Learning from one's own errors and those of others

Janet Metcalfe[1] · Judy Xu[1]

Abstract Three experiments investigated the effects of making errors oneself, as compared to just hearing the correct answer without error generation, hearing another person make an error, or being "on-the-hook," that is, possibly but not necessarily being the person who would be "called-on" to give a response. In all three experiments, generating either an error of commission or generating the correct response, oneself, out loud, compared to being a person who heard another's commission errors (or correct responses), was beneficial for later recall of the correct answer. Experiment 1 suggested that the decrement in recall from self- to other-generation could be partially offset by being "on-the-hook." However, this benefit was fragile and did not hold up either at a delay or when the presence of the other participants was downplayed. The beneficial effect of self-generation, both of correct responses and of errors of commission is consistent with reconsolidation theory. That theory holds that retrieval has a special status as a memory process that renders the retrieved traces *labile*. If the person was correct, reconsolidating the correct trace results in strengthening. If wrong, the malleability of the retrieved trace implied by reconsolidation theory makes it open to enhanced modification and correction. If the person was not the agent who retrieved, though, such as when someone else retrieves information, or when nothing is retrieved as is the case with omission errors (which we argue is truly how the term "unsuccessful retrieval" should be used), the benefit conferred by the special malleability entailed by the postulated reconsolidation process does not obtain.

✉ Janet Metcalfe
jm348@columbia.edu

1 Columbia University, 401B Schermerhorn Hall, New York, NY 10027, USA

There is a growing literature on the counterintuitive finding that when an individual generates an error followed by corrective feedback (rather than just receiving the correct answer without reference to possible errors) memory for the correct answer is enhanced (Kornell, Hays & Bjork, 2009). This finding is puzzling for classical interference theory. The memory for incorrect responses to a probe (a "C" response to a cue "A," in a situation where one is trying to learn "B" to the cue "A") should *interfere* with learning of the correct (B) response (Barnes & Underwood, 1959). If that erroneous response was strengthened by being self-generated (Bertsch, Pesta, Wiscott, & McDaniel, 2007), then the enhanced generation-related memorability of the erroneous response should make learning the correct answer even more difficult. There is a convergent body of literature (Engelkamp & Zimmer, 1997), on the "enactment" effect, showing that memory is better when an individual does an action him or herself, rather than just seeing or hearing it done by others. It follows that this additional strengthening of the erroneous response should make it even more interfering with memory for the correct answer. However, empirically, the presence of the error increases, rather than decreasing, correct responding. We consider the possibility that this puzzling benefit might be specific to self-generated retrieval.[1]

---

[1] Note that other kinds of models than the ones highlighted here—that do not emphasize the role of self-produced retrieval—such as search set models, e.g., Gilland & Shiffrin, 1984; mediational models, Carpenter, 2011, Pyc & Rawson, 2010, and prediction error models, e.g., Rumelhart & McClelland, 1986, also have been proposed as explanations for these effects, with varying degrees of success (see Grimaldi & Karpicke, 2012, Metcalfe, 2017, for discussion).

Although many experiments show that when a person generates an error as opposed to just being presented with the correct answer, recall of the correct answer is enhanced (Huelser & Metcalfe, 2012; Izawa, 1970; Kornell et al., 2009; Kornell, Klein, & Rawson, 2015, Slamecka & Fevreiski, 1983; Richland, Kornell, & Kao, 2009), we were unable to find any experiments showing whether errors committed by another person also produce enhanced performance. However, there is some indirect evidence. Grimaldi and Karpicke (2012) either allowed people to generate their own errors (for example, by guessing "WAVE" to the cue "tide," when the correct response was "beach") or they constrained them to produce the error, "WAVE," by giving them "ocean-WA_ _" and having them complete the second word, before they were presented with the correct pairing, which was "ocean-beach." When participants were forced to make an error (or, to a lesser extent, when they pre-learned the wrong response by being given it), memory for the correct answer was impaired. When they generated the error on their own, however, it was enhanced. Similarly, "blocked" tip-of-the-tongue states in which an individual generates a wrong response to a question (and they know that it is wrong), do not impair access to the correct target item (Kornell & Metcalfe, 2006), but when the experimenter gives people in a tip-of-the-tongue state external erroneous "blockers" (Smith & Blankenship, 1991), access to the correct item is impaired. Furthermore, the literature on externally presented fallacious suggestions (Loftus, Doyle, & Dysart, 2013) indicates that these inaccurate materials produce interference with memory for the correct item.

A theory that is gaining traction at the biological level and that places an emphasis upon retrieval—reconsolidation theory (Nader, Schafe & LeDoux, 2000; Lee, 2008)—may be relevant. The core idea is that when a response is retrieved, rather than necessarily being strengthened, its trace becomes *labile.* Within a limited time window following its exposure, it is out in the open and vulnerable. If the response was correct, the trace could be reconsolidated in its correct form and hence strengthened. If it was wrong, the theory postulates that it is open to change and correction. This enhanced malleability depends upon self-generated retrieval.

In the lab-based error-correction paradigm, the errors were, to our knowledge, always self-generated. However, in crowded classroom situations, much of the time the student would be exposed to other people's errors. It is not known, empirically, whether other-generated errors enhance or impair memory for the correct answers. If the latter—as we suspect—it also is not known whether it might be possible to induce self-generation effects by getting people to generate covertly. This paper addresses an issue that is both interesting for theories of learning and memory as well as educational practice: Does one have to be the author of an overtly produced error for it to produce beneficial memorial consequences? Or can one benefit from others' errors?

We hypothesized, consistent with reconsolidation theory, that memory for the correct answers would be better after self-generating errors compared with just listening to the correct answer. It is worth noting that omission errors and commission errors have a different status within the reconsolidation framework: generating a wrong response (a commission error) is retrieval; failing to respond (i.e., an omission error) is not. So, the enhanced malleability would only be expected with commission errors.

We also suspected—as is consistent with interference theory—that hearing errors generated by others might result in worse memory for the correct answers than would just hearing the correct answers without the interposition of an error. Keeping the classroom situation in mind, we endeavored to devise a manipulation to offset the likely downside of being the observer rather than the agent—the so-called "Hook" condition—in which we tried to induce everyone to generate the answer covertly.

Experiment 1

In these experiments, there was a "Call-on" condition, in which the participant who would answer the question was designated in advance of anyone hearing the question. The other participants could be passive. There was a condition in which people were on-the-"Hook." The question was posed generally, and only after doing so was the person who was designated to answer overtly specified. We expected that the people who were called on in the Hook condition would perform as well as those in the Call-on condition who were called on. We also thought that, because they would covertly generate the answer in anticipation of *perhaps* being called on, people who were not asked to answer might also do well in the Hook condition. Finally, there was a "Listen" condition in which the question and correct answer were simply read by the experimenter. This was the baseline. Within the Call-on and the Hook conditions, one person was the respondent, who answered aloud, and two people did not answer aloud.

Method

Participants Thirty-three Columbia University students (11 males and 22 females) participated for $10. Due to computer error, age of participants was lost; however, participants were restricted to be 18-40 years old. The experiments presented were approved by the Columbia institutional review board and were in accordance with the ethical standards of the Psychonomic Society.

Materials Ninety general information questions, each of which had a verified correct answer, were presented. Questions were randomized across each session of the experiment.

Design and procedure There were three conditions in the experiment: Call-on, Hook, and Listen, with 30 questions in each. The Listen condition served as a baseline. Within the Call-on and the Hook conditions, for any particular participant, one third of the questions were answered by the Self, and two thirds of the questions were answered by the Other. Three people were tested together in each group, and for every three successive questions, one of them was randomly assigned to each participant. When a question was assigned to a particular person it was designated as a "self" question for that person, and the other two questions given to the other participants were "other" questions for that person. Conditions were blocked such that 15 questions were given in each block (i.e., in Listen, Call-on or Hook), and two blocks, of each condition, were presented. The order of blocks was assigned by Latin squares over the entire experiment. Participants were told ahead of each block which condition—Call-on, Hook, or Listen—it would be. In three of the groups, because of no-shows, a confederate trained to mimic the responses that real participants provided took the role of the third participant.

The experiment was conducted on individual computers (in separate rooms) for each participant and for the experimenter, using the multiple-person Skype setup shown in Fig. 1. Each participant was assigned a letter A, B, or C affixed to their computer. Displays were set up such that the experimenter, who read aloud each question, appeared in the middle row along with the other participants, while the participant, him or herself, appeared at the bottom.

In the Call-on condition, the experimenter first indicated aloud whether participant—A, B, or C—would answer. Then, she read the question aloud; waited for an answer; then read the correct answer (which was provided to all questions). In the Hook condition the experimenter read the question aloud (without indicting who would answer), paused for 1-2 seconds, and only then stated which participant would answer. That participant then answered and the experimenter then gave the correct answer. In the Listen condition, the experimenter simply read the question aloud and then gave the correct answer.

After a 5-minute puzzle distractor, participants were individually tested on all 90 questions presented one at a time in written form in a different random order for each participant on the computer. Participants typed their responses into the computer.

Results

Where appropriate, Greenhouse-Geisser corrected degrees of freedom and *p*-values are reported. Degrees of freedom for analyses conditionalized on pretest performance sometimes differ because some participants may not have gotten any correct and/or made both commission and omission errors. Pretest performance across conditions in the three experiments is shown in Table 1.

We compared final recall of the correct feedback in the Hook and Call-on conditions. As is shown in the left panel of Fig. 2, there was an effect of Respondent, such that participants did better on Self than Other questions, $F(1,32)=6.97$, $p=.01$. There was no effect of Condition, $F(1,32) = 2.35$, $p = 0.14$. And, there was a trend towards a Condition x Respondent interaction, $F(1,32) = 3.66$, $p = 0.06$. None of the Self/Hook, Other/Hook, or Self/Call-on conditions were different from the Listen Condition ($t(32) = 0.49$, $p = 0.628$; $d = 0.09$, 95% confidence interval (CI) [−0.04, 0.07], $t(32) = 1.56$, $p = 0.128$; $d = 0.27$, 95% CI [−0.09, 0.01], and $t(32) = 0.77$, $p = 0.446$; $d = 0.13$, 95% CI [−0.07, 0.03], respectively. All three of these conditions revealed higher recall performance than the Other/Call-on condition ($t(32) = 2.73$, $p = 0.010$; $d = 0.47$, 95% CI [0.02, 0.16]; $t(32) = 2.69$, $p = 0.011$; $d = 0.47$, 95% CI [0.02, 0.11]; and $t(32) = 3.96$, $p < 0.001$; $d = 0.69$, 95% CI [0.04, 0.13], respectively). In addition, scores in the Other/Call-on condition—the only case in which, purportedly, the person did not generate but yet heard another's error—also were lower than scores in the Listen condition, $t(32) = 4.26$, $p = 0.0002$; $d = 0.74$, 95% CI [−0.05, 0.15]. These results suggest that there may have been interference from hearing non-self-generated errors.
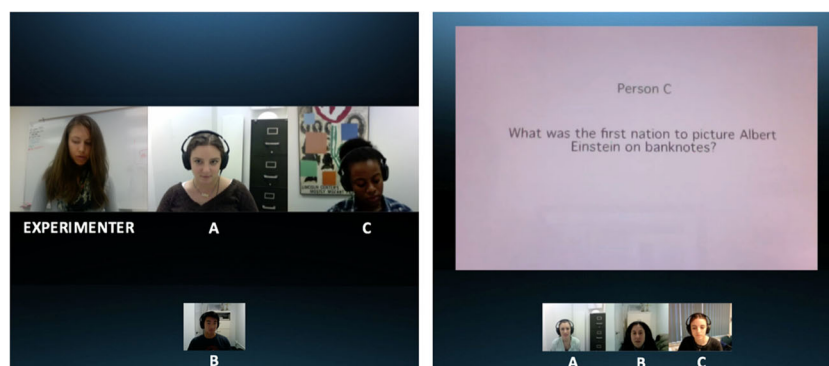

Fig. 1 Skype screenshots from participant B's point of view. Experiments 1 and 2 (left panel); Experiment 3 (right panel). Labeling of participants and experimenter were added for illustrative purposes

Table 1 Proportion of corrects responses, commission, and omission errors on the pretest across Hook and Call-on conditions for all three experiments. Proportions are derived from the responses that participants made within the *Self* Condition and are thus out of 10 items per condition per participant

	Correct		Commission		Omission	
	Call-on	Hook	Call-on	Hook	Call-on	Hook
Experiment 1	0.28 (0.18)	0.29 (0.21)	0.37 (0.21)	0.43 (0.23)	0.35 (0.24)	0.28 (0.22)
Experiment 2	0.24 (0.17)	0.25 (0.16)	0.23 (0.16)	0.24 (0.15)	0.52 (0.22)	0.52 (0.17)
Experiment 3	0.23 (0.14)	0.22 (0.15)	0.30 (0.20)	0.27 (0.16)	0.47 (0.23)	0.51 (0.20)

Standard deviations are in parentheses

For one of the error analyses, we looked at final recall depending upon whether the original response had been correct, a commission error, or an omission error (i.e., Pretest Response Type), treating these "as if" they had been manipulated, while, of course, they depended on the responses of the participants. We collapsed over Hook and Call-on conditions, because there was neither a main effect of Hook versus Call-on, nor was there a three-way interaction among Self/Other, Hook/Call-on, and Pretest Response Type. Additionally, without collapsing we would have lost about a third of our observations because of missing data in all cells. Figure 3 presents final recall of the correct answers for self- compared with other-generated items depending on the Pretest Response. There was a main effect of Respondent, such that participants performed better on questions they had been asked on the pretest (self questions) than those answered by other participants, $F(1,29) = 7.01$, $p = 0.01$. There was an effect of Pretest Response Type, $F(1.59,46.19) = 74.07$, $p < 0.0001$, such that, as often has been shown in experiments reporting retesting of corrects, commission errors and omission errors, performance on corrects was better than on commission errors, $t(32) = 7.01$, $p < 0.0001$, $d = 1.22$, 95% CI [0.12, 0.22], and performance was better on Commission errors than on Omission errors,

$t(32) = 8.14$, $p < 0.0001$, $d = 1.42$, 95% CI [0.15, 0.24]. The interaction between Respondent and Pretest Response Type was not significant, $F(1.53,44.37) = 2.21$, $p = 0.13$.

Finally, the proportion of intrusion errors in the final test, broken into those that were the same as the commission errors from the pretest or that were different from the pretest commission errors, are presented in Table 2. There were few intrusions that were the same as the pretest errors, which might indicate the persistence of those earlier errors. There was no effect of either Condition or Respondent.

Experiment 2

Experiment 2 was identical to Experiment 1 except that participants were tested after a 1-week delay. Thirty-four new participants (12 males and 22 females, $M = 21.5$ years, $SD = 3.13$) participated (and also, on Day 1, completed a questionnaire about anxiety and attention, the results of which are available upon request).

Results

As is shown in the center panel of Fig. 2, there was a main effect of Respondent, in which Self resulted in better recall than Other, $F(1,33) = 9.48$, $p = 0.004$. There was no main effect of Condition, $F(1,33) = 3.64$, $p = 0.07$, nor was there a Respondent by Condition interaction, $F(1,33) = 0.02$, $p = 0.90$. Recall in both the Self/Hook and Self/Call-on conditions was significantly higher than in the Listen condition, $t(33) = 3.33$, $p = 0.002$, $d = 0.57$, 95% CI [0.04, 0.16] and $t(33) = 2.27$, $p = 0.030$, $d = 0.39$, 95% CI [0.01, 0.12], respectively. Neither Other/Hook nor Other/Call-on differed from the Listen condition, $t(33) = 1.75$, $p = 0.09$, $d = 0.30$, 95% CI [−0.01, 0.08], and $t(33) = 0.15$, $p = 0.883$, $d = 0.03$, 95% CI [−0.04, 0.04], respectively.

In investigating the different kinds of initial responses, there was an effect of Respondent, such that recall was better on Self than Other questions, $F(1,31) = 20.44$, $p <$
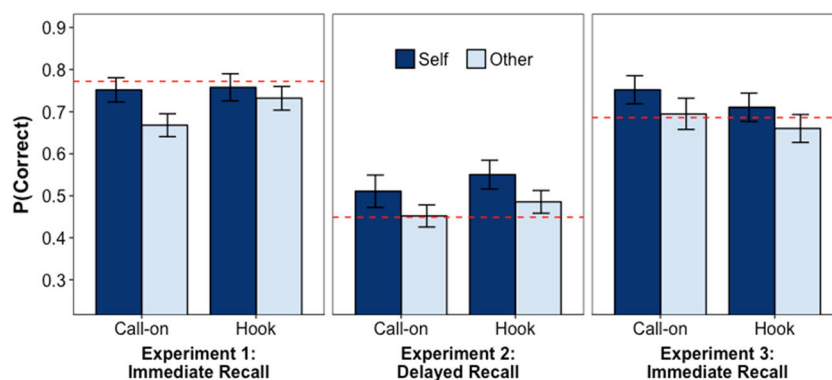

Fig. 2 Proportion correctly recalled on the final test for questions initially answered by the self and by another. The red dashed line represents the mean performance for the Listen condition
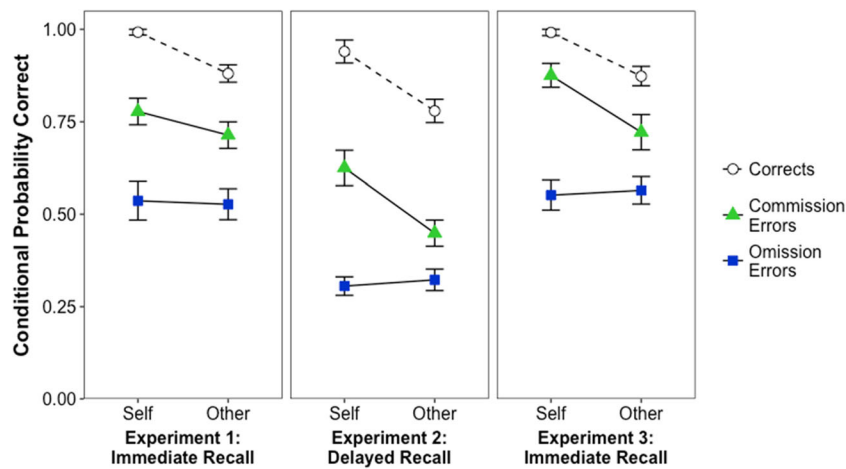
Fig. 3 Conditional probability of correct final recall on items that had originally been Correct, errors of Commission, and errors of Omission, as a function of whether the original response was provided by the Self or the Other

0.0001. There was an effect of Pretest Response Type, $F(1.79, 55.41) = 253.12$, $p < 0.0001$, such that performance was better on Corrects than Commission errors, $t(33) = 11.52$, $p < 0.0001$, $d = 1.98$, 95% CI [0.27, 0.38], and on Commission errors than Omission errors, $t(33) = 7.84$, $p < 0.0001$, $d = 1.34$, 95% CI [0.13, 0.22]. Most importantly, there was an interaction between Respondent and Pretest Response Type, $F(1.51, 46.69) = 8.09$, $p = 0.002$. As shown in the center panel of Fig. 3, while there was a difference between Self and Other on both the Corrects and the Commission errors ($t(33) = 5.63$, $p < 0.0001$, $d = 0.96$, 95% CI [0.10, 0.22], and $t(31) = 3.09$, $p = 0.004$, $d = 0.55$, 95% CI [0.06, 0.28]), there was no memorial benefit for Self compared with Other on errors of Omission, $t(33) = 0.62$, $p = 0.542$, $d = 0.11$, 95% CI [−0.07, 0.04].

Finally, the intrusion errors in the final test are given in Table 2. Among those intrusions that were repeats of the errors

in the pretest, there was neither an effect of Condition nor Respondent. There was no interaction.

Discussion

The memory advantage to having one's own errors corrected as opposed to hearing another person's errors corrected, while found in both of the first two experiments, was especially salient in Experiment 2. When recall was immediate, being On-the-Hook marginally offset the effect of being the person who answered the question, but it did not do so when recall was delayed. Furthermore, in keeping with reconsolidation theory, only those responses that were actually generated (i.e., correct responses and commission errors) showed a Self/Other effect. Omission errors showed no benefit for the person who was supposed to, but failed to, retrieve.

Table 2 Proportion of responses on the final test that were intrusion errors that were either the same as the original error in the pretest or different from the pretest error

		Same as original error		Different from original error	
		Call-on	Hook	Call-on	Hook
Experiment 1	Self	0.01 (0.02)	0.02 (0.04)	0.17 (0.12)	0.12 (0.16)
	Other	0.02 (0.04)	0.01 (0.02)	0.19 (0.12)	0.15 (0.13)
Experiment 2	Self	0.01 (0.04)	0.02 (0.05)	0.25 (0.23)	0.24 (0.18)
	Other	0.03 (0.05)	0.02 (0.04)	0.26 (0.19)	0.25 (0.18)
Experiment 3	Self	0.003 (0.02)	0	0.10 (0.15)	0.13 (0.15)
	Other	0.01 (0.04)	0.01 (0.02)	0.11 (0.12)	0.09 (0.09)

Experiment 3

Experiment 3 served as both a replication with modifications of Experiment 1, and also investigated whether the highly salient social cues – the strong visual presence of the experimenter and other participants – influenced the efficacy of being 'on-the-hook.' Accordingly, instead of showing an experimenter reading the question, we had the question (and answer) presented in written form (and read aloud). The experimenter was not shown. In addition, the size of the images of the other participants was decreased (see Fig. 1). We were interested in developing this method of presentation, in part, as an intermediate step in allowing us to fully automatize the multi-person Skype tutorial procedure.

There were 29 participants (12 males, 15 females, 2 skipped the question; $M = 25.41$ years, $SD = 6.18$). Aside from the method of display, this experiment was the same as Experiment 1.

Results

The right panel of Fig. 2 shows an effect of Respondent, in which Self resulted in better recall than Other, $F(1,28) = 5.51$, $p = 0.03$. There was no effect of Condition, $F(1,28) = 2.33$, $p = 0.14$; nor was there a Respondent by Condition interaction, $F(1,28) = 0.02$, $p = 0.89$. Recall in the Self/Call-on condition was significantly better than in the Listen condition, $t(28) = 2.09$, $p = 0.046$, $d = 0.39$, 95% CI [0.001, 0.13]. Recall in the Self/Hook, Other/Hook, and Other/Call-on conditions were not different from the Listen condition, $t(28) = 0.85$, $p = 0.400$, $d = 0.16$, 95% CI [−0.03, 0.08], $t(28) = 1.16$, $p = 0.255$, $d = 0.22$, 95% CI [−0.07, 0.02], and $t(28) = 0.36$, $p = 0.720$, $d = 0.07$, 95% CI [−0.04, 0.06], respectively.

As shown in the right panel of Fig. 3, there a difference in proportion recalled as a function of the Pretest Response Type, $F(1.99,51.74) = 101.58$, $p < 0.0001$. Participants performed significantly better on Corrects than Commission errors, $t(28) = 4.43$, $p = 0.0001$, $d = 0.82$, 95% CI [0.07, 0.18], and better on Commission than Omission errors, $t(28) = 8.85$, $p < 0.0001$, $d = 1.64$, 95% CI [0.16, 0.26]. There was an effect of Respondent, $F(1,26) = 8.15$, $p = 0.008$, such that Self was better than Other. Importantly, there was a significant interaction between Respondent and Pretest Response Type, $F(1.97,51.31) = 7.97$, $p = 0.001$. Final memory performance was better for Self than for Other on items that initially been Correct, $t(28) = 4.11$, $p = 0.0003$, $d = 0.76$, 95% CI [0.06, 0.18], or Commission errors, $t(28) = 3.39$, $p = 0.002$, $d = 0.65$, 95% CI [0.06, 0.26]. However, no memorial benefit was found for Self over Other on Omissions, $t(28) = 0.32$, $p = 0.749$, $d = 0.06$, 95% CI [−0.09, 0.07].

Finally, the intrusion errors are shown in Table 2. Among those intrusions that were repeats of the errors in the pretest, there was no effect of Condition. In this experiment, however, there was an effect of Respondent, $F(1,26) = 5.66$, $p = 0.02$, showing that more perseverative intrusions were made when the error had been generated in the pretest by the Other rather than the Self.

General discussion

In all three experiments, being the agent who him/herself retrieved either the correct answer or who committed an error, facilitated memory for the correct answer. When the participant merely witnessed someone else make a mistake, or provide a correct response (without covertly generating, as in the Hook condition in Experiment 1) recall performance was lower.

These results are consistent with reconsolidation theory, whereby memory traces are open to modification upon retrieval. The results are not consistent with a simple version of interference theory, or with other theories that posit that self-generation makes the representation that is so retrieved stronger and more memorable. Unlike most theories, reconsolidation theory places a distinct emphasis on the need for retrieval of a response to render the exposed memory trace *vulnerable to modification*. However, the modification is not necessarily strengthening. It can be correction, alteration, or even erasure.

Commission errors and correct responses—both cases in which a response is retrieved—both showed the Self/Other effect, in the last two experiments, although not in Experiment 1. In that experiment, being On-the-Hook appears to have been partially effective in getting people to covertly retrieve even when they were not overtly asked to do so, diluting the retrieval benefit of being the retrieving Self. Omission errors, also, did not show a Self/Other effect in any of the three experiments. According to reconsolidation theory, with omission errors, no trace would be retrieved to enter the malleable reconsolidation state. It is consistent with the theory that no special advantage accrues to being the person who was called upon but who failed to retrieve. It is not enough to be the person in the spotlight who is asked to make a response if no response is evoked.

The tenants of interference theory—being exposed to incorrect answers hurts memory for the correct answers—seem to apply, if at all, only when the trace was not retrieved by the self. Observing errors being made, in Experiment 1, did interfere somewhat with recall of the correct answers (compared with the listen condition). However, in the other experiments, while being the agent who retrieved the responses aloud (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) helped later recall, being the person who merely observed errors generated by other people did not hurt, relative to just hearing the correct response alone.

In the Hook condition, we tried to get people to retrieve covertly, and thereby obtain a generation effect even when the person was not called upon. We had limited success. There was a small generation-like benefit to being on-the-hook in Experiment 1, but the effect was fragile. It only occurred when the test was immediate, when the experimenter who was posing the questions was visually present, and when the other participants were saliently displayed.

These results have implications for practical implementations designed to help learning. They suggest that, to the extent possible, it may be advantageous to individualize testing and feedback. Listening to other people's errors did not advantage the person who had not made the error, and sometimes hurt. However, in a typical classroom situation it may only rarely be the case that the individual's own particular errors are those being addressed. We had thought that putting students on-the-hook might overcome the disadvantage to not being the person generating the error. The effect was weak at best. The results presented indicate that the full benefit of learning from errors appears to depend on the error being one's own error that is overtly retrieved and corrected. These results are consistent with reconsolidation theory, which places a special emphasis on the process of overt retrieval and renders memory traces open to alteration. This theory shows promise as an explanation of human error correction and may prove to be useful in helping to develop methods to optimize students' learning.

References

Barnes, J. M., & Underwood, B. J. (1959). Fate of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97–105.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35,* 201–210.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica, 96,* 42–60.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40,* 505–513.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40,* 514–527.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83,* 340–344.

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.

Kornell, N., Klein, P. J., & Rawson, K. A. (2015). Retrieval attempts enhance learning but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 283–294.

Kornell, N., & Metcalfe, J. (2006). Blockers do not block recall in tip-of-the-tongue states. *Metacognition and Learning, 1,* 248–261.

Lee, J. L. C. (2008). Memory reconsolidation mediates the strengthening of memories by additional learning. *Nature Neuroscience, 11,* 1264–1266.

Loftus, E. F., Doyle, J. M., & Dysart, J. E. (2013). *Eyewitness testimony: Civil and Criminal*. New Providence: LexisNexis.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 671–685.

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology, 68.* Advance publication available online.

Nader, K., Schafe, G., & LeDoux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature, 406,* 722–726.

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

Richland, L. E., Kao, L. S., & Kornell, N. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15,* 243–257.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. I). Cambridge: MIT Press.

Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior, 22,* 153–163.

Smith, S. M., & Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *American Journal of Psychology, 104,* 61–87.