Neuropsychologia 138 (2020) 107296

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Learning from errors is attributable to episodic recollection rather than semantic mediation

Janet Metcalfe *, Barbie J. Huelser

Department of Psychology, Columbia University, 1190 Amsterdam Avenue, New York, NY, 10027, USA

A R T I C L E I N F O

Keywords:
Learning from errors
Episodic memory
Amnesia
Test effects
Semantic mediation
Recursive reminding

A B S T R A C T

Many recent studies have shown that memory for correct answers is enhanced when an error is committed and then corrected, as compared to when the correct answer is provided without intervening error commission. The fact that the kind of errors that produced such a benefit, in past research, were those that were semantically related to the correct answer suggested that the effect may occur because the error provides a semantic stepping stone to the correct answer: the Semantic Mediation hypothesis. This hypothesis seems at odds with the finding that amnesicsgenerate answers, again including those studied by Tulving and his colleagues–who purportedly have spared semantic/implicit memory–experience enormous difficulties when they commit errors. Accordingly, the present experiments investigated whether the error-generation benefit seen in typicals was attributable Semantic Mediation or to Episodic Recollection. In Experiment 1, we used polysemous materials to create Congruent (e.g., wrist-palm) and Incongruent (e.g., tree-palm) cues for target words (e.g., HAND). In the Congruent condition, participants generated errors that were semantically related to the target (e.g., finger), and which could have provided a semantic mediator. In the Incongruent condition they generated errors that were *unrelated* to the target (e.g., coconut), and which, therefore, should not have provided a semantic mediator. The Congruent and Incongruent conditions both produced an error-generation benefit–contradicting the Semantic Mediation hypothesis. Experiment 2 showed that the error-generation benefit only occurred when the original error was also recollected on the final memory test. Indeed, in the Incongruent condition, when the error was not, itself, recalled, error generation resulted in a deficit in memory for the correct response. These results point to episodic/explicit, rather than semantic/implicit memory, as the locus of the ‘learning from errors’ benefits.

1. Introduction

In typical participants, generating an error prior to being provided with the correct answer results in enhanced memory for the correct answer, as compared to simply studying the correct information with no intervening error generation (Finn and Metcalfe, 2010; Grimaldi and Karpicke, 2012; Hays et al., 2013; Huelser and Metcalfe, 2012; Izawa, 1967, 1970; Karpicke and Roediger, 2007; Kornell et al., 2009; Metcalfe and Miele, 2014; Pastötter et al., 2011; Richland et al., 2009; Slamecka and Fevreiski, 1983). This finding has wide-ranging implications for understanding learning and memory and also for educational practice. But it is puzzling in light of findings by Tulving and his many colleagues concerning the deleterious effect of errors on people with amnesia, especially insofar as the error-generation benefit in typical participants is postulated to be attributable to semantic memory, a system that is thought by many to be spared in amnesia.

In the standard ‘error generation’ paradigm of Kornell et al. (2009), a cue is presented and the participant guesses what they think the correct answer will be. After generating an error, the participant is shown the correct answer. In the error-free study only condition, both the cue and target are displayed simultaneously for study, and there is no opportunity to commit an error. The highly robust finding is that participants are more likely to remember the correct target response on a later memory test if they first generated an error, and then had it corrected, as compared to when no error was generated and they only studied the target along with the cue. Among typical participants, error generation improves memory for the correction.

One explanation for the benefit of error generation is that the error may, itself, serve as a semantic mediator, making it more likely that the participant will get to the target. The Semantic Mediation hypothesis was first proposed to explain why testing improves memory more than does merely studying, but it can easily be extended to explain why

* Corresponding author.
E-mail address: jm348@columbia.edu (J. Metcalfe).

https://doi.org/10.1016/j.neuropsychologia.2019.107296
Received 3 June 2019; Received in revised form 8 October 2019; Accepted 3 December 2019
Available online 4 December 2019
0028-3932/© 2019 Published by Elsevier Ltd.

generating errors enhances memory. Pyc and Rawson (2010), Carpenter (2011), Carpenter and Yeung (2017), and Hausman and Rhodes (2018) have all proposed that testing is more beneficial to later recall than merely studying, because, in the testing condition, while people are in the process of trying to retrieve correct answers they produce semantic mediators that later are helpful in getting the learner to the correct answer. As noted by Carpenter (2011, p, 2): “one type of information activated during retrieval might be semantic in nature. For example, the cue Basket:___ may activate concepts such as “Eggs,” “Flour,” or “Wicker.” If the correct target is eventually retrieved (Bread), there is now a structure of mediating semantic information that links Basket to Bread.” This mediation notion is consistent with several semantic network models: the closer two items are in semantic space (i.e. the more related they are), the more activation one item will receive as a result of activation of the other item (Anderson, 1983; Collins and Quillian, 1972; Collins and Loftus, 1975; Dunlosky et al., 2005; Neely, 1976), and the better will be recall (though see Lehman and Karpicke, 2016). The semantic mediation explanation implies that to be effective, the error needs to be semantically related to, or close to, the target, so that it will later put the learner in close semantic proximity to the answer.

The same line of reasoning can easily be extended to the case in which error generation is followed by correct target item feedback. The main difference is that the experimenter, rather than the participant, provides the correct answer. By this view, the error –so long as it is close in semantic space to the target—can provide a route to that target, enhancing its recall. The error-generation benefit should occur only with semantically related errors: unrelated errors are too far away, in semantic space, to provide the proximal spreading activation needed to enhance memory for the target item. Thus, when a person is presented with a prompt such as ‘wrist- _?_‘, and produces an error such as ‘finger,’ this response should activate concepts that are related to ‘finger’ and ‘wrist,’ including the concept ‘hand.’ If ‘hand’ is the to-be-remembered target, the semantic activation due to the error ‘finger’ should help (and see, Ratcliff and McKoon, 1994; Carpenter, 2011; Pyc and Rawson, 2010). In contrast, if an error were unrelated to the target (say it is ‘coconut’) it will not promote activation anywhere near the semantic vicinity of the target. Insofar as it does not, it should not enhance memory for the target. Indeed, it is even possible that the activation from an error in a semantic domain far removed from the correct target item could produce interference that would impair recall.

Consistent with the semantic mediation hypothesis, past research has shown that a self-generated error that is unrelated to the target results in memory performance that is no better than that in the error-free condition (Grimaldi and Karpicke, 2012). Similarly, when Huelser and Metcalfe (2012) used unrelated cue-target pairs which resulted in erroneous unrelated guessing, memory for the target was the same as in a study-only condition, even though, in the same experiment, semantically related errors produced a benefit. Furthermore, the more semantically similar the error is to the correct answer, the more likely the individual is to give the correct target at final test (Slamecka and Fevreiski, 1983).

In addition, it has been shown that high confidence errors, followed by corrective feedback, result in better memory for the corrective feedback than do low confidence errors (Butterfield and Mangels, 2003; Butterfield and Metcalfe, 2001, 2006; Cyr and Anderson, 2013; Eich et al., 2013, Metcalfe et al., 2015, Fazio and Marsh, 2009, 2010; Iwaki et al., 2013; Kang et al., 2011; Kulhavy et al., 1976; Metcalfe et al., 2012; Metcalfe and Eich, 2019; Metcalfe and Finn, 2012; Metcalfe and Miele, 2014; Sitzman et al., 2014; Sitzman et al., 2015; van Loon et al, 2015). But high-confidence errors are also more semantically related to the targets than are low-confidence errors (e.g., Eich, et al., 2013; Metcalfe and Finn, 2011; Sitzman et al., 2015). Sitzman et al., 2015, have argued that this semantic information content differential is responsible for the effect. When latent semantic analyses (LSA, Landauer and Dumais, 1997) have been conducted to assess the semantic nearness to the target of the high versus the low confidence errors that participants actually produced, the consistent finding is that high confidence errors are closer in semantic space to the answers than are low confidence errors. Such results bolster the Semantic Mediation hypothesis.

Even so, the notion that the benefit seen from the generation of errors in typical participants is attributable to semantic memory mediation sits awkwardly with the findings that amnesics–who are thought to have intact semantic memory but not episodic memory (Tulving, 1983, 2005; Tulving et al., 1991)—do not similarly benefit. These patients are hard hit by errors (Baddeley and Wilson, 1994; Hamann and Squire, 1995; Hayman et al., 1993; Rosenbaum et al., 2005). For example, Baddeley and Wilson (1994) administered to amnesic patients a stem-completion task in which they were either asked to first guess the correct answer (and were wrong—the errorful condition) followed by the presentation of the correct answer, or they were simply presented with the correct answer to the stem without any intervening error generation (the errorless condition). Control participants were on ceiling in both conditions. But the amnesic patients, despite performing well in the errorless condition, were devastated by the initial production of errors: performance was about 40% worse than in the errorless condition. Furthermore, effective remedial techniques such as the method of vanishing cues (Glisky and Schacter, 1988; Glisky et al., 1986a, 1986b; and see Hamann and Squire, 1995) have assiduously focused on creating situations in which the patient never generates an error.

Tulving, Hayman and McDonald (1990, p. 611), in describing the effects of testing on amnesic patient, KC’s, memory, have noted that: “In normal subjects, on-line tests of associations are known to have large facilitative effects on subsequent retention … Similar facilitative effects have been obtained with older adults whose retention of material is impaired in comparison with younger adults … But in KC and possibly in other amnesics on-line tests have the opposite effect, probably because they engender incorrect responses which then compete with the correct responses.”

To illustrate the effect of errors on amnesic patient KC, in a situation that is similar to Kornell et al.’s (2009) error generation paradigm, consider the study conducted by Hayman et al. (1993). Cue-riddles such as “An underpaid textile worker ______," “Performs a daily massage ______," or “A talkative featherbrain_________" were presented to K.C. in a pretest to a multi-session learning experiment, in which the goal was to learn the correct answer to each riddle (e.g., silkworm, toothbrush, and parakeet, respectively). Questions on which KC produced the correct answer on this pretest were not included in the study (much as Kornell et al., 2009, had omitted all correct guesses from further consideration). KC frequently produced errors on the pretest. The riddles were separated into those for which KC had produced an error on the pretest, and those for which he had not generated an error—he had said nothing. He was then provided with learning opportunities for the correct responses, over many sessions, in one of two ways—either by repeated testing (in which he tried to generate answers, again, often, producing errors) followed by being given the correct answer (the Test-Study condition), or by simply studying the correct answers along with the riddle (the Study-Only condition). The former condition is, of course, similar to Kornell et al.’s (2009) Error Generation condition, which, in typicals, produces a benefit in memory. The latter is similar to Kornell et al.’s (2009) Study-Only condition. In contrast to the results with typicals, however, not only was KC’s learning not enhanced by error generation, it was severely impaired. Riddles on which he had given an erroneous response on the pretest showed greatly impaired recall of the correct answer. In addition, memory for the correct answers in the Test-Study condition, in which errors were generated during the course of attempted learning, was also impaired. The greatest impairment was found when both occurred. The difference in KC’s memory–attributable to error generation–was in the opposite direction to that of typicals. And, it was severe.

In a similar vein, Schacter et al. (1986), tested K.C. on the Piagetian A not B task—a proactive interference test that can easily be construed as an error correction task. In their experiment, an object was first hidden at location A and, in full view, it was moved to location B—the new correct location. KC persisted in looking at location A. This is easy to understand as a lack of updating–an impairment in error correction.

Hamann and Squire (1995) tested 9 amnesic patients of varying etiologies on a similar paradigm. They used triads such as "medicine cured"-"Hiccup," such that amnesics and controls had both a Study Only condition, in which all three words were shown, and a Test condition, in which the participants generated errors. Consistent with the earlier described results, the amnesics were worse in their final recall of the answers in the Test condition than in the Study Only condition, whereas the controls were better in the Test condition. This pattern did not appear to be a function of the low overall level of recall insofar as it did not reverse for control patents even when they were tested at a long delay that reduced their overall recall level to that of the amnesics.

It is difficult to reconcile the Semantic Mediation hypothesis with the results shown by KC and other amnesics, in whom the memory deficit is thought to be selective to episodic/explicit memory while, at the same time semantic memory is purported to be spared. It is, of course, possible that the construct of semantic memory that is referred to as the basis of the semantic mediation hypothesis is different from the construct of semantic memory indicated by seems at odds with the findinging and colleagues with respect to amnesic patients (see, Rosenbaum et al., 2005). The kinds of semantic tasks spared in amnesic patients usually involve knowledge questions and/or priming effects rather than associative mediation tasks. . But if they are the same construct and the error correction benefit seen in typical participants is attributable to semantic memory (c.f., Lehman and Karpicke, 2016), it is puzzling as to why amnesics show impairment rather than enhancement.

Of course, Semantic Mediation is not the only explanation that can be forwarded to explain the error generation benefit (see Metcalfe, 2017, for review). We will return to alternative hypotheses in a later section of this article. But for now we allude only to the possibility that the error-generation benefit seen in typical subjects but reversed in amnesics, might—under some theoretical frameworks–be attributable, not to semantic/implicit, but rather to episodic/explicit memory.

The aim of the first experiment was to test the Semantic Mediation hypothesis in typical participants. To do so, we used materials that led to two different types of errors: semantically related errors with a close semantic link between the error and target (the Congruent condition), or unrelated errors that lead to the incorrect semantic domain (the Incongruent condition). According to the Semantic Mediation hypothesis a benefit from error generation should occur only in the former but not in the latter condition.

2. Experiment 1

2.1. Method

Forty Columbia University undergraduate students (Mean Age = 22.25 (*SD* = 7.3), 62.5% Female) were instructed to learn 60 word-triplets, where the second word of each triplet was polysemous (having more than one meaning, such as PALM). These were either created from the Nelson et al. (1998) norms or generated by the second author. The materials are given in Appendix A. Half of the triplets presented were congruent: all three words were tightly associated in the same semantic space (wrist – PALM – hand). The other half of the triplets were incongruent: one meaning was inconsistent with the other two, such that the first word of the triplet was of the alternate meaning of the polysemous word (*tree* – PALM– hand). Participants studied these triplets in two different learning conditions. For half of the trials, participants viewed the triplet on the screen (e.g. wrist – palm – hand) for 12 s [Study-only condition]. For the other half of the trials [Error-Generation condition], participants were given 7 s to generate a response to the double-word cue (e.g., wrist – palm _?_), and then the correct triplet (wrist – palm – hand) was displayed for 5 s.

Both Study-Only and Error-Generation trials were 12 s to control for total trial time. As a result, the correct answer was shown for 7 s longer in the Study-Only conditions than in the Error-Generation conditions, which might be seen as a considerable ad. In summary, this was a 2 (Materials: Congruent, Incongruent) x 2 (Learning Mode: Study Only, Error Generation) within-participants design, creating four unique conditions: [Congruent: Error Generation], [Congruent: Study Only], [Incongruent: Error Generation], and [Incongruent: Study Only]. Materials and Learning Mode were counterbalanced between participants. Following the learning phase, there was a short (6 min) visuo-spatial filler before the cued recall test of all 60 triplets (presented in a random order). Both of the original cue words were presented at test. Following the recall test participants were asked to provide a retrospective estimate of their own performance in each of the 4 conditions.

2.2. Results

Participants produced errors in most of the Error-Generation trials, with slightly more errors for Incongruent materials ($M = 0.95$, $SD = 0.06$) than for Congruent materials ($M = 0.91$, $SD = 0.07$), $t(39) = 2.54$, $SE = 0.02$, $p < .05$. Trials in which participants guessed the target correctly were excluded from all analyses. An analysis was conducted to check that the errors participants produced on congruent items were related to the target, while for incongruent items, they were not semantically related. For example, if the to-be-remembered target was 'HAND', 'finger' might be the error generated for a congruent item, (wrist – palm _?_). Alternatively, an incongruent triplet (tree–palm-_?_) might lead to the error 'coconut.' Although a difference in relatedness was predicted from priming results using similar materials (Marcel, 1980), we checked the degree of relatedness of the errors generated in the two conditions by using Latent Semantic Analysis (LSA) (Landauer et al., 1998; Landauer and Dumais, 1997). This analysis confirmed that generated errors for the incongruent items ($M = 0.14$, $SD = 0.05$) were less related to the target than were the errors generated by the congruent items ($M = 0.18$, $SD = 0.05$), $t\,(39) = 3.07$, $SE = 0.01$, $p < .05$).

The main result of interest—memory for the correct answers on the final cued recall test–was analyzed using a 2 (Materials: Congruent, Incongruent) x 2 (Learning Mode: Study Only, Error Generation) repeated measures ANOVA. As is shown in Fig. 1, in contrast to the Semantic Mediation hypothesis, which predicted that the benefit of error generation would apply only to the congruent errors, there was an error generation benefit for *both* congruent *and incongruent* materials such that generating an error during learning led to improved memory for the target ($M = 0.65$, $SD = 0.21$) over simply studying the triplets, ($M = 0.55$, $SD = 0.18$), $F\,(1, 39) = 16.43$, $MSE = 0.03$, $p < .001$, $\eta p^2 = 0.30$. Additionally, Congruent materials led to higher rates of correct

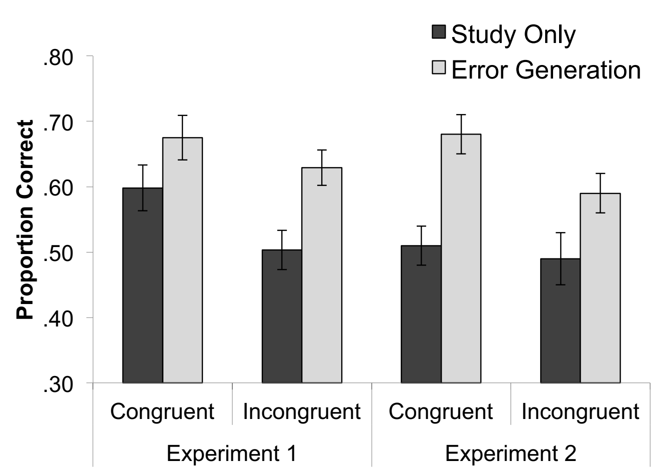

Fig. 1. Proportion correct on final recall test, for Experiments 1 and 2. Error bars indicate SEM.

recall on the final test ($M = 0.64$, $SD = 0.21$) over Incongruent materials ($M = 0.57$, $SD = 0.19$), $F(1, 39) = 12.70$, $MSE = 0.02$, $p < .01$, $\eta p^2 = 0.25$. The benefit of error generation over just studying was significant in the congruent condition *(0.68 versus 0.60,* $t(39) = 2.64$, $SE = 0.03$ $p < .02$), as expected from past research, but it was also significantly beneficial in the incongruent condition (0.63 versus 0.50, $t(39) = 3.97$, $SE = 0.03$ $p < .001$). The interaction between Materials and Learning Mode was not significant.

Furthermore, as is shown in Fig. 2, people's metacognitions concerning their own performance were reversed with regard to their own actual performance, with their estimations indicating that they thought they had recalled more in the Study Only condition than in the Error Generation condition. This result replicates our earlier findings (Huelser and Metcalfe, 2012) showing that people have no insight into the enhancing effects of error generation even after having just experienced them.

2.3. Discussion

The results of Experiment 1 provide disconfirming evidence concerning the Semantic Mediation hypothesis as the reason that error generation benefits correct recall. If Semantic Mediation had been the key to the error correction benefit at time of retrieval then people should have produced superior recall when the error, itself, had been closer in semantic space (in the congruent condition) than when it was further away (in the incongruent condition). If anything, the effect was in the wrong direction.

The findings of Experiment 1, in concert with the findings from amnesics, suggest the possibility that episodic rather than semantic memory may be implicated in error correction. Baddeley and Wilson (1994) noted this possibility specifically: "one of the crucial features of explicit or episodic memory is the capacity it gives to escape errors" (p. 54). Interestingly, while several theories propose an alternative to semantic mediation for the beneficial effect of generating errors (see Metcalfe, 2017, for review), perhaps the most relevant are the 'recursive reminding' model of Jacoby and Wahlheim (2013) and Wahlheim and Jacoby (2013) which specifically makes reference to the importance of making contact with the episodic trace and the episodic memory proposal of Lehman, Karpicke and colleagues (Karpicke et al., 2014; Karpicke et al., 2014; Lehman et al., 2014).

Wahlheim and Jacoby (2013) forwarded a recursive reminding model to investigate the effects of proactive interference. This situation is very similar to the error-correction paradigm, insofar as the participant has learned a particular response A, but is then required to produce a new response, B, instead. In effect, A becomes an error, with B being the correct response. Interestingly, these authors showed that explicit remembering of A is related to *better* memory for the new response, B, just as in the error-generation paradigm, producing the error at time of encoding of the correction or B term, results in better memory for the correction, B.

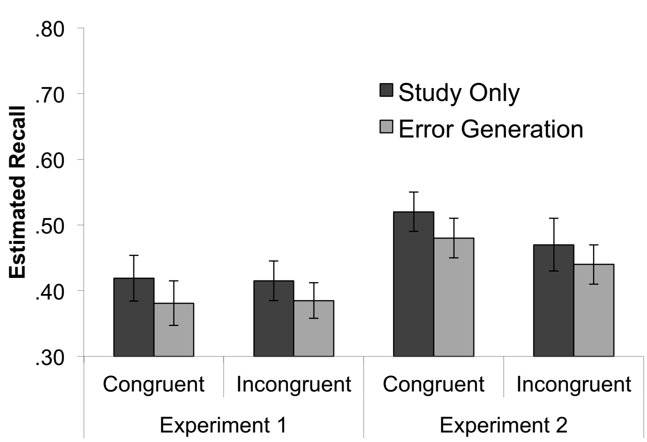

Fig. 2. Retrospective global metacognitive judgments about performance in each condition.

Similarly, episodic reminding effects resulted in facilitation of memory for the designated correct answer in retroactive memory paradigms (Wahlheim and Jacoby, 2013). The authors proposed that a process called 'recursive reminding' is responsible: if the person brings to mind the entire original episodic event at the time of encoding of the new response, and if they later recollect that event, memory will be enhanced via this *episodic* memory route. This theory is attractive because it seems capable of explaining many findings, ranging from spacing and recency effects (Jacoby and Wahlheim, 2013), to effects of proactive and retroactive facilitation (Negley et al., 2018), that, otherwise seem puzzling. It also suggests a way to test whether the error generation benefit–seen so commonly in studies with typical participants, but never in studies with amnesic patients—is episodic in nature. If the effect is episodic we should be able to find evidence that the benefit is linked to episodic retrieval of the error at time of the final test for the correct answer.

Two findings from our lab would seem to go against the recursive reminding explanation. Butterfield and Metcalfe (2006) asked subjects, in an error correction paradigm, to produce *two* answers and then circle the one that was correct. They conditionalized on whether the original error had been generated as one of those two responses. However, in contrast to what might be expected by the recursive reminding framework, they found that there was no relation between correct recall of the target and generation of the original error at time of retrieval. Similarly, Metcalfe and Miele (2014), in investigating whether errors returned over time, also requested that participants provide two responses to each question, in the final (post corrective feedback) test. Like Butterfield and Metcalfe (2006), they conditionalized to determine whether the 'show up' of the original error had an effect on whether or not the person produced the correction. They found no relation.

Although these results are not supportive, they are open to criticism. The instructions that Butterfield and Metcalfe (2006) and Metcalfe and Miele (2014) provided to participants were strikingly similar to those used to tap into *implicit* memory in amnesic patients. The participants were simply asked to write down whatever responses came to mind. They were not asked to recall the previous erroneous event. Only after generating several responses, were they required to say which response was correct.

It is possible that the erroneous responses at the time of final test—even those that were the same as the error that the participant had just generated moments ago—might have come to mind without the subject having accessed the original episode. If so, then our previous data showing that it did not matter whether the person generated the original error or not need not weigh against the Episodic Recollection/Recursive Reminding hypothesis. In Experiment 2, we endeavor to determine whether the locus of the error generation benefit in typical participants is, in fact, episodic—by asking them to recall the error that they themselves had previously generated.

In the experiment that follows, unlike in the experiments of Butterfield and Metcalfe (2006) and of Metcalfe and Miele (2014), people were not asked to generate just any response, erroneous or not, that came to mind, but rather to recollect their own previous error(s) from episodic memory. This experiment, then, provides a more rigorous test of the Recursive Reminding hypothesis and of the role of episodic memory in the error generation benefit. The hypothesis was that error correction would be contingent upon the generation of the original error.

3. Experiment 2

To test the Episodic Recollection hypothesis for the error generation effect, Experiment 2 was the same as Experiment 1 except that in addition to asking participants for the correct answer at final recall, participants were also prompted to provide their original error. If the recollection of the error episode was used in retrieving the correct response regardless of the error's semantic relation to the target, when the participant recollected the error, memory for the correct answer should be enhanced.

3.1. Method

Forty Columbia University undergraduate students (Mean Age = 19.4 (*SD* = 1.5), 52.5% female) participated for course credit. The materials and design were the same as those used in the Experiment 1, with a slight, but crucial, procedural difference on the final cued recall test. For each cue, participants provided the correct target *and* their original error if they had made one previously. If it had been a Study-Only trial where no original error was made, participants typed "NA" for not applicable. Participants either saw "CORRECT ANSWER?" or "ORIGINAL Response?" above the cue, depending on the trial type.

3.2. Results

Rates of correctly guessing the correct response during the original presentation were similar between materials: Congruent: $M = 0.08$, $SD = 0.08$; Incongruent: $M = 0.05$, $SD = 0.06$, $t(39) = 1.55$, $SE = 0.02$, $p = .13$. These trials were excluded from all other analyses. It is important to note, however, that LSA analysis showed that for the non-excluded trials, the errors generated during learning were more related to the congruent target ($M = 0.18$, $SD = 0.05$) than the incongruent target ($M = 0.15$, $SD = 0.05$), $t(39) = 2.13$, $SE = 0.01$, $p < .05$. Order of trial type (being prompted for the correct answer first vs. the original response first) was randomized and did not lead to any differences in cued recall performance, $F < 1$. We, therefore, collapsed over this variable.

As in Experiment 1, we conducted a 2 (Materials: Congruent, Incongruent) x 2 (Learning: Study Only, Error Generation) repeated measures ANOVA on proportion correct on the final cued recall. As is shown in Fig. 1, whereas participants remembered the target more often for Congruent items ($M = 0.60$, $SD = 0.21$) than for Incongruent items ($M = 0.53$, $SD = 0.22$), $F(1, 39) = 10.47$, $MSE = 0.01$, $p = .002$, $\eta p^2 = 0.21$, the Error Generation condition still resulted in better target recall ($M = 0.64$, $SD = 0.21$) than did the Study Only condition ($M = 0.50$, $SD = 0.22$), ($F(1, 39) = 37.92$, $MSE = 0.02$, $p < .001$, $\eta p^2 = 0.49$). This benefit for generation of errors over studying was evident for both congruent and incongruent materials (respectively, $t(39) = 7.12$, $SE = 0.02$ $p < .001$, $t(39) = 3.36$, $SE = 0.02$ $p < .01$); however, it was slightly larger in the congruent condition, as shown by the significant interaction, $F(1, 39) = 5.24$, $MSE = 0.01$, $p = .028$, $\eta p2 = 0.12$. This interaction was not present in Experiment 1. We compared performance between experiments for each of the four conditions. The only difference trending toward significance was that correct recall in the Congruent Study Only Condition was lower in Experiment 2 than Experiment 1; $t(78) = 1.86$, $SE = 0.05$, $p = .067$. All other comparisons were non-significant, $ts < 1$. We have no explanation for this difference in the Study Only condition in the two experiments.

The primary interest in Experiment 2, though, was in whether or not the episodic recollection of the original errors impacted correct recall. Overall, participants remembered their original errors on over half of the trials for both congruent ($M = 0.60$, $SD = 0.24$) and incongruent materials ($M = 0.64$, $SD = 0.25$), with memory for prior errors being slightly higher in the incongruent condition, $t(39) = 2.16$, $SE = 0.02$, $p < 05$. As predicted by the Recursive Reminding hypothesis, there was a strong contingency effect. As is shown in Fig. 3, when the original error was produced, correct recall of the target was much greater ($M = 0.71$, $SD = 0.24$) than when the error was not recollected ($M = 0.45$, $SD = 0.27$) ($F(1, 38) = 33.47$, $MSE = 0.01$, $p = .002$, $\eta p^2 = 0.47$).

Although congruent targets were recalled more (M = 0.63, SD = 0.27) than were the incongruent targets ($M = 0.53$, $SD = 0.25$), $F(1, 38) = 10.59$, $MSE = 0.04$, $p < .01$, $\eta p^2 = 0.22$, this effect was qualified by an interaction between Materials and Error Recall, $F(1, 38) = 6.09$, $MSE = 0.05$, $p = .018$, $\eta p^2 = 0.14$. When the original error was recollected, correct memory for the target was nearly identical in the congruent ($M = 0.71$, $SD = 0.26$) and incongruent conditions ($M = 0.70$, $SD = 0.23$), $t < 1$). However, when participants could ***not*** recall ***the original error***, correct recall of the target was much higher for congruent items ($M = 0.54$, $SD = 0.28$) as compared to the incongruent items ($M = 0.35$, $SD = 0.27$), $t(39) = 3.45$, $SE = 0.05$, $p < .01$.

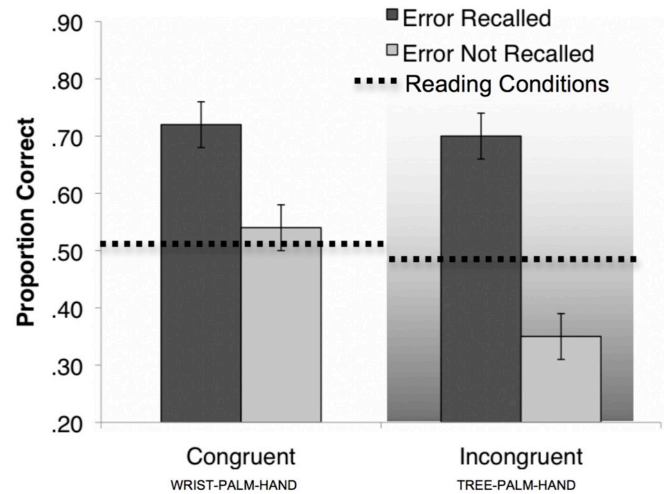

Fig. 3. Final recall of targets in the Error Generation Condition, for congruent and incongruent triads, conditional upon whether the error was recollected or not. Dashed lines indicate the recall level in the Study Only conditions.

Most importantly, when the errors that the individual had previously generated were not recalled, there was no benefit from being in the error-generation condition ($M = 0.45$, $SD = 0.27$) as compared to the read condition ($M = 0.50$ $SD = 0.22$), $F < 1$. Indeed, an interaction revealed that not recalling the error in the incongruent condition had more detrimental consequences than were observed for congruent materials, $F(1,39) = 7.31$, MSE = 0.05, $p = .01$, $\eta p^2 = 0.16$. When the original error was not recalled for Congruent items, performance in the Error Generation condition ($M = 0.54$, $SD = 0.28$) was the same as in the Study Only condition ($M = 0.51$, $SD = 0.20$), $t < 1$. However, for Incongruent items, when the error was not recollected, performance was significantly worse ($M = 0.35$, $SD = 0.27$) than it was in the Study Only condition ($M = 0.49$, $SD = 0.23$), $t(39) = 2.74$, $SE = 0.05$, $p < .01$. Failure to make contact with the episodic trace apparently resulted in the semantically Incongruent erroneous event producing massive interference.

Finally, participants, again, had very poor retrospective metacognitions, as shown in the right panel of Fig. 2. Whereas they had actually performed much better in the Error Generation conditions, they thought they had done better in the Study Only conditions, F(1,39) = 19.37, MSe = 0.06, p < .001, np = .33. Neither the effect of materials (F < 1) nor the interaction between materials and Learning (F(1,39) = 5.24, MSe = 0.06, p = .083, np = .08) were significant.

4. General discussion

The learning benefit that accrues from generating errors does not appear to be attributable to Semantic Mediation. The present results offer no comfort to the idea that the error, which is close in semantic space to the correct answer, provides a semantic stepping stone to the correct answer at time of retrieval. Instead, the error generation benefit appears to rely upon Episodic Recollection.

There are some limitations to our study. First, it would be desirable to have a greater separation in LSA between the errors and the target in the congruous and incongruous conditions. While it was intended that the errors in the incongruous condition be completely unrelated to the target, the results showed that they were still somewhat related. Second, there are still discrepancies in relating our results to past literature. Although our results seem consistent with those of Potts and Shanks (2014), who showed that error generation can have beneficial effects even when the error is unrelated to the target, they seem inconsistent with those of Huelser and Metcalfe (2012) and of Grimaldi and Karpicke (2012) who showed no error generation benefit with unrelated errors. Additionally, these data do not speak to why Bridger and Mecklinger (2014) found a condition in which making errors hurts later correction, even in neurologically typical participants. In their study two-letter stems were used to provoke a whole word response, and then feedback was given that response was incorrect and a different completion was given as the correct answer. The generation of the incorrect response hurt rather than helped later generation of the correct response. Speculatively, it is possible that these results–that seem inconsistent with the present findings–occurred because people may have generated the stem completions at final test without recourse to the episodic memory system. It may be possible that some tasks require or utilize episodic memory while other do not, and that those that do result in errors facilitating recall (at least n neurologically typical people) whereas those that do not result in error generation impairing performance. If the use or lack of use of the episodic memory system determines when error generation will help or hurt, then more research needs to be invested in determining when one or the other system will hold sway.

Limitations to our study notwithstanding, two findings converge on the conclusion that the error generation benefit is due to episodic recolllection rather than semantic mediation. First, the error does not appear to have to be semantically related to the correct target answer to result in an error generation benefit. Even errors that are unrelated can, under the right circumstances, produce enhanced memory for the correct answer. Second, the error generation benefit depends upon the error being episodically recollected at time of retrieval. When the episode in which the error was committed is not recollected, the benefit is not enjoyed. Indeed, interference may even result from initial error generation when episodic retrieval of the error at time of final recall fails. Both lines of evidence point to the error-generation benefit being fundamentally dependent on episodic, not semantic, memory.

Why, then, does it usually matter that the error that the person generates is semantically related to the target? Both Huelser and Metcalfe (2012) and Grimaldi and Karpicke (2012) found that when the cue and the target were related, error generation was beneficial, but when they were unrelated–and consequently the errors that were generated were unrelated to the target–error generation was not beneficial. The semantic mediation hypothesis would seem to provide a natural explanation for these findings. But that explanation is at odds with the results of the present experiments. It might be possible to begin to reconcile these findings by noting that semantically related questions, answers and errors usually afford a smoother and more compelling encoding into memorable distinctive episodic traces than do unrelated questions, answers, and errors. The particularly good episodic encoding induced by the unrelated errors in the incongruent condition in the present experiments may have been the exception to the poor encoding that usually accompanies unrelated items. These particular unrelated errors (and the surrounding event the participants experienced) may have been sufficiently interesting, fun, and stimulating to allow good episodic encoding. By this view, the semantic relatedness of the error usually matters, not because it later provides a semantic mediator or stepping stone during retrieval, but rather because it is more likely to afford a good and memorable initial encoding (and see Potts et al., 2019). Whether or not generating an error helps memory for the correct answer would appear to depend upon (a) the participant having a functional episodic memory system, (b) the task using that episodic system, and (c) the commission of the error itself garnering the person's attention to the correction and fostering an episodically memorable event. If all of these conditions are met, then error generation should enhance memory for the correct answer.

The results that were observed when the original errors could not be recollected resonate with the situation that deeply amnesic patients, purportedly, experience. The most striking finding that we observed was that of severe interference when the error was, itself, incongruent with the target. This effect was puzzling. If the error could not be remembered, why should it have any effect at all? This interfering effect seems akin to the debilitating effects of error generation observed in amnesic patients. In our experiments, then, provided the errors were recollected they enhanced memory for the correct answers. This finding is of importance for understanding human memory and it has wide ranging practical implications. It is important to note, though, that the beneficial effect is attributable to episodic, not semantic, memory. It mattered little whether the error was or was not semantically related to the target, as long as it (and by extension, the entire initial event) was recollected at time of retrieval. Interestingly, when the errors, and especially the incongruent errors, could not be recollected, it was not as if they had never occurred: the effects of having made an error were considerable. But they were detrimental to recall of the correct answer, just as is the case for amnesics.

Acknowledgments

This research was supported by Grant 484800 from the James S. McDonnell Foundation (PI, H. Roediger), Institute of Education Sciences grant R305A150467, and by a National Science Foundation graduate fellowship to BJH. The authors are solely responsible for the data and for the interpretation, herein. We thank Michael Smith, Rachel Burris, David Miele, Karen Kelly, Peter Messa, Lauren Robbins, Frank Messa, Amelia Messa, Teal Eich, and Lisa Son. Data are available upon request to Janet Metcalfe: jm348@columbia.edu.

Appendix A. Materials

Cue 1	Congruent 2	Incongruent 2	Target
WATCH	WRIST	OBSERVE	HAND
MOPED	DEPRESSED	SCOOTER	MELANCHOLY
BLUFF	LIE	CLIFF	FAKE
COBBLER	PEACH	SHOE	PIE
BASS	GUITAR	FISH	BAND
STALK	CORN	FOLLOW	BEAN
HORN	UNICORN	TRUMPET	BULL
DRAFT	COLD	ARMY	WIND
GLASSES	EYES	CUPS	SEE
FIRM	LAW	HARD	COMPANY
ROLL	BASKET	ROTATE	BREAD
STEAL	METAL	THIEF	SOLID

(*continued on next page*)

(*continued*)

Cue 1	Congruent 2	Incongruent 2	Target
SINK	DISHES	FLOAT	FAUCET
MINUTE	SECOND	TINY	HOUR
LOG	FIRE	JOURNAL	TREE
ROSE	CLIMB	THORN	ESCALATE
RACE	WIN	HUMAN	RUN
OBJECT	THING	COMPLAIN	ITEM
JAM	STUCK	SWEET	CROWDED
GROSS	DISGUSTING	INCOME	UGLY
DEGREE	HOT	SCHOOL	TEMPERATURE
SLING	SHOT	BROKEN	THROW
FAIR	GAMES	SKIN	CARNIVAL
COURSE	CLASS	PATH	SCHOOL
SQUASH	VEGETABLE	SMASH	YELLOW
HAMPER	STOP	LAUNDRY	SLOW
RAPID	FAST	RIVER	QUICK
SHED	HAIR	STORAGE	SKIN
TEAR	RIP	CRY	SPLIT
MIGHT	STRONG	POSSIBLE	POWER
SAW	PERCEIVE	TOOL	GLIMPSE
REFUSE	DENY	GARBAGE	REJECT
DRAWER	DRESSER	ARTIST	CLOTHES
CONSOLE	SYMPATHY	TELEVISION	COMFORT
COAST	OCEAN	GLIDE	BEACH
LEAD	FOLLOW	PENCIL	LEADER
BUCK	DOLLAR	DEER	MONEY
CURRENT	RIVER	NOW	WATER
ARMS	WAR	FINGERS	NUCLEAR
STABLE	HORSE	SECURE	BARN
CARDINAL	RED	PRIEST	BIRD
ESSENCE	SMELL	MEANING	PERFUME
GENERAL	SPECIFIC	ARMY	BASIC
CAN	ALUMINUM	ABLE	SODA
CHANNEL	TELEVISION	WATER	STATION
SLUG	WORM	PUNCH	SLIMY
BUREAU	CHEST	GOVERNMENT	DRESSER
POLISH	SHINE	LANGUAGE	BUFF
BOLT	LIGHTNING	SCREW	THUNDER
ORGAN	HEART	CHURCH	BODY
PRESENT	GIFT	CURRENT	BIRTHDAY
CRACK	ADDICT	PAVEMENT	COCAINE
EXPRESS	FAST	SHOW	TRAIN
PRESS	NEWSPAPER	BUTTON	MEDIA
COURT	JUDGE	TENNIS	LAW
HAIL	STORM	PRAISE	ICE
SEWER	THREAD	PIPES	SEAMSTRESS
BAT	CAVE	BASEBALL	NIGHT
BOW	BOAT	RIBBON	HULL
BRISK	FAST	COOL	QUICK

References

Anderson, J.R., 1983. A spreading activation theory of memory. J. Verbal Learn. Behav. 22 (3), 261–285.

Baddeley, A., Wilson, B.A., 1994. When implicit learning fails: amnesia and the problem of error elimination. Neuropsychologia 32, 53–68.

Bridger, E.K., Mecklinger, A., 2014. Errorful and errorless learning: the impact of cue-target constraint in learning from errors. Mem. Cogn. 42, 898–911.

Butterfield, B., Mangels, J.A., 2003. Neural correlates of error detection and correction in a semantic retrieval task. Cogn. Brain Res. 17, 793–817.

Butterfield, B., Metcalfe, J., 2001. Errors committed with high confidence are hypercorrected. J. Exp. Psychol. Learn. Mem. Cogn. 27, 1491–1494.

Butterfield, B., Metcalfe, J., 2006. The correction of errors committed with high confidence. Metacognition Learn. 1, 1556–1623.

Carpenter, S.K., 2011. Semantic information activated during retrieval contributes to later retention: support for the mediator effectiveness hypothesis of the testing effect. J. Exp. Psychol. Learn. Mem. *Cogn.* 38 (6), 1547–1552.

Carpenter, S.K., Yeung, K.L., 2017. The role of mediator strength in learning from retrieval. J. Mem. Lang. 92, 128–141.

Collins, A.M., Quillian, M.R., 1972. Experiments on semantic memory and language comprehension. In: Gregg, L.W. (Ed.), Cognition in Learning and Memory. John Wiley, New York, pp. 117–138.

Collins, A.M., Loftus, E.F., 1975. A spreading activation of semantic processing. Psychol. Rev. 82, 401–428.

Cyr, A.A., Anderson, N.D., 2013. Updating misconceptions: effects of age and confidence. Psychon. Bull. Rev. 20, 574–580.

Dunlosky, J., Hertzog, C., Powell-Moman, A., 2005. The contribution of mediator-based deficiencies to age differences in associative learning. Dev. Psychol. 4, 389–400.

Eich, T.S., Stern, Y., Metcalfe, J., 2013. The hypercorrection effect in younger and older adults. Aging Neuropsychol. Cognit. 20, 511–521.

Fazio, L.K., Marsh, E.J., 2009. Surprising feedback improves later memory. Psychon. Bull. Rev. 16, 88–92.

Fazio, L.K., Marsh, E.J., 2010. Correcting false memories. Psychol. Sci. 21, 801–803.

Finn, B., Metcalfe, J., 2010. Scaffolding feedback to maximize long-term error correction. Mem. Cogn. 38, 951–961.

Glisky, E.L., Schacter, D.L., 1988. Long-term retention of computer learning by patients with memory disorders. Neuropsychologia 26, 173–178.

Glisky, E.L., Schacter, D.L., Tulving, E., 1986a. Learning and retention of computer related vocabulary in memory-impaired patients: method of vanishing cues. J. Clin. Exp. Neuropsychol. 8, 292–312.

Glisky, E.L., Schacter, D.L., Tulving, E., 1986b. Computer learning by memory-impaired patients: acquisition and retention of complex knowledge. Neuropsychologia 24, 313–328.

Grimaldi, P.J., Karpicke, J.D., 2012. When and why do retrieval attempts enhance subsequent encoding? Mem. Cogn. 40, 505–513.

Hamann, S.B., Squire, L.R., 1995. On the acquisition of new declarative knowledge in amnesia. Behav. Neurosci. *109*, 1027–1044.

Hausman, H., Rhodes, M.G., 2018. Retrieval activates related words more than presentation. Memory 26, 1265–1280.

Hayman, C.A., MacDonald, C.A., Tulving, E., 1993. The role of repetition and associative interference in new semantic learning in amnesia: a case experiment. J. Cogn. Neurosci. 5, 375–389.

Hays, M.J., Kornell, N., Bjork, R.A., 2013. When and why a failed test potentiates the effectiveness of subsequent study. J. Exp. Psychol. Learn. Mem. Cogn. 39, 290–296.

Huelser, B.J., Metcalfe, J., 2012. Making related errors facilitates learning, but learners do not know it. Mem. Cogn. 40, 514–527.
Iwaki, N., Matsushima, H., Kodaira, K., 2013. Hypercorrection of high confidence errors in lexical representations. Percept. Mot. Skills 117, 219–235.
Izawa, C., 1967. Function of test trials in paired-associate learning. J. Exp. Psychol. 76, 194–209.
Izawa, C., 1970. Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. J. Exp. Psychol. 83, 340.
Jacoby, L.L., Wahlheim, C.N., 2013. On the importance of looking back: the role of recursive remindings in recency judgments and cued recall. Mem. Cogn. 41, 625–637.
Kang, S.H.K., Pashler, H., Cepeda, N.J., Rohrer, D., Carpenter, S.K., Mozer, M.C., 2011. Does incorrect guessing impair fact learning? J. Educ. Psychol. 131, 48–59.
Karpicke, J.D., Lehman, M., Aue, W.R., 2014. Retrieval-based learning: an episodic context account. In: Ross, B.H. (Ed.), The Psychology of Learning and Motivation, vol. 61. Elsevier Academic Press, San Diego CA, pp. 237–284.
Karpicke, J.D., Roediger, H.L., 2007. Repeated retrieval during learning is the key to long- term retention. J. Mem. Lang. 57, 151–162.
Kornell, N., Hays, M.J., Bjork, R.A., 2009. Unsuccessful retrieval attempts enhance subsequent learning. J. Exp. Psychol. Learn. Mem. Cogn. 35, 989–998.
Kulhavy, R.W., Yekovich, F.R., Dyer, J.W., 1976. Feedback and response confidence. J. Educ. Psychol. 68, *5*22–528.
Landauer, T.K., Dumais, S.T., 1997. A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. Psychol. Rev. 104, 211–240.
Landauer, T.K., Foltz, P.W., Laham, D., 1998. Introduction to latent semantic analysis. Discourse Process 25, 259–284.
Lehman, M., Karpicke, J.D., 2016. Do semantic mediators improve memory? J. Exp. Psychol. Learn. Mem. Cogn. 42, 1573–1591.
Lehman, M., Smith, M.A., Karpicke, J.D., 2014. Toward an episodic context account of retrieval-based learning: dissociating retrieval practice and elaboration. J. Exp. Psychol. Learn. Mem. Cogn. 40, 1787–1794.
Marcel, T., 1980. Recognition of polysemus words: locating the selective effects of prior verbal context. In: Nickerson, R.S. (Ed.), Attention and Performance VIII. Lawrence Erlbaum Associates, New Jersey, pp. 435–457.
Metcalfe, J., 2017. Learning from errors. Annu. Rev. Psychol. 68, 465–489.
Metcalfe, J., Butterfield, B., Habeck, C., Stern, Y., 2012. Neural correlates of people's hypercorrection of their false beliefs. J. Cogn. Neurosci. 24, 1571–1583.
Metcalfe, J., Casal-Roscum, L., Radin, A., Friedman, D., 2015. On teaching old dogs new tricks. Psychol. Sci. 12, 1833–1844.
Metcalfe, J., Eich, T.S., 2019. Memory and truth: correcting errors with true feedback versus overwriting correct answers with errors. Cogn. Res.: Princ. Implications 4, 221–233.
Metcalfe, J., Finn, B., 2011. People's hypercorrection of high-confidence errors: did they know it all along? J. Exp. Psychol. Learn. Mem. Cogn. 37, 437–448.
Metcalfe, J., Finn, B., 2012. Hypercorrection of high confidence in children. Learn. Instr. 22, 253–261.
Metcalfe, J., Miele, D.B., 2014. Hypercorrection of high confidence errors: prior testing both enhances delayed performance and blocks the return of the errors. J. Appl. Res. Mem. Cogn. 3, *1*89–197.
Neely, J.H., 1976. Semantic priming and retrieval from lexical memory: evidence for facilitatory and inhibitory processes. Mem. Cogn. 4, 648–654.
Negley, J.H., Kelley, C.M., Jacoby, L.L., 2018. The importance of time to think back: the role of reminding in retroactive effects of memory. J. Exp. Psychol. Learn. Mem. Cogn. 44, 1352–1364.
Nelson, D.L., McEvoy, C.L., Schreiber, T.A., 1998. The University of South Florida word association, rhyme, and word fragment norms. Retrieved from. http://www.usf.edu/FreeAssociation/.
Pastötter, B., Schicker, S., Niedernhuber, J., Bäuml, K.L., 2011. Retrieval during learning facilitates subsequent memory encoding. J. Exp. Psychol. Learn. Mem. Cogn. 37, 287–297.
Potts, R., Davies, G., Shanks, D.R., 2019. The benefit of generating errors during learning: what is the locus of the effect? J. Exp. Psychol. Learn. Mem. Cogn. 45, 1023–1041.
Potts, R., Shanks, D.R., 2014. The benefit of generating errors during learning. J. Exp. Psychol. Gen. 143, 644–667.
Pyc, M.A., Rawson, K.A., 2010. Why testing improves memory: mediator effectiveness hypothesis. Science 330, 335.
Ratcliff, R., McKoon, G., 1994. Spreading activation versus compound cues theory of memory. Psychol. Rev. 1*01*, 185–187.
Richland, L.E., Kao, L.S., Kornell, N., 2009. The pretesting effect: do unsuccessful retrieval attempts enhance learning? J. Exp. Psychol. Appl. 15, 243–257.
Rosenbaum, R.S., Kohler, S., Schacter, D.L., Moscovitch, M., Westmacott, R., Black, S.E., Gao, F., Tulving, E., 2005. The case of K.C.: contributions of a memory-impaired person to memory theory. Neuropsychologia 43, 989–1021.
Schacter, D.L., Moscovitch, M., Tulving, E., McLachlan, D.R., Freedman, M., 1986. Mnemonic precedence in amnesic patients: an analogue of the AB error in infants? Child Dev. 57, 816–823.
Sitzman, D.M., Rhodes, M.G., Tauber, S.K., 2014. Prior knowledge is more predictive of error correction than is subjective confidence. Mem. Cogn. 42, 84–96.
Sitzman, D., Rhodes, M., Tauber, S., Liceralde, V.R.T., 2015. The role of prior knowledge in error correction for younger and older adults. Aging Neuropsychol. Cognit. 22, 502–516.
Slamecka, N.J., Fevreiski, J., 1983. The generation effect when generation fails. J. Verb. Learn. Verb. Behav. 22, 153–163.
Tulving, E., 1983. Elements of Episodic Memory. Clarendon Press, Oxford.
Tulving, E., 2005. Episodic memory and autonoesis: uniquely human? In: Terrace, H.S., Metcalfe, J. (Eds.), *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*. Oxford University Press, New York.
Tulving, E., Hayman, C.A.G., Macdonald, C.A., 1991. Long-lasting perceptual priming and semantic learning in amnesia: a case experiment. J. Exp. Psychol. Learn. Mem. Cogn. 17, 595–617.
Van Loon, M.H., Dunlosky, J., van Gog, T., van Merrienboer, J.J.G., de Bruin, A.B.H., 2015. Refutations in science texts leads to hypercorrections of misconceptions held with high confidence. Contemp. Educ. Psychol. 42, 39–48.
Wahlheim, C.N., Jacoby, L.L., 2013. Remembering change: the critical role of recursive remindings in proactive memory. Mem. Cogn. 41, 1–15.