Adapted Primary Literature for Teaching

How To Correct False Knowledge

Original research report: Butler, A. C., Fazio, L. K.,and Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. Psychonomic Bulletin and Review. 10.3758/s13423-011-0173-y

Adapting author: W. R. Klemm

Abstract

Authors begin by reminding us that people often have misconceptions that are well-learned and firmly believed. Surprisingly, previous research suggests that the more confident one is about a misconception, the easier it is for proper feedback to correct it. Here, they tested the hypothesis that this "hypercorrection effect" can be demonstrated after a one-week delay after a misconception was initially corrected. Subjects were given a general-knowledge quiz about science and for each question they rated how confident they were with their answer. Immediately afterwards, students received corrective feedback on those questions they missed. Half of the group re-answered the questions immediately and the other half did so a week later. Results revealed the hypercorrection effect in both cases, though the degree of error correction was less when the test was delayed a week. In the delayed test, when students missed a question, they sometimes gave the same wrong answer they did originally. If the correct answer had been forgotten, high-confidence answers were more likely than low-confidence answers to be repeated on the delayed test.

Introduction

The paper begins by listing common sources of misinformation such as novels, movies, and multiple-choice tests. As an example, they cite a study of recent Harvard graduates almost all of whom had erroneous explanations for the climate seasons, even though the graduates felt very confident their understanding was correct. So, the authors want to know why such gross errors persist in supposedly smart people.

The authors cite research showing that a common way to correct factual error is to provide feedback, but that such feedback is less likely to change memory if the error is strongly believed and easily remembered. They explain this in terms of "proactive interference," a mechanism in which well-learned information interferes with learning new related information.

Contrary to this effect, recent studies indicate that errors made with high confidence are more likely to be corrected by feedback than low-confidence answers. The authors hypothesized that these contradictory results might be explained by the lag time between events. They noted that previous studies had used a short interval (less than 5 minutes) between corrective feedback and re-testing. Interference theory predicts that the hypercorrection effects might be short-lived. Memory of corrected high-confidence answers might remain intact early on but later might fade, resulting in a return of wrong answers. They cite a paper that supports this notion.

However, they cite a conflicting study in which high-confidence answers were more correctable, irrespective of whether the final test was given right away or a week later. Two problems occurred

with this study. First, there were few high-confidence answers on the initial test. Second, the investigators did not check to see if errors on the final test were the same as those made initially.

Thus, the main goal of this present study was to see if error correction changes over time. Specifically, does correction of high-confidence errors persist over a long delay or does the memory of original errors tend to re-appear?

Methods

About 50 Duke University undergraduates participated, either for course credit or pay. The one independent variable was the retention interval, either 6 minutes or 1 week.

Students were given an initial 120 general-knowledge questions about science in short-answer format (such as, What is stored in a camel's hump? Answer: fat). Single words or short phrases sufficed as answers. Authors picked questions from pilot studies on the basis of questions that produced a big range in accuracy and confidence ratings.

Students were divided randomly into two groups, one that was to be re-tested in six minutes and the other one week later. Students first took the test, being required to give an answer even if they were not sure. At the same time, they rated their confidence about the answer on an ordinal scale of 1 to 7. Then students were given the correct answer for each question and told to study it for six seconds. Then, students were given a "filler task," visuo-spatial puzzles, for six minutes and then re-took the test, either at the six-minute point or one-week later. Again, students were forced to give an answer. After taking the test, students were immediately quizzed again to see if they remembered what answer they had given at original testing.

Grading was done independently by two graders who were blind to the experimental condition. Consistency of rating was high, as measured by the "Cohen's kappa" test. In the few cases where scorers differed, the senior author of the paper ruled on the correctness of the answer.

Results

Initial Test. Average correct scores were low (38%). This made it easier to assess the effect of error correction. No difference between test groups occurred, presumably because at this point no corrective feedback had been given. Table 1 shows the confidence ratings for answers that were correct or incorrect.

Table 1. The numbers of responses on the initial test as a function of confidence rating, response outcome (correct/incorrect), and retention interval group

Response Outcome	Retention Interval Group	Confidence Rating							Total
		1	2	3	4	5	6	7	
Correct	6 Min	46	30	39	139	119	155	554	1,082
	1 Week	50	41	52	130	157	187	564	1,181
	Total	96	71	91	269	276	342	1,118	2,263
Incorrect	6 Min	614	159	158	440	222	118	207	1,918
	1 Week	536	229	188	346	200	119	201	1,819
	Total	1,150	388	346	786	422	237	408	3,737
Grand total		1,246	459	437	1,055	698	579	1,526	6,000

Final Test. Subjects who were tested six minutes after corrective feedback had a higher percentage of correct answers than those who were tested one week later (data not presented).

Several "conditional analyses" were conducted. They first measured the relationship between response confidence and error correction (Fig. 1). The greater the confidence in the error, the more likely it was to be corrected (that is, the hypercorrection effect was demonstrated). This held for both retention intervals. Since the data were ordinal scale, they used a non-parametric statistical test (Goodman-Krusal correlation) to measure the consistency of confidence ratings on the initial test and the response outcomes on the final test (correct or incorrect). The average correlations were significantly different from zero. However, when data were compared across the two groups, correlations did not differ significantly from each other.

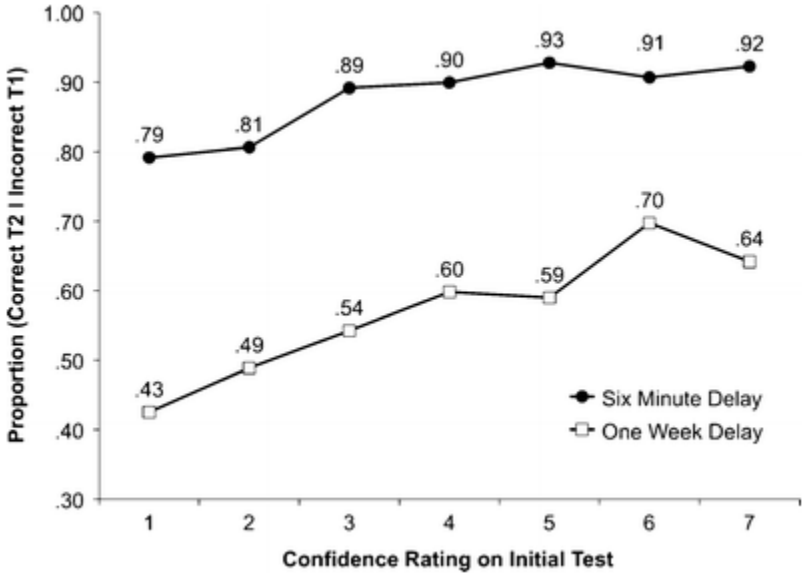

Fig. 1 Mean proportions of errors on the initial test that were corrected on the final test, as a function of confidence on the initial test and retention interval group

The proportion of errors corrected was lower in the group tested one week later than in the 6-minute group (56% vs. 86%). That is, over the one-week delay students forgot many of the correct responses. Did the original errors return? Yes, in the one-week group but not in the 6-minute group (Fig. 2). Moreover, the greater the original confidence in the error, the more likely it was to be repeated on the final test one-week later. The correlation was statistically significant. Equivalent correlation testing for the 6-minute group could not be made because there were not enough repeated errors.

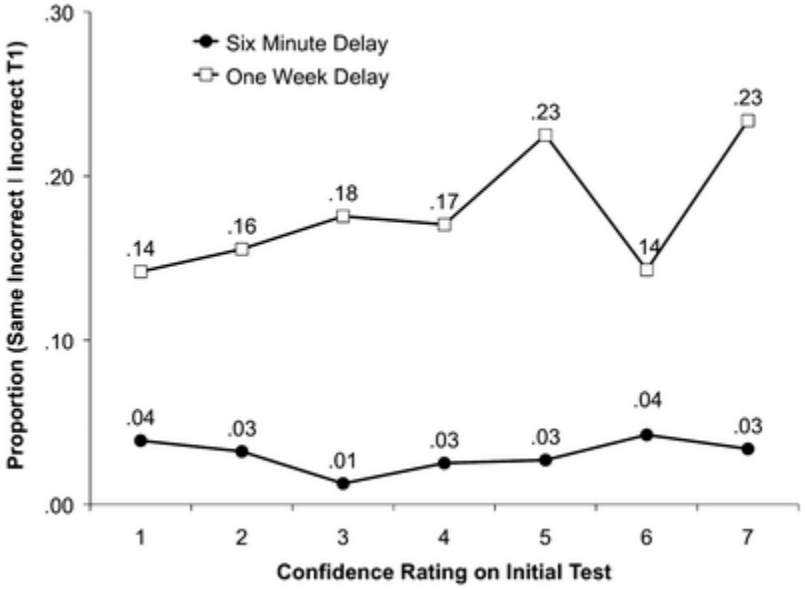

Fig. 2 Mean proportions of errors on the initial test for which the same error was produced on the final test, as a function of confidence on the initial test and retention interval group

Recall of Initial Test Responses. When subjects had answered the initial questions correctly, they remembered the correct answer 98% of the time on the final test. They were also good at remembering their errors on the first test (data not provided). Subjects in the six-minute group were more accurate than those in the one-week group (85% vs. 61%). Statistical correlation testing revealed that in both groups the higher the level of confidence in an erroneous answer the more likely it was to be made again on the final test.

For each subject, they measured the proportion of initial errors that were corrected on the final test as a function of whether or not the subject could recall the original error after the final test. Aggregating the data from both groups, subjects corrected 72% of the errors compared with 65% if they had not remembered it.

Discussion

The hypercorrection effect persisted over both short and long intervals. This confirms previous reports for early re-testing, and presents new evidence that hypercorrection can occur even for a longer retention interval of one week. Even so, the data showed that subjects forgot many of the correct responses after one week and began to re-produce the errors they made on the initial test. Moreover, conditional analysis showed that the greater the confidence in an error on the initial test, the more likely it was to be reproduced on the final test. Also, some evidence supported the notion that subjects were more likely to correct their answers if they remembered the initial error.

The apparent contradiction with literature on the proactive interference effect may be resolved by the finding that high-confidence answers are more likely to be reproduced if the correct answer is forgotten. Thus, a shift occurs gradually over time as correct answers are forgotten, making it more likely for high-confidence errors to return. In other words, proactive interference increases as a function of retention interval.

The authors put this finding in the context Bjork and Bjork's "theory of disuse." That is, "memory storage strength" (how well something is learned) and "retrieval strength" (how readily something can be recalled) are two different things. According to the theory, both storage and retrieval strength increase with each repeated exposure to the information. However, retrieval strength decreases over time because of exposure to other pieces of information.

The present results are thus interpreted to indicate that high-confidence errors have both high storage and retrieval strength. Feedback after an error should reduce retrieval strength of the error and increase the retrieval strength of the correct answer. Thus, when memory is tested soon after error correction, the correct answer should be retrieved. But as retrieval strength decays over time, high-confidence error is more likely to be retrieved because the storage strength decays over time. High-confidence errors tend to persist because they have more storage strength as a result of being rehearsed so often, compared to the brief presentation of the correction.

Since people seem to be quite good at recalling earlier errors, one might expect that remembering the error interferes with learning a correction. This is consistent with an earlier report that learning a correction is more likely when corrective feedback is given later when an error is more likely forgotten. But, the present study found a small but statistically significant advantage in error correction when subjects could remember the error they made on the initial test. Authors provide no good explanation for this discrepancy.

The implications for teachers deal with the common problem that student misconceptions tend to persist despite corrective instruction. Authors stress the inadequacy of just one corrective feedback instance, especially with high-confidence errors. One option, tested by others, is to refute misconception and practice retrieving correct responses. They cite a study in which a test was given, corrective feedback given for all answers, and then students were re-tested right away. This reinforcement of correction resulted in few mistakes on the re-test. The most effective procedure, advocated in another report, could be to provide repeated testing on the same material with corrective feedback given soon after each test.

In summary, correct misconceptions is neither quick nor easy. But the hypercorrection effect can be used to improve learning.
