

Directed forgetting and feedback in written instruction

JAMES M. WEBB, WILLIAM A. STOCK, RAYMOND W. KULHAVY,
ROBERT C. HAYGOOD, D. N. D. ZULU, and DANIEL H. ROBINSON
Arizona State University, Tempe, Arizona

In three experiments, we examined the effects of directed forgetting messages on item recall. Using a variety of materials and conditions, we observed no main effects of such messages on recall. However, analyses of question and feedback study time, response certitude, and conditional probability revealed a number of significant effects. Effects of feedback were consistently strong and may have masked the benefits of directed forgetting. Other results were consistent with previous research in the feedback paradigm.

Bjork proposed that being able both to retain and to forget information is important to human learning (e.g., Bjork, 1972; Geiselman, Bjork, & Fishman, 1983). Using a so-called directed forgetting paradigm, he has reported that to-be-remembered items are retrieved better when other items are designated as to-be-forgotten. Also, cues to forget items inhibit access to them. According to Bjork, set differentiation and selective rehearsal are encoding processes underlying forgetting. Set differentiation occurs when a cue to forget causes separate clustering of to-be-remembered and to-be-forgotten items. Selective rehearsal of the to-be-remembered items increases their likelihood of being retrieved later.

Set differentiation and selective rehearsal may also underlie an error correction process specified in a recent model of instructional feedback described by Kulhavy and Stock (1989), who proposed that learners use feedback to identify errors and new alternatives (suggesting set differentiation). Subsequently, learners are assumed to forget former incorrect responses and to strengthen new responses (selective rehearsal).

We conjecture that error correction as proposed by Kulhavy and Stock (1989) does depend on the processes of set differentiation and selective rehearsal posited by Bjork (1972). Thus, in a feedback paradigm, we predict that cues to forget errors and remember correct responses will increase the differentiation of errors and correct responses and facilitate posttest performance. Three separate investigations of this expectation are reported.

EXPERIMENT 1

Method

Subjects and Design. The subjects were 76 undergraduates participating for course credit. The design had a between-subjects factor, cue

We wish to thank Marston C. White for his assistance in collecting the data for this experimental series. Correspondence should be addressed to James M. Webb, Instructional Science Research Facility, Payne Hall 322, Arizona State University, Tempe, AZ 85287-0611.

conditions ("forget" or "remember" [$N = 38$] versus "noted" [$N = 38$]), and a within-subjects factor, test occasion (original versus posttest).

Procedure. Materials were presented via microcomputer (Behrens & Stock, 1990). Items were 45 eight-alternative fact questions randomly drawn from those reported by Nelson and Narens (1980) and locally normed by Webb, Stock, Kulhavy, and White (1990). Before presentation, we trained subjects to rate response certitude (Webb, Kulhavy, & Stock, 1990). In the first presentation of items, the subjects saw each item, responded to it, and rated their certitude of a correct response on a 5-point scale (1 = not certain; 3 = somewhat certain; 5 = absolutely certain; 2 and 4 not labeled). Next, the subjects viewed post-response messages. In one cue condition, the subjects were shown cues either to forget or to remember their responses (for error and correct responses, respectively). In the second condition, the subjects were always shown cues that their responses had been noted. Thereafter, all subjects received feedback (a question, with alternatives; the correct alternative was marked). A second presentation of items immediately followed the first. The subjects answered each question and rated their certitude. In all presentations of items, order was randomized.

Results

We examined overall correct performance. We also sorted original responses into four categories (corrects rated high in certitude [certitudes of 4 or 5]; corrects rated low [certitudes of 1, 2, or 3]; errors rated high; and errors rated low). For these categories, we examined mean study times for questions and feedback, mean posttest certitude, and conditional probability of correct posttest responding. All tests were conducted with $\alpha = .05$. Table 1 contains summary statistics for all measures.

Overall, correct responding was higher on the posttest [$F(1,74) = 667.27, MS_e = 26.07$]. We found this effect of feedback in each study.

On the average, subjects took more time on questions answered incorrectly [$F(1,67) = 36.59, MS_e = 6.85$], and with low certitude [$F(1,67) = 215.30, MS_e = 12.58$]. Accuracy (corrects versus errors) and level of certitude (high vs. low) [$F(1,67) = 6.67, MS_e = 4.67$] interacted, because for low-certitude responses, errors were studied 1.23 sec longer than corrects, whereas for high-certitude responses, they differed by 2.58 sec.

On the average, errors were followed by more feedback study time than corrects were [$F(1,67) = 44.30, MS_e$

Table 1
Mean Response Accuracy, Question and Feedback Study Times,
Posttest Certitude Estimates, and Conditional
Probabilities for Experiment 1

Response Category	Noted			Directed Forgetting		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Response Accuracy						
1st Test	38	30.08	5.20	38	30.05	6.07
2nd Test	38	43.79	1.77	38	43.47	2.48
Question Study Time						
RH	35	7.65	1.43	34	8.33	2.70
RL	35	15.20	4.39	34	14.65	3.73
EH	35	10.39	3.05	34	10.75	4.30
EL	35	15.68	3.68	34	16.64	4.71
Feedback Study Time						
RH	35	1.53	.35	34	1.77	.87
RL	35	2.56	1.48	34	2.37	1.81
EH	35	3.36	1.71	34	4.28	2.71
EL	35	2.61	1.04	34	3.20	1.46
Posttest Certitude Estimates						
RH	35	4.96	.06	34	4.97	.11
RL	35	4.62	.42	34	4.85	.34
EH	35	4.89	.26	34	4.87	.27
EL	35	4.65	.32	34	4.74	.35
Conditional Probabilities						
RH	35	.99	.02	34	.99	.04
RL	35	.96	.11	34	.99	.05
EH	35	.96	.12	34	.96	.09
EL	35	.95	.08	34	.95	.09

Note—RH = correct responses rated high in certitude at Time 1; RL = corrects rated low; EH = incorrect responses rated high in certitude; EL = incorrects rated low. The loss of observations was due to an absence of items in one of these four categories.

= 2.66]. An interaction between accuracy and level of certitude [$F(1,67) = 43.18$, $MS_e = 1.20$] confirmed a pronounced effect for high-certitude responses.

Nevertheless, although posttest certitudes were high, subjects were more certain of items originally answered correctly [$F(1,67) = 4.13$, $MS_e = .07$] and with high certitude [$F(1,67) = 36.64$, $MS_e = .08$]. Cue conditions and level of certitude [$F(1,67) = 5.81$, $MS_e = .05$] interacted. Cue groups did not differ on responses originally rated high in certitude, but did on those rated low (noted subjects had the lower mean). Finally, the conditional probability of being correct on the posttest was higher for original corrects than it was for errors [$F(1,67) = 5.77$, $MS_e = .01$].

EXPERIMENT 2

Because of evident ceiling effects, Experiment 2 was implemented. To reduce the elevation of scores on the posttest, in Experiment 2 we used more difficult items and separated the two presentations of items with an interpolated task.

Method

The subjects were 60 undergraduates participating for course credit. The design remained the same ($N = 30$ in each cue group). The proce-

dures duplicated those of Experiment 1, except that a 3-min arithmetic task separated the presentations of items, and 45 new items were randomly selected from among those with difficulties less than or equal to 50%.

Results

Measures and analyses duplicated those of Experiment 1 (see Table 2). Correct performance was higher on the posttest [$F(1,58) = 925.48$, $MS_e = 19.75$].

On the average, subjects took longer for questions answered incorrectly [$F(1,50) = 4.64$, $MS_e = 4.25$] and with low certitude [$F(1,50) = 49.28$, $MS_e = 16.20$]. An interaction between cue conditions and accuracy [$F(1,50) = 9.01$, $MS_e = 4.25$] occurred because subjects receiving "forget"/"remember" cues studied corrects .24 sec longer than they did errors, whereas those receiving "noted" cues studied errors 1.48 sec longer than they did corrects. Accuracy and level of certitude [$F(1,50) = 5.84$, $MS_e = 9.32$] interacted, because, for low-certitude responses, corrects were studied .41 sec longer than errors were, whereas for high-certitude responses, errors were studied 1.64 sec longer than were corrects.

For mean feedback study time, a significant interaction involving cue conditions, accuracy, and level of certitude [$F(1,50) = 10.15$, $MS_e = .42$] moderated an interaction between accuracy and level of certitude [$F(1,50) = 24.73$,

Table 2
Mean Response Accuracy, Question and Feedback Study Times,
Posttest Certitude Estimates, and Conditional
Probabilities for Experiment 2

Response Category	Noted			Directed Forgetting		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Response Accuracy						
1st Test	30	11.23	4.56	30	10.70	4.50
2nd Test	30	36.90	5.36	30	34.30	8.02
Question Study Time						
RH	26	10.51	3.55	26	9.44	2.98
RL	26	15.30	4.53	26	14.53	6.79
EH	26	12.95	3.78	26	10.28	3.32
EL	26	15.81	4.34	26	13.21	5.24
Feedback Study Time						
RH	26	1.69	.48	26	1.85	.60
RL	26	2.79	.88	26	2.01	.70
EH	26	2.87	1.65	26	2.59	.73
EL	26	2.50	.94	26	2.43	.77
Posttest Certitude Estimates						
RH	26	5.00	.02	26	4.88	.30
RL	26	4.24	.68	26	4.48	.54
EH	26	4.74	.35	26	4.53	.56
EL	26	4.26	.50	26	4.05	.64
Conditional Probabilities						
RH	26	.98	.05	26	.93	.17
RL	26	.86	.18	26	.88	.16
EH	26	.91	.21	26	.86	.20
EL	26	.80	.14	26	.73	.21

Note—RH = correct responses rated high in certitude at Time 1; RL = corrects rated low; EH = incorrect responses rated high in certitude; EL = incorrects rated low. The loss of observations was due to an absence of items in one of these four categories.

$MS_e = .42$], as well as the main effect of accuracy [$F(1,50) = 22.19$, $MS_e = .62$]. Subjects receiving "forget"/"remember" cues studied feedback less than those receiving "noted" cues. The former spent more time on errors than on corrects, regardless of certitude, in contrast with subjects receiving "noted" cues, who spent the most time on high-certitude errors and low-certitude corrects.

Posttest certitudes were higher for items originally answered correctly [$F(1,50) = 18.08$, $MS_e = .19$] and with high certitude [$F(1,50) = 76.96$, $MS_e = .19$]. An interaction between cue conditions and accuracy [$F(1,50) = 5.20$, $MS_e = .19$] was a consequence of a larger difference in ratings between corrects and errors for subjects receiving "forget"/"remember" cues (.39) than for those receiving "noted" cues (.12). Finally, conditional probabilities were higher for items originally answered correctly [$F(1,50) = 12.84$, $MS_e = .03$] and with high certitude [$F(1,50) = 27.81$, $MS_e = .02$].

EXPERIMENT 3

Because the subjects controlled study time in the first two studies, substantial processing of each item might have masked the effects of cues. In a final attempt to produce an effect for cues, we constrained the time available for studying questions and cues.

Method

The subjects were 60 undergraduates participating for course credit. The design remained the same ($N = 30$ in each cue condition). Methods and materials duplicated those of Experiment 2, except that presentation time was fixed at 10 sec per question and at 2 sec per cue. After display of an item, a prompt solicited a response. Response time (from the prompt) and feedback study time were not constrained.

Results

Data were scored and analyzed as before (see Table 3). Posttest scores were higher than original scores [$F(1,58) = 1893.98$, $MS_e = 11.43$]. On the average, subjects spent more time responding to questions answered incorrectly [$F(1,39) = 5.13$, $MS_e = .16$] and with low certitude [$F(1,39) = 44.54$, $MS_e = .20$].

Subjects receiving "forget"/"remember" cues studied feedback longer than those receiving "noted" cues [$F(1,39) = 4.67$, $MS_e = 8.27$]. Also, feedback study time was longer following items answered incorrectly [$F(1,39) = 4.83$, $MS_e = .74$] and with low certitude [$F(1,39) = 17.21$, $MS_e = .66$], although an interaction between accuracy and level of certitude [$F(1,39) = 17.31$, $MS_e = .66$] moderated these effects. For errors, feedback study time did not differ for high- and low-certitude responses (both means equaling 3.42 sec), whereas for corrects, high- and low-certitude responses differed measurably (means of 2.60 and 3.66 sec, respectively).

Posttest certitudes were higher for items originally answered correctly [$F(1,39) = 16.35$, $MS_e = .27$] and with

Table 3
Mean Response Accuracy, Response Latency, Feedback Study Time, Posttest Certitude Estimates, and Conditional Probabilities for Experiment 3

Response Category	Noted			Directed Forgetting		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Response Accuracy						
1st Test	30	9.83	5.50	30	9.67	4.16
2nd Test	30	36.37	3.85	30	36.87	5.00
Question Response Time						
RH	21	.78	.32	20	1.08	.36
RL	21	1.28	.46	20	1.56	.99
EH	21	1.03	.32	20	1.16	.53
EL	21	1.45	.51	20	1.62	.52
Feedback Study Time						
RH	21	2.03	.91	20	3.19	1.81
RL	21	3.14	1.63	20	4.20	2.14
EH	21	3.03	1.37	20	3.84	1.81
EL	21	3.00	1.53	20	3.86	1.41
Posttest Certitude Estimates						
RH	21	4.95	.14	20	4.75	.54
RL	21	4.40	.76	20	4.40	.78
EH	21	4.56	.52	20	4.43	.53
EL	21	4.21	.27	20	3.98	.64
Conditional Probabilities						
RH	21	.93	.16	20	.96	.10
RL	21	.87	.16	20	.85	.25
EH	21	.89	.17	20	.88	.19
EL	21	.78	.09	20	.77	.13

Note—RH = correct responses rated high in certitude at Time 1; RL = corrects rated low; EH = incorrect responses rated high in certitude; EL = incorrects rated low. The loss of observations was due to an absence of items in one of these four categories.

high certitude [$F(1,39) = 21.36$, $MS_e = .35$]. Finally, with respect to conditional probabilities, correct responses on the posttest were more likely for items originally answered correctly [$F(1,39) = 6.77$, $MS_e = .03$] and with high certitude [$F(1,39) = 12.89$, $MS_e = .03$].

GENERAL DISCUSSION

These studies failed to demonstrate a main effect for directed forgetting cues on overall correct performance. In three studies, subjects performed comparably in both cue conditions. In Experiment 3, there was an effect of "forget"/"remember" cues on feedback study time, with subjects receiving such cues devoting more time to feedback. Consequently, there was little substantiation of our original conjecture that the error correction mechanism proposed by Kulhavy and Stock (1989) corresponds directly to the basic memory processes proposed by Bjork (1972). However, we propose that additional studies be conducted wherein item, cue, and feedback processing time are all constrained.

Subjects studied for longer periods questions they subsequently answered incorrectly or to which they assigned low certitudes. Consistent with Kulhavy and Stock's (1989) model, feedback was studied longer following responses that were errors or rated low in certitude. Also in line with the model, high-certitude errors were generally associated with the longest feedback, but not question, study times.

Results dealing with conditional probability of correct responding on the posttest are also consistent with the feedback model. For the most

part, the subjects showed a substantial improvement in correctly answering the test items upon their second exposure to them. The subjects made fewer errors and were more confident of their answers.

REFERENCES

- BEHRENS, J. T., & STOCK, W. A. (1990). *PRESENTER: Microcomputer programs for text-based experiments* (Tech. Rep. ZN-16). Tempe, AZ: Arizona State University, Instructional Science Research Facility.
- BJORK, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217-235). Washington, DC: Winston.
- GEISELMAN, R. E., BJORK, R. A., & FISHMAN, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, *112*, 58-72.
- KULHAVY R. W., & STOCK, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*, 279-308.
- NELSON, T. O., & NARENS, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning & Verbal Behavior*, *19*, 338-368.
- WEBB, J. M., KULHAVY, R. W., & STOCK, W. A. (1990). *Developing a certitude module* (Tech. Rep. ZN-14). Tempe, AZ: Arizona State University, Instructional Science Research Facility.
- WEBB, J. M., STOCK, W. A., KULHAVY, R. W., & WHITE, M. C. (1990). *Feeling-of-knowing ratings for 240 general information facts* (Tech. Rep. ZN-15). Tempe, AZ: Arizona State University, Instructional Science Research Facility.

(Manuscript received May 8, 1990.)

Notices and Announcements

20th Annual Meeting of the Society for Computers in Psychology New Orleans, Louisiana November 15, 1990

The 20th Annual Meeting of the Society for Computers in Psychology will be held at the Hyatt Regency Hotel in New Orleans on November 15, 1990. The meeting will include presentations, discussions, and times to preview software and hardware. All areas of psychology are featured, including research, education, clinical practice, and industrial applications.

For further information, contact Sarah Ransdell, Department of Psychology, University of Maine, Orono, ME 04469 (BITNET address RANSDSELL@MAINE) or C. Michael Levy, Department of Psychology, University of Florida, Gainesville, FL 32611 (BITNET address MLEVY@UFFSC).

A New Series Knowledge: Disciplinarity and Beyond Call for Proposals

The University Press of Virginia announces a new series. *Knowledge: Disciplinarity and Beyond* will feature books that examine, critically and imaginatively, the organization and production of knowledge—for instance, disciplinary histories, institutional topographies, and studies of counter- and extradisciplinary enterprises. The series is edited by Ellen Messer-Davidow (University of Minnesota), David R. Shumway (Carnegie-Mellon University), and David J. Sylvan (University of Minnesota).

Please send proposals (approximately 10 pages) for monographs and collections to The Editors, Disciplinarity Series, The University Press of Virginia, Box 3608, University Station, Charlottesville, VA 22903 (telephone, 804-924-3361).