Condition. One participant in the Constant Delay Group did not contribute at least two items to the Low-Curiosity Condition and one participant did not contribute at least two to the High-Curiosity Condition. Two participants in the Varied Delay Group did not contribute at least two items to the Low-Curiosity Condition and two did not contribute at least two items to the High-Curiosity Condition.

[7] The same pattern of results emerged when final test accuracy was based on participants who contributed at least one response to the High- and Low-Curiosity Conditions. Results of the same 2 x 2 ANOVAs limiting the data to these 66 participants in the Immediate Group, 67 participants in the Constant Delay Group, and 67 participants in the Varied Delay Group revealed a significant main effect of level of curiosity (both $F$s > 10.93, both $p$s < .05), but not feedback timing (both $F$s < 0.55). When the feedback delay was constant, the Feedback Timing x Curiosity Level interaction was not significant [$F(1, 131) = 1.60, p > .05$] (i.e., delayed feedback produced no benefit over immediate feedback for either high-curiosity or low-curiosity items (both $t$s < 0.89). When the feedback delay was varied however, the Feedback Timing x Curiosity Level interaction was significant [$F(1, 131) = 4.93, p < .05$]. Planned comparisons showed that while delayed feedback produced no benefit over immediate feedback for low-curiosity items [$t(131) = 0.30, p > .05$], a significant benefit emerged for delayed over immediate feedback for high-curiosity items [$t(131) = 1.82, p = .07$, two-tailed, $p = .04$, one-tailed].

# REFERENCES

Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L., & McDermott, K. B. (2007). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22,* 861-876.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 276,* 1293-1295.

Berlyne, D. E. (1954). An experimental study of human curiosity. *British Journal of Psychology, 45,* 256-265.

Berlyne, D. E. (1966). Conditions of prequestioning and retention of meaningful material. *Journal of Educational Psychology, 57*(3), 182-132.

Botham, N. (2006). *The book of useless information.* New York, NY: Penguin Group.

Bourne, L. E. Jr. (1957). Effects of delay of information feedback and task complexity on the identification of concepts. *Journal of Experimental Psychology, 54,* 201-207.

Brackbill, Y., (1964). The impairment of learning under immediate reinforcement. *Journal of Experimental Child Psychology, 1,* 199-207.

Brackbill, Y., & Kappy, M. S. (1962). Delay of reinforcement and retention. *Journal of Comparative and Physiological Psychology, 55*(1), 14-18.

Brackbill, Y., Bravos, A., & Starr, R. H. (1962). Delayed-improved retention of a difficult task. *Journal of Comparative and Physiological Psychology, 55*(6), 947-952.

Brackbill, Y., Isaacs, R. B., & Smelkinson, N. (1962). Delay of reinforcement and the retention of unfamiliar, meaningless material. *Psychological Reports, 11,* 553-554.

Brosvic, G. M., Epstein, M. L., Dihoff, R. E., & Cook, M. L. (2006). Retention of Esperanto is affected by delay-interval task and item closure: A partial resolution of the delay-retention effect. *The Psychological Record, 56,* 597-615.

Bull, S. G., & Dizney, H. F. (1973). Epistemic-curiosity-arousing presquestions: Their effect on long-term retention. *Journal of Educational Psychology, 65,* 45-49

Carpenter, S. K., & Vul, E. (2011). Delaying feedback by three seconds benefits retention of face-name pairs: The role of active anticipatory processing. *Memory & Cognition.*

Carpenter, S. K., Pashler, H., Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology, 23*, 760-771.

Dawson, M. E., Schell, A. M., & Courtney, C. G. (2011). The skin-conductance response, anticipation, and decision-making. *Journal of Neuroscience, Psychology, and Economics, 4,* 111-116.

Fazio, L. K., Huelser, B. J., Johnson, A., & Marsh, E. J. (2010). Receiving right/wrong feedback: Consequences for learning. *Memory, 18*(3), 335-350.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49*(268), 732-764.

Hockman, C. H., & Lipsitt, L. P. (1961). Delay-of-reward gradients in discrimination learning with children for two levels of difficulty. *Journal of Comparative and Physiological Psychology, 54,* 24-27.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963-973.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989-998.

Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 68*, 505-512.

Lhyle, K. G., & Kulhavy, R. W. (1987). Feedback processing and error correction. *Journal of Educational Psychology, 79*, 320-322.

Lintz, L. M., & Brackbill, Y. (1966). Effects of reinforcement delay during learning on the retention of verbal material in adults. *Journal of Experimental Psychology, 71*(2), 194-199.

Maw, W. H., & Maw E. W. (1961). Information recognition by children with high and low curiosity. *Educational Research Bulletin, 40*, 197-224

Monat, A., Averill, J. R., & Lazarus, R. S. (1972). Anticipatory stress and coping reactions under various conditions of uncertainty. *Journal of Personality and Social Psychology, 24*(2), 237-253.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology, 31*(1), 3-8.

Perin, C. T. (1943). A quantitative investigation of the delay of reinforcement gradient. *Journal of Experimental Psychology, 32,* 37-51.

Pressey (1926). A simple apparatus which gives tests and scores – and teaches. *School and Society, 15,* 549-552.

Pubols, B. H. (1958). Delay of reinforcement, response perseveration, and discrimination reversal. *Journal of Experimental Psychology, 56*(1), 32-40.

Rankin, R. J., & Trepper, T. (1978). Retention and delay of feedback in a computer-assisted instructional task. *The Journal of Experimental Education, 46*(4), 67-70.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255.

Saltzman, I. J. (1951). Delay of reward and human verbal learning. *Journal of Experimental Psychology, 41,* 437-439.

Sassenrath, J. M., & Yonge, G. D. (1969). Effects of delayed information feedback and feedback cues in learning on delayed retention. *Journal of Educational Psychology, 60,* 174-177.

Schroth, M. L. (1995). Variable delay of feedback procedures and subsequent concept formation transfer. *The Journal of General Psychology, 122*(4), 393-399.

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

Skinner, B. F. (1958). Teaching machines. *Science, 128*, 969-977.

Sturges, P. T., Sarafino, E. P., & Donaldson, P. L. (1968). Delay-retention effect and informative feedback. *Journal of Experimental Psychology, 78,* 357-358.

Vargas, J. S. (1986). Instructional design flaws in computer-assisted instruction. *Phi Delta Kappan,* 738-744.

Waugh, N. C. (1967). Presentation time and free recall. *Journal of Experimental Psychology, 73*(4), 39-44.

## APPENDIX A. QUESTIONS USED IN EXPERIMENTS 1 AND 2

What does camera shutter speed "B" stand for? (Bulb)

What is IBM's motto? (Think)

What is the side of a hammer called? (Cheek)

How many spaces are on a Scrabble board? (225)

What was the top-rated television series from 1957-1961? (Gunsmoke)

What is a prestidigitator? (Magician)

What does a horologist measure? (Time)

What was the first novel written on a typewriter? (Tom Sawyer)

What size of shoe did Robert E. Lee wear? (4.5)

In what year was the Golden Gate Bridge first opened? (1937)

In what country did doughnuts originate? (Holland)

What was the biggest-selling toy in 1957? (The Hula Hoop)

What is the oldest known vegetable? (Pea)

What is the hottest chili in the world? (Habanero)

Where did voodoo originate? (Haiti)

What is a group of kangaroos called? (Mob)

What does a deltiologist collect? (Postcards)

What is a baby bat called? (Pup)

What does a librocubicularist do? (Reads in Bed)

How many muscles does a caterpillar have? (4000)

What was the first domesticated animal? (Goose)

Which country is the largest exporter of frogs' legs? (Japan)

What does the word calendar mean? (To Call Out)

The ruins of Troy are located in what country? (Turkey)

To which fruit family does an almond belong? (Peach)

Who was the first female monster to appear on the big screen? (Bride of Frankenstein)

What year in a marriage is the leather anniversary? (Third)

How many U.S. presidents were the only child in their families? (0)

What are the plastic things on the end of shoelaces called? (Aglets)

Who was the sun god of ancient Egypt? (Ra)

What is the smallest unit of time? (The Yoctosecond)

What does Pokémon stand for? (Pocket Monster)

What does karaoke mean in Japanese? (Empty Orchestra)

What was Adolf Hitler's favorite movie? (King Kong)

Which U.S. president wrote 37 books? (Theodore Roosevelt)

What is the average number of houses a person looks at before buying one? (8)

Where did the Beatles perform their first U.S. concert? (Carnegie Hall)

What is considered to be the sister language of English? (German)

The world's smallest painting is on the surface of what? (Grain of Corn)

What is the most common name for a goldfish? (Jaws)

Which zoo has the largest collection of animals in the world? (San Diego Zoo)

What color is a grasshopper's blood? (White)

How long did Leonardo da Vinci spend painting the Mona Lisa's lips? (12 Years)

Who coined the word "nerd"? (Dr. Seuss)

Who appeared on the cover of *Life* magazine more than anyone else? (Elizabeth Taylor)

What letter does not appear on the periodic table of elements? (J)

What comedic actor was voted the least likely to succeed in high school? (Robin Williams)

The most presidents have been born in which state? (Virginia)

Which U.S. city sells more popcorn than anywhere else in the country? (Dallas)

What is the world's most read comic strip? (Peanuts)

What do the Olympic rings represent? (The Continents)

Where did the yo-yo originate? (The Philippines)

What do phobatrivaphobics fear? (Trivia About Phobias)

What is the most popular ice cream flavor? (Vanilla)

In which season do most burglaries occur? (Winter)

What is the most preferred reading material for the bathroom? (Reader's Digest)

What animal cannot contract or carry the rabies virus? (Squirrels)

Which famous composer wrote "Twinkle Twinkle Little Star" at the age of five? (Mozart)

What is the longest one-syllable word in the English language? (Screeched)

What is the oldest word in the English language? (Town)

What is the top-grossing Disney movie of all time? (The Lion King)

Coca-Cola was originally what color? (Green)

What is the largest object ever found in the Los Angeles sewer system? (Motorcycle)

Which U.S. president was once a male model? (Gerald Ford)

Which country offered Albert Einstein its presidency in 1952? (Israel)

What song is sung most often in America? (Happy Birthday)

What is the most common name in the world? (Mohammed)

Which native Mexican group went to battle with wooden swords so as not to kill their enemies? (Toltecs)

Thurl Ravenscroft was the voice of what cartoon character? (Tony the Tiger)

What do dendrologists study? (Trees)

## APPENDIX B. WORD SEARCH USED IN EXPERIMENT 2

Try to find each of the 35 words in the puzzle. When you find each word, circle it in the puzzle and cross it out in the list below.

```
P  G  J  E  R  H  W  F  N  N  M  I  D  D  L  E  E  G  L  L
H  V  D  Z  F  W  T  H  P  M  F  G  J  R  F  U  Y  I  H  X
O  P  E  E  G  C  Q  R  R  N  S  B  M  X  R  O  W  C  I  R
N  W  Q  R  P  N  V  E  I  U  N  E  W  B  W  O  U  K  Z  W
E  D  G  N  Y  M  B  G  D  H  O  N  X  K  R  O  O  Z  T  D
R  I  D  I  C  B  U  S  R  D  S  C  L  R  C  N  E  E  U  X
L  C  R  R  U  T  A  J  C  T  Y  H  A  Q  U  I  E  T  O  E
F  R  O  R  O  J  R  S  S  E  P  A  R  G  X  P  G  K  C  P
S  H  X  H  X  P  W  A  L  I  P  R  O  H  Y  G  N  S  U  W
G  S  C  E  N  I  P  D  F  E  W  E  I  T  T  M  I  W  P  K
A  I  V  G  W  M  H  E  O  F  E  S  R  V  S  M  R  H  S  Z
A  L  H  C  R  A  P  W  D  M  I  P  E  C  E  T  Z  R  M  Q
I  O  R  F  O  L  D  E  R  O  X  C  J  V  E  R  K  E  Y  S
I  P  K  M  O  V  A  L  Z  N  S  D  C  J  E  N  Y  S  V  H
Y  A  R  N  O  L  G  D  Q  K  A  G  B  F  X  N  T  N  M  I
T  Y  V  M  K  U  M  E  K  E  T  T  S  E  X  O  B  A  J  G
L  I  S  I  Y  M  C  I  P  Y  X  S  Y  U  Z  B  L  I  C  R
T  O  T  D  J  B  Z  H  I  N  S  E  R  T  V  Y  I  L  K  O
A  S  G  N  C  E  F  E  N  E  R  G  Y  X  D  X  F  K  X  S
E  M  N  S  J  R  J  K  C  I  T  S  P  U  R  P  L  E  G  S
```

| | | | |
|---|---|---|---|
| arch | grapes | oval | seven |
| arrow | gross | percent | shirt |
| bench | insert | phone | sleep |
| boxes | jumped | polish | stick |
| couch | keys | purple | traffic |
| cups | lumber | quiet | type |
| dropped | middle | ring | very |
| energy | monkey | river | yarn |
| folder | nail | rubber | |