Delaying feedback by three seconds benefits retention of face–name pairs: the role of active anticipatory processing

Shana K. Carpenter · Edward Vul

Abstract In three experiments, we used face–name learning to examine the puzzling *feedback delay benefit*—the tendency for feedback to be more effective when it is delayed rather than presented immediately. In Experiment 1, we found that feedback presented after a 3-s blank screen was more effective than feedback presented immediately, even after controlling for the exposure time to the material. In Experiment 2, we replicated the benefit of a feedback delay even when participants were given extra time to view the feedback or to try to retrieve the answer, indicating that this benefit is specific to a delay *before* feedback. Finally, in Experiment 3, we showed that the 3-s delay is beneficial only if it involves a blank screen, not if the delay is filled with an unrelated distracter task. These results suggest that the feedback delay benefit in this paradigm could arise from an active anticipatory process that occurs during the delay.

Keywords Feedback delay · Feedback timing · Feedback delay benefit · Delay retention effect · Memory · Associative learning · Recall

S. K. Carpenter (✉)
Department of Psychology, Iowa State University,
W112 Lagomarcino Hall,
Ames, IA 50011, USA
e-mail: shacarp@iastate.edu

E. Vul
University of California,
San Diego, CA, USA

Several decades of memory research have left little doubt that feedback is essential to learning (e.g., Ammons, 1956; Butler & Roediger, 2008; Fazio & Marsh, 2009; Metcalfe & Kornell, 2007; Renner, 1964). Recent studies have shown that feedback is critical for error correction, since learners are very unlikely to correct errors they make while learning unless they are informed of the correct answers (e.g., Pashler, Cepeda, Wixted, & Rohrer, 2005; Pashler, Rohrer, Cepeda, & Carpenter, 2007). In the present study, we were concerned with a question that has received much less empirical attention: *When* should feedback be provided?

If feedback provides only corrective information, then the timing of feedback should not matter. In contrast, if the benefits of feedback are linked to reinforcement schedules, then feedback should be most effective when provided immediately (e.g., Pressey, 1926; Skinner, 1954, 1958). Interestingly, however, the results of some empirical studies have argued against both of these intuitive accounts, showing instead that *delaying* feedback by 24 hr can lead to better retention than can providing it immediately (e.g., Butler, Karpicke, & Roediger, 2007; Kulhavy & Anderson, 1972). For example, Metcalfe, Kornell, and Finn (2009) demonstrated that 6th grade children retained knowledge of new vocabulary better when they were given a vocabulary test and then received the correct answers one day later, as compared with taking the test and receiving the answers on the same day.

Given that the turnaround time for classroom exams is usually at least 24 hr, these studies provide some encouraging news that waiting 1 day to receive feedback about one's performance does not appear to harm students' retention of information, but instead might actually improve it.

What happens when feedback is delayed by only a very brief time interval? For example, when students use computer-assisted instructional software, should they receive an answer to a test question immediately, or after a few seconds? When a teacher uses review questions to prepare her students for an upcoming exam, should she read the correct answer immediately, or should she pause for a moment before giving students the answer? It is not clear from past research whether such short-term feedback delays are beneficial to retention. As will be detailed in the following paragraphs, the results of the few studies that have addressed this question are difficult to interpret and could have been driven by factors other than the timing of feedback, per se.

Early work by Brackbill and colleagues (e.g., Brackbill, Bravos, & Starr, 1962; Brackbill & Kappy, 1962; Brackbill & Lintz, 1967; see also Markowitz & Renner, 1966) reported superior retention in a pair discrimination task under conditions in which knowledge of the correct answer was delayed by 10 s during learning. In these experiments, third-grade children were presented with two line drawings of familiar items (e.g., a boat or star) and had to guess which one was correct. After making their selection, they were informed of which option was correct by the presence of a flashing light above the appropriate drawing. For some children, the flashing light occurred immediately after they made their selection, whereas for others it occurred after a 10-s delay. On a subsequent relearning test, the children who received feedback after 10 s outperformed those who received it immediately. Similar results were obtained by Sturges, Sarafino, and Donaldson (1968), who used the same type of apparatus and found that 10-s delayed feedback led to better relearning of U. S. state-capitol pairs than did immediate feedback.

One interpretive difficulty with these studies, however, is the fact that the material in the delayed feedback condition was presented for a longer amount of time than in the immediate feedback condition. During the 10-s delay interval, the materials remained exposed to participants and thus allowed 10 extra seconds of time to view and rehearse the items. The benefit of delayed feedback, therefore, could have been due to this extra rehearsal time and not to the delay of feedback, per se. These studies also assessed relearning as the dependent measure, and so it is unknown what effects, if any, feedback timing has on the proportion of information recalled. Although relearning is an important measure of memory performance, proportion recalled is more likely to be used as a performance measure in educational settings.

Although some studies have examined retention as a function of the proportion of answers that could be recalled after receiving either immediate or delayed feedback, these studies are difficult to interpret. For example, Berlyn (1966)

presented high school students with 28 quotes from famous historical figures from the last 200 years. Students were required to guess the correct author of the quote out of two or three alternatives. For some students, the correct answer was read to them immediately after they made their guess and before going on to the next question. Other students offered a guess for all 28 quotes, and then received feedback by going back through the list and hearing the name of the correct author that was associated with each quote. On an immediate final test requiring students to write down the author of each quote, those who received delayed feedback recalled more correct authors than did those who received immediate feedback. Along similar lines, O'Neil, Rasor, and Bartz (1976) found that the answers to multiple-choice general knowledge questions were retained better on an immediate retention test if participants received feedback for all items at the end of the initial test, rather than after each item on the initial test.

These studies appear to have confounded the effect of feedback delay with recency, however. Immediate feedback was provided after each item on the initial test, whereas delayed feedback was provided after all of the initial test questions had been answered. In both studies, the final test was then administered right after the presentation of the delayed feedback. The apparent advantage for delayed feedback over immediate feedback, therefore, could be nothing more than an advantage for the items that were encountered most recently before the final test.

Thus, despite several studies reporting benefits of short-term feedback delays on retention, it is difficult to know whether these effects were due to feedback timing, per se or to exposure time or recency effects. In the present study, we controlled for these design issues and explored whether short-term feedback delays are beneficial for retention, and what the theoretical mechanisms underlying these effects might be.

We explored the benefit of a brief (3-s) delay of feedback in three experiments on face–name learning. In all three experiments, we controlled for serial position effects by randomly assigning items to conditions, administering the feedback conditions in random order, and by requiring participants to complete a distracter task prior to the final test. In Experiment 1, we controlled stimulus exposure time, and in Experiment 2, we controlled the total time on task across the delayed and immediate feedback conditions. To our surprise, a 3-s delay prior to feedback was still beneficial, indicating that the delay itself confers an advantage. In Experiment 3, therefore, we explored whether this benefit is driven by the activity that occurs during the delay. Although a 3-s blank screen produced the feedback delay benefit, this benefit was eliminated when the delay was filled with an unrelated distracter task. Together, these experiments suggest that the feedback delay benefit is

driven by an active process in which participants engage during the delay period that may serve to better prepare them for the upcoming feedback.

Experiment 1

In Experiment 1, we compared retention of names that were learned through immediate versus delayed feedback while exposure time was controlled. Both immediate and delayed feedback conditions allowed participants 8 s to recall the name in response to a face, and then provided feedback in the form of the correct name for 2 s. For items learned through delayed feedback, all stimuli were removed during the 3-s delay interval, ensuring that the exposure to the material in both feedback conditions was 10 s. Later, participants were tested for their retention of names learned through immediate versus delayed feedback, as well as of items that received no feedback.

Method

Participants Thirty-one undergraduate students volunteered in order to fulfill partial requirements for an introductory psychology course at Iowa State University. Participants were tested individually on personal computers.

Materials Thirty-six color photographs of nonfamous faces (18 males and 18 females) were chosen from the online database at the University of Stirling (2003). In addition, 36 first names (18 males and 18 females) were chosen from the 100 most common U. S. names according to the 1990 Census (2008). Highly common names were avoided by sampling those that ranked between 50 and 80 (e.g., *Diane*, *Rose*, *Justin*, *Bruce*). For each participant, names within a gender were randomly assigned to faces within that same gender.

Design and procedure Participants first experienced a presentation phase in which all 36 face–name pairs were presented in the center of the computer screen, one at a time, for 5 s each. Immediately following, participants were tested over each face–name pair. In order to control the exposure time for each item, the test involved a specific time interval during which the face alone was presented in the center of the computer screen, with the instructions, "Please try to remember the name of this person." In Experiment 1, this time interval was 8 s. During this 8-s period, participants could not enter a response or advance the program. After the 8 s had passed, however, participants were given the instructions, "Type in your answer now, followed by the ENTER key." Participants were then allowed as much time as needed to enter a response, and

their responses appeared on the screen directly below the face. For all three experiments, we include analyses of response times (RTs) to verify that there were no systematic differences in overall time on task between the various conditions.

As soon as participants typed in their responses and pressed ENTER, they experienced one of three conditions: (a) delayed feedback, in which all stimuli disappeared from the screen for a 3-s time interval, followed by a re-presentation of the face and the correct name for a duration of 2 s; (b) immediate feedback, in which the face and correct name immediately appeared on the screen for a duration of 2 s; and (c) no feedback, in which all stimuli disappeared from the screen for a 1-s time interval, and the program then advanced to the next item. Figure 1 provides a graphic depiction of the feedback conditions. For each participant, a random set of 12 face–name pairs (six males and six females) was assigned to each condition.

Each of the 36 items was tested once in this fashion, and then the test phase repeated two more times. On each repetition, each item appeared in the same condition as before, and the order in which all items were presented was randomized and different for each participant. After completion of the third and final test phase, participants were asked to type in the names of as many U. S. states as they could think of within a 5-min time interval. Afterward, a final test was given in which all 36 faces appeared, one at a time in random order, and participants were asked to type in the name that belonged to each face. Participants were given unlimited time to enter their responses, and feedback was not provided.

Results and discussion

Scoring All responses across the three experiments were handscored. Responses were considered correct if they were an exact match with the correct name, or contained minor spelling errors (e.g., *Theresa* instead of *Teresa*; *Luis* instead of *Louis*). One-third of the responses from each of the three experiments was randomly selected and scored by two raters who were blind to the conditions in which each item had appeared. Interrater correlations were highly significant across all four test trials for each of the conditions (rs > .97, ps < .001). The remainder of the scoring was completed in blind fashion by a single rater.

Initial test Mean accuracy across all test trials, for all conditions, is shown in Fig. 2. A one-way repeated measures ANOVA revealed that, although accuracy did not differ between any of the three conditions on the first test trial, $F(2, 60) = 1.91$, $p > .05$, the no-feedback condition

Fig. 1 Feedback conditions used in Experiments 1, 2 and 3. The *x-axis* shows trial time, with *gray arrows* indicating the temporal duration of each event. All trials began with the presentation of the test face, followed by a timed prompt to enter the associated name. Participants were given unlimited time to respond after this prompt appeared, and an analysis of response times indicated no significant differences in the time required to respond to items in the immediate versus delayed feedback conditions
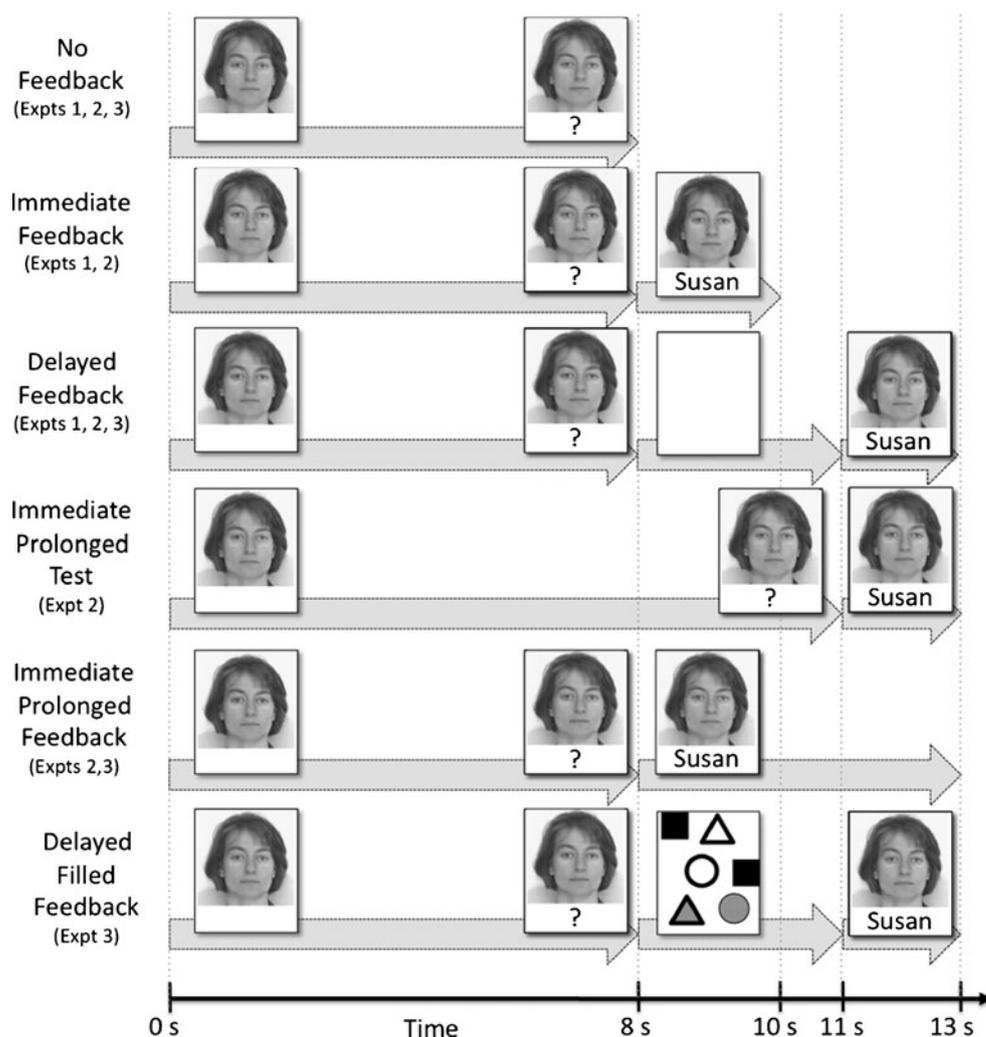
exhibited significantly poorer performance than did either of the two feedback conditions on the second test trial (both $ts > 5.96$, $p < .001$) and on the third test trial (both $ts > 10$, $p < .001$), with no significant differences between the two feedback conditions on either the second or third test trial (all $ts < 1$).

Since participants were given unlimited time to respond after the mandatory 8-s prompt, we analyzed RTs to test for systematic differences across conditions in the time participants spent responding. Table 1 contains the mean response times (RTs) across all three initial test trials for all conditions. A one-way repeated measures ANOVA revealed that RTs did not differ between any of the three conditions on the first and second test trials (both $Fs < 1.3$). The only significant difference that occurred was between delayed feedback and no feedback on the third test trial, $t(30) = 2.69$, $p < .02$. RTs, therefore, appeared to be fairly stable across the three test trials and were unlikely to have contributed to differences in accuracy between immediate versus delayed feedback conditions on the final test.

Final test Final test accuracy was significantly affected by learning condition, as revealed by a one-way repeated measures ANOVA, $F(2, 60) = 193.74$, $p < .001$, $MSE = .015$. Planned comparisons revealed that delayed feedback led to significantly higher accuracy than did immediate feedback, $t(30) = 2.96$, $p < .01$, and that both feedback conditions led to higher accuracy than did no feedback (both $ts > 13$, $ps < .001$).

Across all three experiments, we examined whether there was a different effect of feedback delay on answers that were initially correct versus incorrect. If delayed feedback is beneficial because it allows 3 s of additional time to rehearse the correct answer, then the feedback delay benefit would arise only for those items that were initially correct. Instead, we find that items that received an incorrect response on the first test showed the same results as in the overall analysis, $F(2, 60) = 268.96$, $p < .001$, $MSE = .018$. Delayed feedback ($M = .75$, $SD = .20$) was more effective than immediate feedback ($M = .68$, $SD = .24$), $t(30) = 3.23$, $p < .01$, and both feedback conditions were more effective than no feedback ($M = .03$, $SD = .06$) ($ts > 15$, $ps <$

Fig. 2 Proportion of names correctly recalled in each of the three experiments. Across experiments, the delayed feedback condition proves to result in greater final test accuracy than do all other conditions. See the text and Fig. 1 for a description and motivation for the conditions in each experiment. In the panel labeled Experiment 3*, the delayed filled feedback condition includes only those items for which participants completed all three trials of the distracter task. This excludes data for three participants who did not have more than one item meeting this criterion, and one whose distracter task responses were lost due to a programming error. *Error bars* represent +/− 1 within-subjects standard error of the mean (obtained by factoring out across-subject variation)
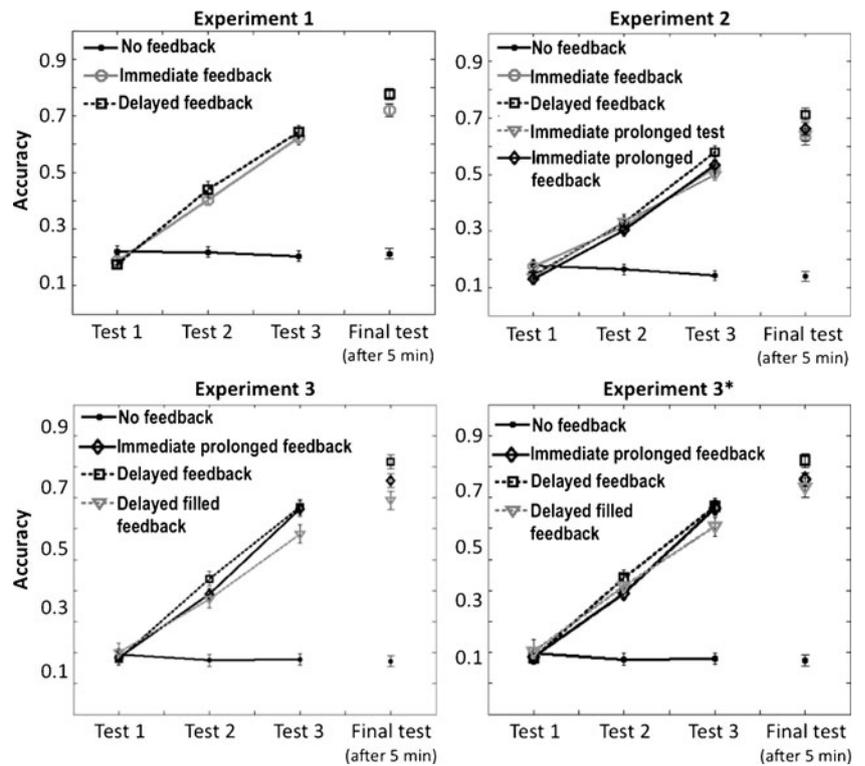
.001). Analyses based on initially correct responses also showed a slight advantage for delayed feedback ($M = .98$, $SD = .10$) over immediate feedback ($M = .97$, $SD = .09$), and for both feedback conditions over the no feedback condition ($M = .87$, $SD = .26$). However, these data should

be interpreted cautiously, since only 19% of responses were correct on the initial test.

The results of Experiment 1 confirmed the feedback delay benefit under conditions in which the exposure time to the stimuli was equated between delayed and immediate

Table 1 Mean response times and their standard deviations as a function of test trial and condition

Condition	Test 1		Test 2		Test 3	
	M	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1						
Delayed Feedback	5407.81	3540.12	4858.01	3522.32	3556.50	1534.41
Immediate Feedback	5674.96	3551.47	4261.72	2364.62	4138.20	2159.07
No Feedback	5282.60	3131.14	4487.47	2856.72	4607.53	2605.87
Experiment 2						
Delayed Feedback	4338.25	2786.24	3754.29	1798.24	3379.73	1417.97
Immediate Feedback	4224.52	3203.86	3948.78	1597.08	3521.86	1845.69
Immediate Prolonged Test	3905.60	2623.08	3453.72	1652.76	3588.55	2277.33
Immediate Prolonged FB	4120.40	2113.00	4051.50	2570.90	3992.00	2711.70
No Feedback	4153.64	2773.73	4007.19	2170.56	3854.28	2339.05
Experiment 3						
Delayed Feedback	4562.46	3251.35	4472.54	3541.17	3266.54	1697.73
Delayed Filled Feedback	4687.46	3369.57	3554.24	1879.64	3816.28	2721.70
Immediate Prolonged FB	5018.62	5230.38	3545.51	1732.78	3585.22	2360.39
No Feedback	4348.60	2959.46	3707.01	3166.30	3237.39	2043.66

feedback (see Brackbill & Kappy, 1962). By presenting the question prompt for 8 s, followed by a feedback presentation of 2 s, the results of Experiment 1 ensured that the exposure time to the stimuli was 10 s for delayed and immediate feedback. Although participants took some extra time to enter their responses after the prompt appeared (e.g., about 4 s, on average; see Table 1), the time required for participants to enter their responses did not vary systematically as a function of condition.

The presence of a 3-s blank screen in the delayed feedback condition added additional time, however, so that the total amount of time-on-task was actually 3 s longer in the delayed feedback condition than in the immediate feedback condition. It is possible that these three extra seconds afforded participants the opportunity to engage in processing that was beneficial to later memory, even if this benefit was not due to the delay of feedback, per se. Experiment 2 was designed to address this issue.

Experiment 2

Experiment 2 was designed to test two alternative explanations of the feedback delay benefit in Experiment 1. First, if participants are able to retrieve the correct answer, then the 3-s blank screen in the delayed feedback condition allows three extra seconds of time to rehearse the answer, as compared with immediate feedback. It is a well-known finding that additional rehearsal time is beneficial to memory (e.g., Bugelski, 1962), so the feedback delay benefit could be partially driven by extra time to think about the correct answer.

Second, regardless of whether participants recall the name initially, the subsequent 3 s of blank screen could allow them additional time to engage in retrieval practice before the correct answer appears. Recent research has shown that retrieval practice is beneficial to memory retention (e.g., Karpicke & Roediger, 2008), even under circumstances in which participants generate incorrect responses (e.g., Carpenter, 2009; Kornell, Hays, & Bjork, 2009). The delayed feedback condition in Experiment 1 could have provided participants with three additional seconds to try to generate the answer, relative to the immediate feedback condition.

To explore these possibilities, in Experiment 2, we replicated the same delayed and immediate feedback conditions from Experiment 1, in which an 8-s prompt was followed by a 2-s feedback presentation. In addition, two new conditions were added: (a) In the immediate prolonged feedback condition, participants were given an 8-s prompt followed by immediate feedback that lasted for 5s, and (b) in the immediate prolonged test condition,

participants were given an 11-s prompt, followed by immediate feedback that lasted for 2 s. This way, the total time on task was controlled (13 s, plus the time it takes to respond on each trial) for the delayed feedback, immediate prolonged feedback, and immediate prolonged test conditions (see Fig. 1). Note, however, that the delayed feedback condition is the only one that involves a delay between the participant's response and the presentation of feedback.

These conditions allow a direct exploration of the extent to which the benefits of delayed feedback are driven by factors such as rehearsal of the correct answer or time spent engaging in retrieval practice. If the advantage of delayed feedback over immediate feedback in Experiment 1 reflects nothing more than 3 s of additional time to rehearse the correct answer, then immediate prolonged feedback should produce the same benefit as delayed feedback, because the former presents the feedback for three additional seconds relative to the latter. If the feedback delay benefit reflects additional time spent engaging in retrieval practice, then the immediate prolonged test condition should produce the same benefit as delayed feedback, because it allows three additional seconds for participants to attempt retrieval of the answer. Finally, if the feedback delay benefit is driven specifically by the 3-s delay that occurs between entering a response and receiving exposure to the correct answer, then the delayed feedback condition should be superior to the other conditions, all of which involve immediate feedback.

Method

Participants Fifty-one undergraduate students were recruited from the same participant pool as in Experiment 1.

Materials The same 36 face–name pairs from Experiment 1 were used, plus two additional face–name pairs of each gender that were sampled from the same sources, yielding 40 face–name pairs in total.

Design and procedure Participants first saw a presentation of all 40 face–name pairs for 5 s each, followed by three test trials over each pair. A random set of eight items (four male face-name pairs and four female face-name pairs) was assigned to each of the five conditions for each participant. The delayed, immediate, and no feedback conditions were identical to those in Experiment 1. The two new conditions—immediate prolonged feedback and immediate prolonged test—were identical to the immediate feedback condition, except for the fact that the feedback was presented for 5 s instead of 2 in the immediate prolonged feedback condition, and the test prompt was presented for 11 s instead of 8 in the immediate prolonged test condition. All other procedural details were identical to those in Experiment 1.

Results and discussion

Initial test Mean accuracy and RTs are given in Fig. 2 and in Table 1, respectively. A one-way, repeated measures ANOVA revealed that, although accuracy did not differ between any of the five conditions on the first test trial, $F(4, 200) = 1.60$, $p > .16$, all of the feedback conditions outperformed the no feedback condition on the second test trial (all $ts > 4.4$, $ps < .001$) and on the third test trial (all $ts > 10$, $ps < .001$). The feedback conditions did not differ from one another on the second test trial (all $ts < 1$), but on the third test trial, the delayed feedback condition outperformed the immediate prolonged test condition, $t(50) = 2.45$, $p < .02$. RTs did not differ between any of the five conditions across the three test trials (all $Fs < 1.5$).

Final test Final test accuracy was significantly affected by learning condition, as revealed by a one-way repeated measures ANOVA, $F(4, 200) = 118.17$, $p < .001$, $MSE = .024$. Planned comparisons revealed that delayed feedback led to significant advantages over immediate feedback, $t(50) = 3.10$, $p < .01$, immediate prolonged test, $t(50) = 2.98$, $p < .01$, and immediate prolonged feedback, $t(50) = 1.93$, $p < .03$. However, none of the conditions involving immediate feedback differed from one another (all $ts < 1$). All of the feedback conditions outperformed the no feedback condition (all $ts > 12$, $ps < .001$).

This effect was again replicated for those items that received an incorrect response on the initial test, $F(4, 200) = 53.58$, $p < .001$, $MSE = .07$. Delayed feedback ($M = .69$, $SD = .29$) was superior to immediate feedback ($M = .58$, $SD = .33$), $t(50) = 3.92$, $p < .01$, immediate prolonged test ($M = .60$, $SD = .27$), $t(50) = 2.88$, $p < .01$, immediate prolonged feedback ($M = .63$, $SD = .28$), $t(50) = 1.90$, $p < .07$, and no feedback ($M = .03$, $SD = .08$). There were no significant differences between any of the immediate feedback conditions ($ts < 1.45$), and all feedback conditions outperformed the no feedback condition ($ts > 12.2$, all $ps < .01$). Analyses based on initially correct responses did not reveal any clear systematic differences between delayed feedback ($M = .93$, $SD = .25$), immediate feedback ($M = .92$, $SD = .25$), immediate prolonged test ($M = .96$, $SD = .19$), or immediate prolonged feedback ($M = 1.0$, $SD = 0$; all participants remained at 100% accuracy for initially correct items in this condition), although each of these conditions appeared to outperform the no feedback condition ($M = .65$, $SD = .43$). Again, however, these data should be interpreted cautiously, since they are based on only 15% of responses.

In Experiment 2, we replicated the same feedback delay benefit from Experiment 1, and further demonstrated that this effect does not appear to be driven by extra time to think about the correct answer or to engage in retrieval practice. Experiment 3 was conducted to further examine specific hypotheses for why the feedback delay benefit occurs in this paradigm.

Experiment 3

Why would a 3-s delay between a response and feedback be beneficial to later retention? One possibility is that a delay allows some forgetting of information that was generated during the initial retrieval attempt. Markowitz and Renner (1966) proposed that during a brief delay, participants begin to forget their own answer and thus pay more attention to the feedback when it appears. The likelihood of forgetting one's answer, and of benefitting from this increased attentiveness, is reduced when feedback is provided immediately.

Along similar lines, Kulhavy and Anderson (1972) suggested that a delay allows forgetting of error-related information that may have been generated during retrieval, reducing the chances that this information will interfere with encoding the correct answer when it appears. When feedback is immediate, error information does not have a chance to dissipate and may therefore be encoded in lieu of, or in addition to, correct information. Both of these accounts predict that the feedback delay benefit is a passive process, such that the key to the manipulation is the time between response and feedback, rather than any cognitive actions that participants may undertake during the delay.

In contrast, another possibility is that the feedback delay benefit is the consequence of an active cognitive process that occurs during the delay. After entering a response, participants may think about or anticipate some aspect of the answer, thus increasing their preparedness for the forthcoming feedback. When feedback is provided immediately, this preparedness has no chance to develop. In contrast with a passive account of the feedback delay benefit, an active account predicts that the feedback delay would cease to be advantageous if participants were occupied with an unrelated task during the delay.

In Experiment 3, we included the same delayed feedback and immediate prolonged feedback conditions from Experiment 2 (ensuring that the total time on task was controlled) and added a delayed filled feedback condition. This condition was identical in all respects to the delayed feedback condition, except for the fact that the 3-s delay involved a distracter task in which participants were asked to count an assortment of shapes (see Fig. 1).

If the benefit of 3-s delayed feedback is due to passive processes such as the forgetting of response- or error-related information, then inserting a distracter task during the delay should have no effect on this benefit. In fact, such a distracter task may even improve this benefit, by

increasing the probability that this information would be forgotten through interference. On the other hand, if the 3-s delay is beneficial because of some active cognitive process that occurs during the delay, then the distracter task in the delayed filled feedback condition would presumably suppress this activity, leading to an advantage of delayed feedback over delayed filled feedback.

Method

Participants Thirty-six participants were sampled from the same participant pool as in the previous experiments.

Materials, design, and procedure In Experiment 3, we used the same 40 face–name pairs from Experiment 2, with 10 items randomly assigned to one of the four conditions for each participant. The delayed feedback and immediate prolonged feedback conditions were identical to those in Experiment 2. For the delayed filled feedback condition, as soon as participants entered a response, they were shown a random assortment of 10 shapes on the computer screen. The shapes contained between one and four triangles, squares, and circles. Participants were instructed to count one of the shape types via instructions that appeared above the shapes (e.g., "How many triangles do you see? Press 1–4"). They did not know ahead of time which type of shapes they would be asked to count (triangles, circles, or squares).

At the beginning of the experiment, participants were given three practice trials on the shape counting task, and they were encouraged to try to respond as quickly as possible. Regardless of whether participants made a response to the shape-counting task during the delayed filled feedback trials, the program automatically advanced after 3 s to present the correct name for a duration of 2 s. This way, the delay interval and feedback duration were identical for the delayed feedback and delayed filled feedback conditions. The only difference was in how participants spent the time during the delay—doing nothing for 3 s in the delayed feedback condition, and counting shapes for 3 s in the delayed filled feedback condition. As in the previous experiments, we also included the no feedback condition. All other procedural details were identical to those in the previous experiments.

Results and discussion

Initial test Mean accuracy across all test trials, for all conditions, is displayed in the bottom left panel of Fig. 2. Accuracy did not differ between any of the four conditions on the first test trial, $F(3, 105) = .17$, $p > .05$; however, all of the feedback conditions outperformed the no feedback condition on the second test trial (all $ts > 5.85$, $ps < .001$) and on the third test trial (all $ts > 10.34$, $ps < .001$). The feedback conditions did not differ from one another on the second test trial (all $ts < 1.65$), although by the third test trial, delayed feedback was superior to delayed filled feedback, $t(35) = 2.19$, $p < .04$. Across the three test trials, RTs did not differ between any of the four conditions (all $Fs < 2.57$) (see Table 1).

Final test Final test accuracy was significantly affected by learning condition, $F(3, 105) = 159.87$, $p < .001$, $MSE = .02$, and planned comparisons revealed that delayed feedback ($M = .82$, $SD = .19$) yielded higher accuracy than did immediate prolonged feedback ($M = .76$, $SD = .21$), $t(35) = 2.55$, $p < .02$, or delayed filled feedback ($M = .69$, $SD = .26$), $t(35) = 3.49$, $p < .01$. There was no significant difference between the latter two, $t(35) = 1.92$, $p > .05$. All of the feedback conditions outperformed the no feedback condition (all $ts > 12.90$, $ps < .001$).

As in the previous experiments, we examined final test accuracy for responses that were correct versus incorrect on the initial test. This analysis revealed the same results reported previously, in that final test accuracy of initially incorrect responses was greater for delayed feedback ($M = .80$, $SD = .21$) than for immediate prolonged feedback ($M = .73$, $SD = .24$), $t(35) = 2.06$, $p < .05$. Delayed feedback was also superior to delayed filled feedback ($M = .66$, $SD = .27$), $t(35) = 3.50$, $p < .01$, but no significant difference emerged between immediate prolonged feedback and delayed filled feedback, $t(35) = 1.74$, $p > .05$. All feedback conditions were superior to no feedback ($ts > 14.43$, $ps < .001$).

Analyses based on initially correct responses revealed no clear systematic differences on the final test between delayed feedback ($M = .93$, $SD = .16$), delayed filled feedback ($M = .90$, $SD = .26$), and immediate prolonged feedback ($M = .96$, $SD = .13$), although all three feedback conditions appeared to be superior to no feedback ($M = .81$, $SD = .31$). Like in the previous experiments, however, no firm conclusions can be drawn from these data, since they are based on only 19% of responses.

In Experiment 3, we replicated the feedback delay benefit under conditions that controlled for the total amount of time on task, and demonstrated that the effect appears to be driven by the activity that takes place during the delay interval. A 3-s blank screen during the delay led to significant benefits on later retention, as compared with no delay. However, when the 3-s delay was filled with the shape-counting task, the feedback delay benefit was eliminated so that delayed feedback was no better than immediate feedback.

We undertook additional analyses of the delayed filled feedback condition to ensure that processing of the feedback in that condition was not disrupted by the distracter task. Out of the 30 total distracter trials during

learning (10 items in the delayed filled feedback condition that were tested three times), participants entered a response within the 3-s time window on 25.29 trials, on average (SD = 4.53). The fact that some distracter trials were not completed during the 3-s time window raises the possibility that on these trials, processing of the 2-s feedback presentation may have been disrupted by the incomplete distracter task.

We tested this possibility in two ways. First, for each participant, we calculated the total number of distracter trials that were completed within the 3-s time window, and correlated this number with final test accuracy in the delayed filled feedback condition. If failure to complete the distracter task within the 3-s time window indicates that the distracter task disrupted processing of the feedback, then one would expect final test performance to be worse when fewer distracter trials were completed. However, no such correlation was evident (r = .028, p >.87), indicating that the learning that occurred from feedback following the distracter task did not appear to depend on completion of the distracter task. This analysis excluded one participant for whom responses on the distracter task were lost due to a programming error.

In a second analysis, we assessed for each participant whether each of the 10 items in the delayed filled feedback condition received a response within the 3-s time window on each of the three training trials. We could thus identify for each participant the items in the delayed filled feedback condition that always received a response on the distracter task before the feedback was presented. We could then assess performance on just those items. For this analysis, we excluded three participants who did not have more than one item that met this constraint, and one whose distracter task responses were lost due to a programming error.

The bottom-right panel of Fig. 2 displays accuracy in the delayed filled feedback condition based on those items for which the distracter task was always completed within the 3-s time window. A one-way, repeated measures ANOVA revealed that accuracy did not differ between any of the four conditions on the first test trial, $F(3, 93)$ = .15, p > .05, but all of the feedback conditions outperformed the no feedback condition on the second test trial (all ts > 5.49, ps < .001) and on the third test trial (all ts > 9.95, ps < .001). The feedback conditions did not differ from one another on the second test trial (all ts < 1.79) or on the third test trial (all ts < 1.45).

Final test accuracy was significantly affected by learning condition, $F(3, 93)$ = 153.69, p < .001, MSE = .019, and planned comparisons revealed that delayed feedback led to significantly higher accuracy than did either immediate prolonged feedback, $t(31)$ = 2.07, p < .05, or delayed filled feedback, $t(31)$ = 2.32, p < .03, with no significant difference between the latter two, $t(31)$ = .68, p > .05. All

of the feedback conditions outperformed the no feedback condition (all ts > 13.79, ps < .001).

It appears, therefore, that delayed feedback was still superior to delayed filled feedback, even under conditions in which the distracter task in the delayed filled feedback condition was completed 100% of the time and was unlikely to have disrupted processing of the feedback that was presented afterward.

General discussion

In three experiments, we demonstrated superior face–name retention under conditions in which feedback—that is, presentation of the correct name—was delayed by 3 s after participants tried to retrieve the answer. This replicates a number of studies that have reported similar benefits for delaying feedback by 10 s in a pair discrimination task (e.g., Brackbill & Kappy, 1962; Brackbill et al., 1962; Markowitz & Renner, 1966), or by a few minutes on a multiple choice task (e.g., Berlyn, 1966; O'Neil et al., 1976).

In the present experiments, we controlled for confounding factors present in these previous studies. We replicated the benefit of a short-term feedback delay even when controlling for stimulus exposure time (Experiment 1), total time-on-task (Experiments 2 and 3), and serial position effects (Experiments 1, 2 and 3). Together, these results indicate that there is something particularly useful about a brief delay between the participant's response and corrective feedback. The present article helps advance theoretical development in this area by demonstrating that reliable short-term feedback delay benefits can be obtained after controlling for these factors, and that these benefits appear to vanish if the delay period is filled with a distracting task.

A delay filled with a distracter task failed to provide a benefit over immediate feedback (Experiment 3). This indicates that passive processes such as forgetting of response- or error-related information (e.g., Kulhavy & Anderson, 1972; Markowitz & Renner, 1966) are unlikely to be the cause of the feedback delay benefit in this paradigm. Instead, the benefit seems to arise from active cognitive processes that occur during the delay period and that are disrupted by a distracter task. Taken along with the results of Experiments 1 and 2, it appears that these active processes operate only after participants make a response and before feedback is presented, since additional time before the response or after feedback confers no such advantage.

There are several candidate cognitive processes that may underlie this benefit. First, the delay interval may serve to

heighten participants' curiosity to know the correct answer. Berlyn (1966, 1968) proposed that answering a test question induces *epistemic curiosity*—a heightened sense of arousal that results from an inquisitive drive to know the answer. As long as the answer is unknown, the drive will persist. Discovering the answer, however, provides reinforcement by reducing this drive. According to Berlyn, "the higher the drive before such reduction, the greater the amount of reinforcement" (1954, p. 183). It is possible, therefore, that the feedback delay interval is beneficial because it allows time for this curiosity to grow stronger, which results in the reinforcement (i.e., feedback) having a more powerful effect because it is delivered during a time when the drive to know the answer is greater. Such curiosity has less of an opportunity to develop in the case of immediate feedback, or in the case of delayed feedback that is preceded by a distracter task that consumes attentional resources.

A second possibility is that participants may use the delay interval to think about various aspects of their answer, making it possible that their own response acts as a cue for remembering the correct response. For example, Brackbill and Kappy (1962) reported that participants in their study sometimes continued to vocalize their response over the delay interval. Such rehearsal of one's own response— whether overt or covert—may benefit retention by acting as a bridge or "mediator" between the stimulus and the correct answer. Indeed, recent work by Pyc and Rawson (2010) has demonstrated that mediators generated by participants (i.e., information that links a cue to a target) play a critical role in successful retrieval of a target from a cue. The rehearsal of this mediating information would seem less likely in the case of immediate feedback and would be diminished in the case of delayed feedback that is preceded by a distracter task.

It is also possible that thinking about one's initial answer over an unfilled delay increases subjective confidence in that answer, and the subsequent feedback may be more effective because of hyper-correction effects—the tendency for high-confidence errors to be corrected more often than low confidence errors (Butterfield & Metcalfe, 2001). Participants may also actively attempt to retrieve the face stimulus during the unfilled delay interval, so that the face stimulus benefits from retrieval practice (e.g., Carpenter, Pashler, & Vul, 2006; Karpicke & Roediger, 2008) and may serve as a better cue for the correct name.

Finally, another possibility is that after making a response, participants can continue generating guesses. Vul and Pashler (2008) found that participants have additional information beyond that provided by the first guess. Thus, a 3-s delay interval could provide participants with an opportunity to generate more guesses, increasing the probability that the correct answer might be retrieved and reinforced by the feedback.

These explanations need not be mutually exclusive, and still other explanations could exist for the benefits of delaying feedback by 3 s during learning. The same theoretical mechanism(s) may not be expected to generalize to longer feedback delays of 1 day or more, however (e.g., Metcalfe et al., 2009). In such paradigms, feedback delay is manipulated over time periods of at least 24 hr, which by their very nature are filled with other activities that would render anticipatory processing unlikely. It is quite likely that the concept of active anticipatory processing is linked to paradigms involving relatively brief delays similar to those used in the present study. Future research will hopefully shed new light on the precise nature of this processing, and how best to utilize it to optimize retention and reduce forgetting.

References

Ammons, R. B. (1956). Effects of knowledge of performance: a survey and tentative theoretical analysis. *The Journal of General Psychology, 54*, 279–299.

Berlyn, D. E. (1954). A theory of human curiosity. *British Journal of Psychology, 45*, 180–191.

Berlyn, D. E. (1966). Conditions of prequestioning and retention of meaningful material. *Journal of Educational Psychology, 57*, 128–132.

Berlyn, D. E. (1968). Effects of prior guessing on intentional and incidental paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 7*, 750–759.

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

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

Brackbill, Y., & Lintz, L. M. (1967). The delay-retention effect in children: roles of reward and knowledge of results. *The Journal of Psychology, 66*, 149–155.

Bugelski, R. R. (1962). Presentation time, total time and mediation in paired associate learning. *Journal of Experimental Psychology, 63*, 409–412.

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

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

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

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 35*, 1563–1569.

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13*, 826–830.

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

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.

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, 63*, 505–512.

Markowitz, N., & Renner, K. E. (1966). Feedback and the delay-retention effect. *Journal of Experimental Psychology, 72*, 452–455.

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

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

O'Neil, M., Rasor, R. A., & Bartz, W. R. (1976). Immediate retention of objective test answers as a function of feedback complexity. *Journal of Educational Research, 70*, 72–75.

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

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

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

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science, 330*, 335.

Renner, K. E. (1964). Delay of reinforcement: a historical review. *Psychological Bulletin, 61*, 341–361.

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.

U. S. Census Bureau. *Frequently occurring first names and surnames from the 1990 Census*. Retrieved December 19, 2008, from http://www.census.gov/genealogy/names/

University of Stirling. (2003). *The psychological image collection at Stirling* (*PICS*). Retrieved August 8, 2003, from http://lionel.stir.uk

Vul, E., & Pashler, H. (2008). Measuring the crowd within: probabilistic representations within individuals. *Psychological Science, 19*, 645–647.