

RESEARCH ARTICLE

Can we learn from errors? Retrieval facilitates the correction of false memories for pragmatic inferences

María J. Maraver ^{*}, Ana Lapa, Leonel Garcia-Marques , Paula Carneiro, Ana Raposo 

Faculdade de Psicologia, CICPSI, Universidade de Lisboa, Lisbon, Portugal

^{*} mjmaraver@psicologia.ulisboa.pt OPEN ACCESS

Citation: Maraver MJ, Lapa A, Garcia-Marques L, Carneiro P, Raposo A (2022) Can we learn from errors? Retrieval facilitates the correction of false memories for pragmatic inferences. PLoS ONE 17(8): e0272427. <https://doi.org/10.1371/journal.pone.0272427>

Editor: Alessandra S. Souza, University of Zurich, SWITZERLAND

Received: December 3, 2021

Accepted: July 19, 2022

Published: August 2, 2022

Copyright: © 2022 Maraver et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The datasets presented in this study can be found in the following Open Science Framework project identifier: DOI [10.17605/OSF.IO/JVU4K](https://doi.org/10.17605/OSF.IO/JVU4K).

Funding: AR, PC: PTDC/PSI-ESP/28414/2017, Fundação para a Ciência e a Tecnologia (Foundation for Science and Technology of Portugal). <https://www.fct.pt/> AL: PD/BD/114066/2015, Fundação para a Ciência e a Tecnologia (Foundation for Science and Technology of Portugal). <https://www.fct.pt/> The funders had no

Abstract

Errorful learning suggests that, when perfect learning has not yet been attained, errors can enhance future learning if followed by corrective feedback. Research on memory updating has shown that after retrieval, memory becomes more malleable and prone to change. Thus, retrieval of a wrong answer might provide a good context for the incorporation of feedback. Here, we tested this hypothesis using sentences including pragmatic sentence implications, commonly used for the study of false memories. Across two experiments with young adults, we hypothesized that corrective feedback would be more efficient at reducing false memories if provided immediately after retrieval, when memory is more malleable than after being exposed to the material. Participants' memory was assessed as a function of the type of learning task (Experiment 1: retrieval vs. restudy; and Experiment 2: active vs. passive recognition); and whether participants received corrective feedback or not. In both experiments, we observed that retrieval not only improved correct recall (replicating the testing effect) but also promoted the correction of false memories. Notably, corrective feedback was more effective when given after errors that were committed during retrieval rather than after restudy (Experiment 1) or after passive recognition (Experiment 2). Our results suggest that the benefits of retrieval go beyond the testing effect since it also facilitates false memories correction. Retrieval seems to enhance memory malleability, thus improving the incorporation of feedback, compared to the mere presentation of the information. Our results support the use of learning strategies that engage in active and explicit retrieval because, even if the retrieved information is wrong—when immediate feedback is provided—memory updating is promoted and errors are more likely to be corrected.

Introduction

Making errors is uncomfortable, even though we frequently make them. Especially in educational settings, educators may be reluctant to ask students to guess before learning the correct information for fear that their incorrect guesses would be confused with true items and be harmful to future learning. On the other hand, many theories of associative and predictive learning [1,2] assume that learning can only occur after performance or predictive errors

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

(followed by accurate feedback). What is the most optimal strategy for promoting proper learning is still an unresolved question: whether and in what conditions, if any, it is better to avoid making errors than to commit, explore, and correct mistakes [3]. This study aimed to explore the consequences of making errors during the learning process and to contribute to the current debate on whether and under which conditions, generating errors can be beneficial or detrimental to learning.

Retrieving previously learned information from memory has been widely evidenced as one of the most effective learning strategies. This robust finding, known as the “testing effect” or “retrieval practice effect”, shows that when learned information is retrieved by taking memory tests, it is more likely to be correctly recalled in the future than when the information is reread or restudied [4,5]. The testing effect has been highly replicated both in the laboratory and school settings. Yet, these studies have traditionally focused on successful retrieval [4–7], that is, when the retrieved information is correct. Much less is known about the impact of unsuccessful retrieval on subsequent learning. When a wrong answer is given in a memory test, it remains unclear what happens to the subsequent retrievability of the correct answer.

In the last two decades, an increasing number of studies have shown that generating errors, as long as followed by feedback, can enhance later memory for the correct response [3,8,9]. These results contradict classic interference theories that suggest that a generated wrong response might interfere and hinder the learning of the correct answer because of an enhanced memorability of the erroneous response [10,11]. This rationale is supported by traditional theories of learning and memory that advocate that errors should be avoided, and teaching should follow an *errorless learning* procedure encouraging students to make as few errors as possible [10,12–14].

However, associative and predictive learning models argue that the revision of associative weights (associative learning) between predictive and criterion items are a function of performance or predictive errors, such that learning cannot improve without errors [1,2]. Early studies by Izawa [15,16] introduced the hypothesis that in comparison to errorless learning, the generation of errors, as long as it is followed by corrective feedback, results in better memory for the correct response. These *errorful learning* perspectives propose that failed retrieval attempts can improve learning given that the occurring mismatch between the wrong response and the corrective feedback would increase attention to the encoding of the feedback and, therefore, strengthen learning [9]. When an error is committed, providing feedback with the availability of the correct answer is paramount to process, understand and integrate the corrected information into memory [17,18].

Various accounts have been put forward to explain the benefits of error generation and error correction. Retrieving a memory trace, even if wrong, could potentially strengthen retrieval pathways to related content, identifying a need for additional information, and turning memory into a fertile ground susceptible to the encoding of new information [19]. According to the reconsolidation hypothesis, retrieving a consolidated memory can increase its malleability within a limited time window, in which the memories become susceptible to modification [20,21]. One of the most adaptive properties of our memory is its function to update and revise the contents. In this sense, memory updating allows our memory to stay current and up to date, eliminating information that is no longer needed to avoid error perpetuation [22]. The Memory Updating After Retrieval (MUAR) framework explains how retrieval is a strong modifier of memory. It proposes that retrieval attempts make memory more malleable, promoting an optimal environment for the updating of information relevant for retrieval, in this case, the incorporation of the feedback [22]. The finding that knowledge can be successfully updated after self-generated errors has been demonstrated with experimental paradigms such as last word generation in sentences [23], word-definition or translation pairs [9,24],

generation of antonyms [25], or related word pairs [26–29], on immediate and delayed follow-up tests. Kornell and colleagues demonstrated that producing errors led to better learning in comparison to merely being exposed to the correct response using the standard error-generation paradigm with different materials for cue-target pairs (i.e., general-knowledge questions or weak associates word pairs). Across experiments, the authors found that guessing a wrong answer before being provided with corrective feedback enhanced learning on a final memory test, compared to conditions in which the question and the answer were simultaneously presented [26]. Such a beneficial effect of error generation has been replicated with similar guessing errors that were semantically related to the correct answer [29–33]. The guessing methodology has the advantage of ensuring a high number of unsuccessful retrieval attempts. However, guesses do not represent all types of errors, and in most cases, responses are only right or wrong in the context of the experiment. For example, in the weak-associate paradigm, the generation of the strong associate after the presentation of the cue (e.g., snow-white) is often considered a guessing error, whereas the cue-weak associate pair (e.g., snow-balls) represents the correct response. Therefore, these types of mistakes are not genuine generated errors, but arbitrary guesses. The differences between guessing and retrieval may be of great importance to the understanding of the beneficial effects of error correction. According to the semantic mediation hypothesis, the benefit of error generation occurs when the error can act as a semantic mediator, thereby increasing the retrievability of the correct item and facilitating the target retrieval (see for instance, [34,35]). This explanation would account for the finding that, when the guessing error and the target are semantically related, retrieval is enhanced [26,32,35–37]. Other research however has found no effect or even a detrimental effect of guessing when unrelated word pairs were learned [30,31,38–40]. As Metcalfe and Huesler [28] argue: “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 [41–43] do not similarly benefit” [28]. Further limitations of the semantic mediation hypothesis come from Metcalfe and Huesler [28] who showed that guessing errors that were semantically incongruent with the target (i.e., “tree-palm” for the target “hand” vs. “wrist-palm” as the congruent condition), still provided performance benefits as long as both (error and item) were recalled. Thus, an explanation in terms of semantic mediation is clearly incomplete and a more comprehensive interpretation must include the role of episodic recollection. That is, when information is retrieved from memory (even if it is wrong), the answer is computed through the search and navigation into a richer episodic network, probably requiring a more elaborated form of processing than the one underlying (uninformed) guessing.

Because of the limitations of guessing errors, it remains unexplored what the benefits of error correction are when using materials that generate responses that involve an episodic retrieval of a previous context or a learning event. This question is key to understand the benefits of error correction given that research has suggested that a guessing error is only beneficial to learning when it can be used as a semantic mediator [26,32,36,37] but not when unrelated word pairs are learned [30,31,38,39]. However, the MUAR framework suggests that when information is retrieved from memory (even if incorrectly), memory becomes more vulnerable and amenable to change, making it more prone to be updated with new information [22] –as would be the case of corrective feedback–suggesting that retrieval might be a promotor for error correction. Carneiro and colleagues provided evidence for this hypothesis, using DRM lists that required the study of semantically related words (e.g., “bed, tired, rest, pillow, dream”), that often led to falsely remembering a non-studied critical lure (e.g., “sleep”) at a later memory test. The authors compared the effect of presenting new information, (i.e., corrective feedback) when provided after retrieval vs. after restudy [44]. Consistent with the

MUAR framework, their results suggested that retrieval followed by feedback is a better learning environment than restudy. But even if this result was found using a paradigm that involved episodic retrieval, it disregarded two important aspects. First, retrieval and restudy practices were not assessed when no corrective feedback was given, making it difficult to disentangle what is the actual impact that providing corrective feedback had on learning practices. And second, previous research has shown that the associative memory errors generated with the DRM lists correlate poorly with other types of memory errors such as false-event suggestion (remembering an event that did not happen) or misinformation (updating a previous memory with false information), that involve the episodic recall of real-life events [45]. Therefore, it becomes critical to extend the results obtained by Carneiro et al. [44] and study the effect of retrieval on error correction but using materials that involve episodic retrieval and are more ecologically valid.

The present work attempted to overcome these limitations and to address in-depth the effect of retrieval followed by feedback on the correction of memory errors, by adopting a paradigm with greater ecological validity. Sentences including pragmatic implications are a useful tool as they involve an episodic retrieval of a previous context or a learning event and represent more closely the type of errors that occur in real-world situations. For example, the sentence “*The boy lost balance on the skate*” is usually falsely remembered as “*The boy fell from the skate*”, or “*The baby stayed awake all night*” is very often misremembered as “*The baby cried all night*”, when in both cases the remembered outcomes were never explicitly presented [46]. Pragmatic inferences have been widely used for the study of false memories for everyday actions since memory errors are easily created, vivid, and held with high confidence [46–48]. Importantly, they allow for the dissociation between a semantic (i.e., the concepts of baby and to cry are semantically related), and an episodic component (i.e., the explicitly presented sentence “*The baby stayed awake all night*”). Therefore, by using pragmatic inferences we could evaluate the generation of errors that stem from episodic retrieval, and the effects of their correction on subsequent learning.

We tested the hypothesis of whether the retrieval of memory errors generated from pragmatic inferences could improve learning compared to conditions in which the information was merely presented, while contrasting the presence and absence of corrective feedback. Two experimental studies with young adults were conducted using pragmatic inference sentences in Portuguese [47]. Both experiments included a first learning encoding phase, a second intermediate phase in which our key manipulations were applied, and a final cued-recall memory test in which learning was assessed. In Experiment 1, we compared the effects of the learning task (retrieval vs. restudy) and the presentation of feedback (present vs. absent). During the intermediate phase of Experiment 1, participants either performed a cued-recall retrieval task in which the fragment that triggered the pragmatic inference of the sentence had to be completed (i.e., “*The baby _____ all night.*”); or a restudy task in which half of the sentences were presented in their correct form (i.e., the same sentence that was presented during encoding “*The baby stayed awake all night*”) and the other half in an incorrect form (i.e., different from encoding “*The baby cried all night*”). As for the feedback manipulation, Mullet & Marsh suggest that two factors are essential to correct false memories [49]. First, the learner needs to realize a mistake has been made. Therefore, feedback should be presented immediately after an error is committed to allow learners to notice the discrepancy between the error and the correct information. And second, there is empirical and theoretical consensus that for feedback to be effective, learners need to know the correct information and not only that a mistake was made [18,22], so that explicit corrective feedback is essential to promote successful error correction [49]. Thus, in the current experiments, when corrective feedback was provided, it was

presented immediately after the responded or restudied sentence, and it contained the correct information, (i.e., the sentence studied at the initial encoding phase).

Retrieval and restudy conditions of Experiment 1 led to an unbalanced proportion errors and to different types of errors across conditions. Specifically, in the retrieval condition, the proportion of errors generated varied across participants, whereas in the restudy condition, for all participants, 50% of the sentences presented contained an error. Therefore, to increase experimental control and to even the errors' ratio that participants committed versus were exposed to, we conducted a second experiment in which we implemented a yoked design and used a recognition instead of a cued-recall task in the intermediate phase. In Experiment 2, experimental conditions were compared as a function of feedback (present vs. absent as in Experiment 1), and the recognition paradigm (active vs. passive). Participants in the active recognition condition were presented with sentences that could match or mismatch the sentences presented at encoding and had to respond whether it was correct (same as encoding) or incorrect (different from encoding), as in a typical recognition test. Participants in the passive recognition condition were simply instructed to read the stimuli that was about to be presented. They were presented with the same sentences that participants in the active condition saw, but instead of having to provide an answer, they were shown the responses that a participant in the active condition had given. Lastly, in both Experiments 1 and 2, participants completed a final cued-recall test where the proportion of correct responses, pragmatic inferences errors, intrusions and omissions were assessed to measure the impact that the learning conditions and feedback had on memory performance.

For both studies, we hypothesized that, overall, retrieval would lead to a better final memory performance for sentences correctly recalled at the intermediate phase thus, replicating the testing effect [5]. Moreover, we anticipated that, when no feedback was provided, retrieval would lead to an enhancement of false memory errors compared to the conditions in which the information was merely presented (restudy for Experiment 1, and passive recognition for Experiment 2). However, when corrective feedback was provided, we predicted retrieval at the intermediate phase to decrease false memory errors in the final memory test. Because after retrieval memory seems to be more prone to change [22], even when the retrieved information is incorrect, it might result in a deeper and more elaborated processing which, in turn, would lead to a richer context for encoding the corrective feedback. Henceforth, we expected corrective feedback to be better incorporated into the retrieved memories, than when it is presented after information that was restudied (Experiment 1) or passively recognized (Experiment 2).

Experiment 1

Methods

Participants. The sample size was determined according to sample resources availability on a priori grounds—based on previous studies with similar experimental protocols manipulating error correction [44] or using pragmatic inferences as experimental material [50]. 120 university students from the University of Lisbon participated in the experiment ($M_{\text{age}} = 21.05 \pm 6.24$; 100 female) and were rewarded with 10€ vouchers for their participation. Participants were randomly assigned to one out of the four experimental conditions by crossing the two manipulated factors: learning condition (retrieval ($n = 30$) vs. restudy ($n = 30$)), and corrective feedback (present ($n = 30$) vs. absent ($n = 30$)).

All experimental procedures were approved by the Ethics Committee of the Faculty of Psychology of the University of Lisbon. Prior to the start of the experimental session, participants received information about the study and provided written consent for their participation in accordance with the Declaration of Helsinki [51].

Materials. Sixty sentences including pragmatic inferences in Portuguese from Carneiro et al. [47] were previously pretested to assess the proportion of pragmatic inferences falsely recalled. The thirty sentences with the highest proportion of false recall were selected and used in the current study. These thirty sentences were pretested again to assess their average proportion of pragmatic inferences when only thirty sentences composed the task. Pragmatic inferences for these sentences were generated, on average, 54% of the time.

Design. This experiment followed a 2 (learning condition: retrieval vs. restudy) \times 2 (corrective feedback: present vs. absent) factorial design, with both variables manipulated between participants.

Procedure. Participants performed the task individually on computers at the laboratory of the Faculty of Psychology of the University of Lisbon. The experimental task included three separate phases, and the procedure is illustrated in Table 1. During the encoding phase, all participants were instructed to read the sentences presented on the computer screen and to memorize them (e.g., *The baby stayed awake all night*). Sentences were presented in a black font, for 4.5 seconds, in random order. After the presentation of each sentence, participants had 5 seconds to perform an easy arithmetic operation (e.g., $23-5 = ?$) to prevent retrieval, as used in previous protocols [50]. Before the start of the task, participants completed a practice trial with five sentences (without pragmatic implications). After the encoding phase, participants performed a distractor task for 5 minutes, where they were asked to count the differences between 4 pairs of images.

Afterwards, participants completed the intermediate phase in which the key manipulations were applied. In the retrieval condition, participants performed a cued-recall test: each sentence from encoding was now presented in a random order with a critical fragment missing (e.g., *The baby ____ all night*). This cue was presented for 10.5 seconds in blue font. Participants were asked to complete the missing fragment with the information presented at encoding and had a maximum of 20 seconds to provide an answer. In the restudy condition, participants were told that they would be presented with the recall output of another person from a previous experimental session, and their task was simply to read that sentence. Participants were presented with sentences from encoding, in a random order, but to mimic the errors that participants in the retrieval condition are expected to make at this phase, 50% of the presented sentences were correct (presented in the same format as encoding) whereas the other 50% were incorrect (presented in a different format from encoding). This ratio of correct to incorrect sentences matched approximately the percentage of pragmatic inference errors expected for those in the retrieval condition, based on the pilot study for the same sentences (0.54 rate of false recall). The correct/incorrect format of each sentence was counterbalanced across participants. Each sentence was presented for 10.5 seconds in blue font. Across both

Table 1. Procedure scheme of Experiment 1.

Encoding phase				
<i>The baby stayed awake all night</i>				
$5 + 4 = 12$				
Intermediate phase				
Task	Retrieval	<i>The baby ____ all night</i>	Feedback	<i>The baby stayed awake all night</i>
			No feedback	$23-7 =$
	Restudy	<i>The baby cried all night</i>	Feedback	<i>The baby stayed awake all night</i>
			No feedback	$36 + 6 =$
Final cued-recall test				
<i>The baby ____ all night</i>				

<https://doi.org/10.1371/journal.pone.0272427.t001>

conditions, after retrieval or restudy of each sentence, participants in the feedback condition, were provided with corrective feedback: the correct sentence (i.e., in the same format as encoding) was presented for 4.5 seconds, in a black font, matching the color of encoding; participants in the no feedback condition were not presented with the correct sentence, instead, performed an easy arithmetic operation after each sentence for 4.5 seconds. After the intermediate phase, all participants performed another distractor task, like the previous one, for 1 minute. Finally, all participants performed a final cued-recall test in which all 30 sentences from encoding were presented with the critical fragment missing, and to be completed (i.e., *The baby _____ all night*). The sentences were presented in a random order, participants had 60 seconds to provide each answer and no feedback was provided. After completing the final cued-recall test, participants were thanked and rewarded for their time.

Statistical analyses. The list of experimental sentences used and the corresponding coding criteria for correct responses and pragmatic inference errors can be found in the supplementary material. Like in previous protocols [50], responses at the intermediate cued-recall test for those in the retrieval condition, and in the final cued-recall test for all participants were recorded following an adaptation of the standard coding procedure from Brewer [46], resulting in four response types: i) correct responses, which corresponded to those answers matching the original sentence or synonyms maintaining the original meaning of the sentence (i.e., “*put his lips towards*”, “*approached*” for the sentence “*The charming prince gently put his lips towards Snow White’s cheek*”); ii) pragmatic inference responses, which included responses that matched the expected pragmatic inferences or their synonyms (i.e., “*kissed*”), meeting the but-not test (i.e., “*The charming prince gently put his lips towards Snow White’s cheek, but did not kiss her*”); iii) intrusions, corresponding to other alternative answers (i.e., “*touched*”, “*smelled*”, “*looked*”); and iv) omissions, for responses left blank. The proportion of each response type was calculated, resulting in four dependent variables.

To compare memory retrieval at the intermediate phase between participants in the feedback and no feedback conditions, independent samples *t*-tests were performed over the proportion of recall of the four response types (correct responses, pragmatic inferences, intrusions, and omissions). Next, between-subjects ANOVAs were performed over the proportion of recall of each response type as a function of the learning condition (retrieval vs. restudy) and the presence of feedback at the intermediate phase (present vs. absent). Finally, two indexes were calculated as a function of the responses given at the intermediate and final cued-recall phases in order to analyze the persistence of correct responses and the correction of errors. Analyses were run using IBM SPSS Statistics software version 25 (IBM Corp., Armonk, NY, United States).

Results

Memory performance at the intermediate phase. At the intermediate phase, we observed no differences in memory performance between participants that did and did not receive feedback after the retrieval task (see Table 2).

Memory performance at the final cued-recall test. Table 3 summarizes the proportion of cued-recall at the final test for each experimental condition and response type for Experiments 1 and 2.

Results of the ANOVA revealed a higher proportion of correct responses in the final cued-recall test for the retrieval condition compared to the restudy ($F(1,116) = 4.97, p = 0.03, \eta^2_p = 0.04$), replicating the testing effect. The presentation of feedback also resulted in more correct responses compared to not receiving feedback ($F(1,116) = 282.88, p < 0.01, \eta^2_p = 0.71$). And a significant interaction between the retrieval task and feedback ($F(1,116) = 36.28, p < 0.01, \eta^2_p =$

Table 2. Memory performance at the retrieval condition of the intermediate phase.

	Feedback	No feedback	t(58)	p	95% Confidence Interval
Correct responses	0.29 ±0.15	0.23±0.11	1.74	0.87	[-0.01–0.13]
Pragmatic inferences	0.61±0.15	0.64±0.11	0.95	0.34	[-0.10–0.03]
Intrusions	0.04±0.07	0.08±0.07	1.87	0.07	[-0.07–0.00]
Omissions	0.06±0.05	0.05±0.05	0.68	0.50	[-0.02–0.03]

Descriptive statistics (mean ± standard deviations) and independent samples *t*-test statistics for each response type in the feedback and no feedback groups of the retrieval condition at the intermediate phase.

<https://doi.org/10.1371/journal.pone.0272427.t002>

0.24) showed that the beneficial effect of giving feedback was larger for retrieval ($F(1,58) = 440.82, p < 0.01, \eta^2_p = 0.88$) than for the restudy condition ($F(1,58) = 41.38, p < 0.01, \eta^2_p = 0.42$).

A similar pattern of results was observed for pragmatic inference errors. Participants who performed the retrieval task during the intermediate phase generated a lower proportion of pragmatic inference errors at the final test than those who restudied the material ($F(1,116) = 7.92, p < 0.01, \eta^2_p = 0.06$). Receiving feedback similarly reduced the proportion of pragmatic inferences compared to no feedback ($F(1,116) = 254.50, p < 0.01, \eta^2_p = 0.69$). The significant interaction ($F(1,116) = 20.44, p < 0.01, \eta^2_p = 0.15$) revealed that the reduction of pragmatic inference errors following feedback was greater for the retrieval ($F(1,58) = 322.63, p < 0.01, \eta^2_p = 0.85$) than for the restudy condition ($F(1,58) = 48.39, p < 0.01, \eta^2_p = 0.45$).

Table 3. Memory performance at the final cued-recall test.

Correct responses						
	Experiment 1			Experiment 2		
	Retrieval	Restudy	Total	Active recognition	Passive recognition	Total
Feedback	0.83±0.10	0.61±0.21	0.72±0.20	0.71±0.19	0.51±0.28	0.61±0.26
No feedback	0.22±0.12	0.32±0.12	0.28±0.13	0.27±0.15	0.23±0.17	0.25±0.16
Total	0.53±0.33	0.47±0.22	0.50±0.28	0.50±0.30	0.37±0.27	0.43±0.28
Pragmatic inferences						
	Experiment 1			Experiment 2		
	Retrieval	Restudy	Total	Active recognition	Passive recognition	Total
Feedback	0.14±0.09	0.33±0.19	0.23±0.17	0.22±0.16	0.40±0.25	0.31±0.22
No feedback	0.67±0.14	0.63±0.14	0.65±0.14	0.64±0.13	0.65±0.17	0.65±0.15
Total	0.40±0.29	0.48±0.22	0.44±0.26	0.43±0.25	0.52±0.25	0.47±0.25
Intrusions						
	Experiment 1			Experiment 2		
	Retrieval	Restudy	Total	Active recognition	Passive recognition	Total
Feedback	0.01±0.02	0.03±0.03	0.02±0.03	0.05±0.05	0.08±0.07	0.07±0.06
No feedback	0.06±0.07	0.03±0.04	0.05±0.06	0.08±0.08	0.08±0.07	0.08±0.08
Total	0.04±0.05	0.03±0.03	0.04±0.05	0.06±0.07	0.08±0.07	0.07±0.07
Omissions						
	Experiment 1			Experiment 2		
	Retrieval	Restudy	Total	Active recognition	Passive recognition	Total
Feedback	0.02±0.03	0.03±0.08	0.02±0.06	0.01±0.03	0.01±0.03	0.01±0.03
No feedback	0.04±0.07	0.02±0.04	0.03±0.05	0.00±0.02	0.03±0.08	0.02±0.06
Total	0.03±0.05	0.02±0.06	0.03±0.06	0.01±0.03	0.02±0.06	0.01±0.05

Descriptive statistics (mean ± standard deviations) of the four response types assessed at the final cued-recall test as a function of the experimental conditions: Feedback vs. no feedback and learning condition (retrieval vs. restudy for Experiment 1, and active vs. passive recognition for Experiment 2).

<https://doi.org/10.1371/journal.pone.0272427.t003>

For intrusions, we did not observe a main effect of retrieval task ($F(1,116) = 1.26, p = 0.26, \eta^2_p = 0.01$), but there was a main effect of feedback ($F(1,116) = 11.33, p < 0.01, \eta^2_p = 0.09$) and an interaction ($F(1,116) = 7.10, p < 0.01, \eta^2_p = 0.06$). This indicates that while for the restudy group there was no difference in the proportion of intrusions in the feedback and no feedback conditions ($F(1,58) = 0.37, p = 0.55, \eta^2_p < 0.01$), for the retrieval group there were fewer intrusions in the feedback compared to the no feedback condition ($F(1,58) = 13.66, p < 0.01, \eta^2_p = 0.19$).

Lastly, for omissions, we did not observe a main effect of retrieval task ($F(1,116) = 0.022, p = 0.64, \eta^2_p < 0.01$) or feedback ($F(1,116) = 0.32, p = 0.57, \eta^2_p < 0.01$), but there was an interaction between the factors ($F(1,116) = 4.08, p = 0.05, \eta^2_p = 0.03$). While no difference was observed between feedback and no feedback for the restudy conditions ($F(1,58) = 0.88, p = 0.35, \eta^2_p = 0.01$), presenting feedback after retrieval led to fewer omissions than having no feedback ($F(1,58) = 4.14, p = 0.05, \eta^2_p = 0.07$).

Persistence of correct responses and error correction. To further investigate the effects of the experimental manipulations and to directly test our hypotheses, we analyzed memory performance at the final cued-recall test as a function of participants responses during the intermediate phase. To that purpose, we calculated two different memory indexes (see Fig 1) and introduced them as dependent variables in a 2×2 between subjects ANOVA with learning condition (retrieval vs. restudy) and feedback (present vs. absent) as factors.

To assess the persistence of correct responses, we considered the sentences that had been recalled/restudied correctly at the intermediate phase and inspected if they continued to be correctly recalled in the final test. We calculated this index through the sum of correct responses in the final test that had been correctly recalled/restudied in the intermediate phase divided by the total number of sentences correctly recalled/presented in the intermediate phase. Results showed that participants that performed the retrieval task had a higher rate of persistence of correct responses ($M = 0.83 \pm 0.22$) compared to those who restudied the sentences ($M = 0.53 \pm 0.21$; $F(1,116) = 66.47, p < 0.01, \eta^2_p = 0.07$), replicating the testing effect. Receiving feedback also led to a greater persistency of the correct responses ($M = 0.74 \pm 0.25$) than not

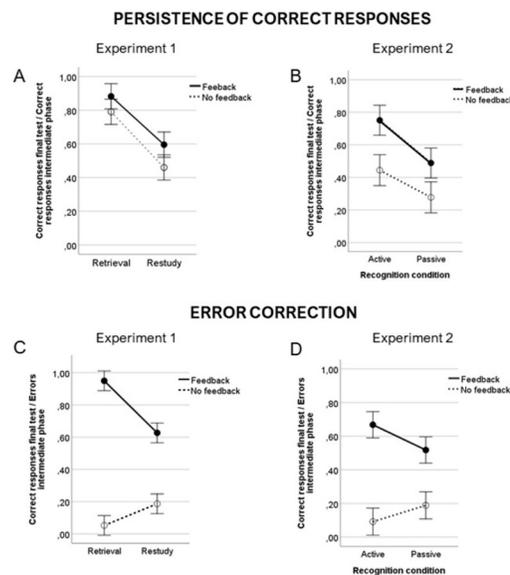


Fig 1. Memory performance. Memory indexes calculated from performance at the final cued-recall test as a function of performance in the intermediate phase. A, B: Persistence of correct responses; C, D: Error correction. Left panels depict data for Experiment 1 and right panels represent performance in Experiment 2.

<https://doi.org/10.1371/journal.pone.0272427.g001>

receiving feedback ($M = 0.62 \pm 0.27$, $F(1,116) = 9.00$, $p < 0.01$, $\eta^2_p = 0.07$). We did not observe an interaction between the factors ($F(1,116) = 0.33$, $p = 0.57$, $\eta^2_p < 0.01$, see Fig 1A).

The most interesting measure of our analysis was the assessment of error correction in those sentences that, at the intermediate phase, were incorrectly recalled/restudied but at the final cued-recall test were correctly recalled. We computed this index as the sum of correct responses in the final test phase that had previously been pragmatic inferences errors in the intermediate phase divided by the total number of pragmatic inferences in the intermediate phase. Results showed that, compared to restudy ($M = 0.41 \pm 0.29$), performing a retrieval task during the intermediate phase led to a higher proportion of error correction ($M = 0.50 \pm 0.47$, $F(1,116) = 9.34$, $p < 0.01$, $\eta^2_p = 0.07$). The presence of feedback led to a much larger proportion of corrected errors ($M = 0.79 \pm 0.27$) compared to the absence of feedback ($M = 0.12 \pm 0.13$, $F(1,116) = 468.60$, $p < 0.01$, $\eta^2_p = 0.80$). But most interestingly, the interaction between the factors ($F(1,116) = 54.74$, $p < 0.01$, $\eta^2_p = 0.32$) shows that providing feedback after retrieval resulted in higher error correction (Feedback: $M = 0.95 \pm 0.19$; No feedback: $M = 0.05 \pm 0.06$; $F(1,58) = 599.59$, $p < 0.01$, $\eta^2_p = 0.91$) compared to restudy (Feedback: $M = 0.63 \pm 0.23$; No feedback: $M = 0.19 \pm 0.14$; $F(1,58) = 78.30$, $p < 0.01$, $\eta^2_p = 0.57$, see Fig 1C), suggesting that feedback was more effective in correcting errors after retrieval than after restudy.

Discussion

Our results showed that retrieval and the presentation of corrective feedback were advantageous tools for learning, even when retrieval was unsuccessful. Compared to restudy, retrieval promoted learning both by perpetuating correct responses and increasing correction of pragmatic inference errors. In the absence of corrective feedback, retrieval led to a higher proportion of false memories compared to restudy, as had already been demonstrated by McDermott [52]. But when feedback was provided, retrieval strengthened error correction in contrast to restudy. The consistent interaction found between learning conditions and feedback suggests that the administration of corrective feedback was more effective when it was presented after retrieval than after restudying practices. These results go in line with an errorful approach to learning and the proposal that after retrieval, memory becomes more amenable to change, promoting memory updating.

However, caution is warranted before further interpretation of our findings since this design had some limitations. In the retrieval condition several types of errors were generated during the intermediate phase (i.e., pragmatic inferences, intrusions, and omissions), while only one type of error was presented in the restudy condition (i.e., pragmatic inferences). Additionally, while at the intermediate phase, 50% of the sentences presented in the restudy condition contained a pragmatic inference error, in the retrieval condition participants committed a pragmatic inference error in 63% of the sentences, leading to an uneven proportion of errors in the two groups, which in turn may have impacted performance at the final cued-recall test. Moreover, in the retrieval condition participants generated false information, whereas those at restudy were presented to it during the experiment which might be a cause for source confusion [31]. Specifically, for those in the retrieval condition, sentences at encoding and as feedback were presented in the experiment setting, coming from an external source, while the pragmatic inferences committed during the intermediate phase were generated internally, coming from an internal source. Yet, for those in the restudy condition, all information (encoding, feedback, and the pragmatic inferences presented during intermediate restudy) was presented in the experiment setting and all came from a shared external source. So, at the final cued-recall test, it is likely that it was more difficult for participants in the restudy condition to identify the source of their memories, than it was for those in retrieval.

To address these limitations, we conducted a second experiment in which we increased the methodological control of the key experimental manipulations during the intermediate phase. To control for the different response types, proportion of responses and the source of the information, we employed a recognition paradigm at the intermediate phase. Participants could engage in one out of two recognition conditions: active or passive recognition. The active recognition condition was analogous to the previous retrieval condition, in which participants had to take a memory test, where they had to decide whether the sentences being presented were correct (like encoding) or incorrect (different from encoding). In contrast, passive recognition was analogous to the previous restudy condition, in which participants were exposed to the same sentences than those at active recognition did but, critically, instead of having to provide an answer, they were presented with the responses given by another participant (from the active condition) and they were instructed only to read the information (i.e., the sentences and the answers given). Therefore, an adapted yoked design was implemented, where the responses of a participant in the active recognition group were presented to another participant in the passive recognition condition, thus forming yoked pairs, and guaranteeing that across the two learning conditions participants were exposed to the same information, in the same proportion, and in the same experimental setting. Presence and absence of corrective feedback was manipulated as in Experiment 1. Following the testing effect and the errorful learning literatures, we predicted that when followed by feedback, the active recognition would lead to better memory performance (i.e., more correct responses) and a higher rate of corrected errors in the final cued-recall test compared to the passive recognition task.

Experiment 2

Methods

Participants. One hundred and twenty-eight university students from the University of Lisbon participated in this study and received 10 vouchers as a reward for their time. Because of the COVID-19 pandemic, data collection was conducted online, so participants were asked to evaluate their commitment to the experiment and to report any relevant incidents (see procedure below). Eight participants had to be excluded for the following reasons: 2 reported being interrupted during the experiment, 2 did not have a yoked pair, and 4 reported low levels of attention. Thus, the final sample size included 120 young adults ($M_{\text{age}} = 27.47 \pm 8.92$; 65 females). Participants were randomly assigned to one out of four conditions from the crossing of the factors, recognition task and feedback: active recognition with feedback ($n = 31$), active recognition without feedback ($n = 29$), passive recognition with feedback ($n = 31$), passive recognition without feedback ($n = 29$). Similarly to Experiment 1, all participants gave their informed consent in the online link of the study. All procedures were approved by the Ethics Committee of the Faculty of Psychology of the University of Lisbon.

Materials. Thirty-two pragmatic implication sentences in Portuguese were selected from those adapted by Carneiro et al. [47], and used in Maraver et al. [50] (see [S1 Dataset](#)). To improve the experimental control of the materials and to adapt them to a recognition test, sentences were counterbalanced as a function of the format of the sentence presented at encoding (critical vs. filler) and the format presented at the intermediate phase (match vs. mismatch), generating four different sets of the materials. Each participant saw only one set, with the four sets being counterbalanced between participants. In each set, 16 critical sentences were presented at encoding intermixed with 16 fillers. The critical sentences were presented in their original format to generate a pragmatic inference (i.e., “*The karate champion hit the cinder block*”). In contrast, filler sentences (e.g., “*The baby cried all night*”) were created by modifying the critical sentences from another set (e.g., “*The baby stayed awake all night*”), such that they

were similar to the critical sentences in all respects, except that the pragmatic implication was explicitly presented in the sentence. By mixing critical and filler sentences we minimized participants' attention to the inferences that the critical sentences induce and hence prevented participants' awareness of the experimental manipulation. Afterwards, at the intermediate phase, the presented sentences either matched or mismatched the format of the sentences presented at encoding. A sentence was a match when presented at the intermediate phase in the same format as it was presented at encoding (i.e., "The karate champion hit the cinder block"), while a mismatch was a sentence presented at the intermediate phase in a different format from encoding (i.e., "The karate champion broke the cinder block"). This complex counterbalance of the material was developed to set a high level of task difficulty, encourage errors to be made, and guarantee that participants would not predict the experimental manipulations. However, because our interest was on the correction of memory errors arising from pragmatic inferences, our analyses focused on the 16 original sentences in which inference generation was promoted.

Design. Experiment 2 followed a 2 (recognition: active vs. passive) × 2 (feedback: present vs. absent) design, with both variables manipulated between-participants.

Procedure. The experiment was conducted online, and participants completed the experiment in their own computers. They were instructed to complete the study in a calm and quiet environment without distractions. Like Experiment 1, the study was programmed and run through the online platform Qualtrics (Qualtrics, Provo, UT).

The structure of the experiment was similar to Experiment 1, except that the learning conditions at the intermediate phase consisted of an active or passive recognition paradigm, instead of a retrieval or restudy tasks (see Table 4). Therefore, during the encoding phase, participants were instructed to read and memorize the 32 sentences (16 critical and 16 fillers) presented at the center of the screen, in black font, for 4.5 seconds. After each sentence they solved a simple math operation for 5 seconds. Sentences were presented in a random order. The key manipulations of the two experimental factors occurred at the intermediate phase. During this phase, in the active recognition condition, participants performed a typical recognition test in which they evaluated whether the sentences presented (in blue font and in a random order) were correct (the same as presented at encoding) or incorrect (different from the sentence presented at encoding). Participants were exposed to each sentence for a minimum of 6.5 seconds and had a maximum of 10 seconds to provide their answer. In the passive recognition condition, participants were exposed to the same sentences (presented in a blue font and in a random order) than those in the active recognition condition but were simultaneously shown the responses given by a previous active participant. Each sentence and corresponding answer were presented

Table 4. Procedure scheme of Experiment 2.

Encoding phase				
<i>The baby stayed awake all night</i>				
$5 + 4 = 12$				
Intermediate phase				
Task	Active recognition	<i>The baby cried all night</i> Correct / Incorrect? (Participant gives the response)	Feedback	<i>The baby stayed awake all night</i>
			No feedback	$23 - 7 =$
	Passive recognition	<i>The baby cried all night</i> Correct / Incorrect (Participant reads the response)	Feedback	<i>The baby stayed awake all night</i>
			No feedback	$36 + 6 =$
Final cued-recall test				
<i>The baby ____ all night</i>				

<https://doi.org/10.1371/journal.pone.0272427.t004>

for 6.5 seconds, and participants' task was to read them. Similar to Experiment 1, after giving a response or reading an answer to each sentence at the intermediate phase, the presentation of feedback was manipulated. Half of the participants in the active and passive conditions received corrective feedback that consisted in the presentation of the sentence in the same format as encoding for 4.5 seconds in a black font. The other half did not receive feedback, and after each sentence had 4.5 seconds to solve simple math operations. At the end of the intermediate phase, and after a 5-minutes distractor task, all participants performed a final cued-recall memory test in which they were instructed to complete the missing information in the sentences presented, like in Experiment 1. They were given 60 seconds to complete each sentence.

At the end of the study, participants responded to a brief self-report questionnaire about their attention and environmental conditions during the study (similar to Experiment 2 in [50]). Participants were asked to rate their level of attention, the quality of their data, and had a blank space to provide any comments in relation to their performance. After completing this questionnaire, participants were thanked and rewarded for their time.

Statistical analyses. Statistical analysis for Experiment 2 focused on the same dependent variables as in Experiment 1. First, we compared recognition memory at the intermediate phase between participants in the feedback and no feedback conditions, through independent samples *t*-tests. Second, we analyzed the different types of response (proportion of correct responses, pragmatic inference errors, intrusions, and omissions) at the final cued-recall test. Lastly, we assessed the memory indexes reflecting persistence of correct responses and error correction. For each dependent variable we conducted an ANOVA with recognition (active vs. passive) and feedback (present vs. absent) as between-subjects factors.

Results

Memory performance at the intermediate phase. At the intermediate phase, the overall accuracy in the active recognition test did not differ between participants who received feedback ($M = 0.63 \pm 0.12$) and those who did not ($M = 0.63 \pm 0.08$; $t(53.93) = 0.14$, $p = 0.88$, 95% CI [-0.05–0.06]).

Memory performance at the final cued-recall test. Descriptive statistics of the different response types assessed at the final cued-recall test as a function of the experimental conditions can be found in Table 3. Results of the between subjects ANOVA showed that participants in the active recognition group had a higher proportion of correct responses than those who completed the passive recognition task at the intermediate phase ($F(1,116) = 10.91$, $p < 0.01$, $\eta^2_p = 0.09$). Receiving feedback also generated more correct responses compared to receiving no feedback ($F(1,116) = 89.68$, $p < 0.01$, $\eta^2_p = 0.44$). The significant interaction between the factors ($F(1,116) = 4.42$, $p = 0.04$, $\eta^2_p = 0.04$) revealed that the beneficial effect of feedback was larger for the active recognition ($F(1,58) = 95.25$, $p < 0.01$, $\eta^2_p = 0.62$), than for the passive recognition group ($F(1,58) = 20.93$, $p < 0.01$, $\eta^2_p = 0.26$).

For pragmatic inference errors we observed a similar pattern of results. Participants in the active recognition condition had a lower proportion of pragmatic inference errors compared to those in the passive recognition condition ($F(1,116) = 7.89$, $p < 0.01$, $\eta^2_p = 0.06$); and receiving feedback led to a decrease in the proportion of pragmatic inferences committed compared to receiving no feedback ($F(1,116) = 99.99$, $p < 0.01$, $\eta^2_p = 0.46$). The interaction ($F(1,116) = 5.87$, $p < 0.02$, $\eta^2_p = 0.05$) revealed that the beneficial effect of feedback in reducing the proportion of pragmatic inference errors was larger for the active recognition ($F(1,58) = 122.42$, $p < 0.01$, $\eta^2_p = 0.68$) than for the passive recognition group ($F(1,58) = 20.95$, $p < 0.01$, $\eta^2_p = 0.26$).

No significant effects emerged for intrusion errors (effect of recognition group ($F(1,116) = 1.18$, $p = 0.28$, $\eta^2_p = 0.01$); effect of feedback ($F(1,116) = 0.73$, $p = 0.39$, $\eta^2_p = 0.01$); interaction

($F(1,116) = 0.58, p = 0.45, \eta^2_p < 0.01$) or for omissions (effect of recognition group $F(1,116) = 3.65, p = 0.06, \eta^2_p = 0.03$); effect of feedback ($F(1,116) = 1.08, p = 0.30, \eta^2_p = 0.01$); interaction ($F(1,116) = 2.13, p = 0.15, \eta^2_p = 0.02$), see Table 3).

Persistence of correct responses and error correction. Similar to Experiment 1, we analyzed memory performance at the final cued-recall test as a function of whether the responses were correct or incorrect at the intermediate phase, in order to analyze the persistence of correct responses and error correction.

To analyze the persistence of correct responses, we considered the correct responses at the final cued-recall test that had already been correctly responded/presented at the intermediate phase, divided by the total of sentences correctly recognized/presented at the intermediate phase. Results showed that performing the active recognition test led to a higher persistence of the correct responses ($M = 0.60 \pm 0.28$) compared to a passive recognition task ($M = 0.39 \pm 0.29; F(1,116) = 20.66, p < 0.01, \eta^2_p = 0.15$). Likewise, receiving feedback ($M = 0.62 \pm 0.30$) resulted in more persisting correct responses compared to receiving no feedback ($M = 0.36 \pm 0.26, F(1,116) = 30.00, p < 0.01, \eta^2_p = 0.21$). No significant interaction between the factors was observed ($F(1,116) = 1.02, p = 0.31, \eta^2_p = 0.01$, see Fig 1D).

To assess the correction of errors, we considered the number of sentences that were recognized incorrectly in the intermediate phase but correctly recalled in the final test and divided them by the total number of errors made at the intermediate phase. While no effect of the recognition group was found ($F(1,116) = 0.44, p = 0.51, \eta^2_p = 0.08$), we observed that receiving feedback ($M = 0.59 \pm 0.27$) led to a higher rate of error correction than receiving no feedback ($M = 0.14 \pm 0.32; F(1,116) = 127.60, p < 0.01, \eta^2_p = 0.52$). A significant interaction between the factors ($F(1,116) = 9.49, p < 0.01, \eta^2_p = 0.07$) showed that error correction after feedback compared to no feedback was greater for those in the active recognition condition (Feedback: $M = 0.67 \pm 0.22$; No feedback: $M = 0.09 \pm 0.10; F(1,58) = 165.92, p < 0.01, \eta^2_p = 0.74$), than for those in the passive recognition condition (Feedback: $M = 0.52 \pm 0.30$; No feedback: $M = 0.19 \pm 0.20; F(1,58) = 24.50, p < 0.01, \eta^2_p = 0.30$, see Fig 1F).

Discussion. The results extended the findings of Experiment 1 demonstrating the benefit of retrieval in the correction of errors compared to the passive processing of information. Increased methodological control was implemented in this study, since the comparison between conditions relied on a matched proportion of errors between the active and passive recognition tasks as a result of the yoked design. Besides, we replicated the beneficial effect of feedback in the correction of errors when it is presented after an active rather than a passive recognition task. Finally, correct responses persisted to a greater extent for active recognition compared to passive, supporting a consistent testing effect for pragmatic inferences. Taken together, when followed by corrective feedback, unsuccessful testing attempts are useful, and we highlight the consistency of this pattern of results in a highly controlled setting.

General discussion

In a learning task, we have observed that false memory errors were more easily corrected following the retrieval of a wrong response, than when people were simply exposed to errors. Retrieval seemed to be a more favorable environment to correct errors than the passive processing of information by promoting a better incorporation of the corrective feedback. We have replicated this benefit across two independent studies, both using a cued-recall (Experiment 1) and a recognition paradigm (Experiment 2).

Throughout the two experiments, we have found that responses correctly recalled during the intermediate phase continued to be correctly retrieved at the final memory test, confirming a robust testing effect [4–6]. To the extent of our knowledge, no previous studies have

extended the testing effect to pragmatic inference sentences, and we provide additional evidence in support of active memory tests as successful learning events for knowledge acquisition [19]. It is also worth noting that this effect was found even though memory was assessed in an immediate test. The testing effect has been typically found for long term learning, being generally weaker on immediate tests [53], highlighting the robustness of the testing effect.

Besides, as expected, accurate recall was higher when participants received feedback than no feedback, regardless of the learning condition. This corroborates the view that when participants make an error, it is crucial that the correct answer is available to guarantee its processing and integration into memory, rather than simply indicating if the responses were right or wrong [3]. The memory benefit observed following feedback (compared to a no feedback condition) suggests that participants paid attention to the information given, understood that they had made an error, and, as a result, memory was updated, and performance improved. This is consistent with a large body of previous findings [3,8,53], and with the observation that surprising feedback improved memory for the content and its source when presented in response to correct guesses or errors made with high confidence [54]. In our study, we have not compared different features of feedback but we have chosen the parameters that have been proven to be most effective for the correction of false memories, such as explicitly providing the correct answer [18,55], being distinctive [8,17], and appearing immediately after the related false memory [44]. Moreover, future studies should further extend these findings with the current paradigm exploring how the confidence degree of the retrieved errors and the monitoring of the error source modulates its correction. Overall, we provide additional support for the enhanced processing of feedback as a key consequence of error generation [24].

Across two experiments, our main finding was that in the learning process, even when errors occur, retrieving an error benefits subsequent learning compared to the passive processing of the error, as long as corrective feedback is provided. The interaction between the learning conditions and feedback indicates that feedback is more effective when presented immediately after error retrieval than after the passive processing of information. According to the interference account and errorless learning perspective, errors should be avoided during practice to prevent from perpetuating wrongful responses [14]. In contrast, the errorful learning hypothesis states that errors are beneficial for learning as long as feedback is provided, because incorrect responses draw attention to their mismatch from corrective feedback and promote the encoding and incorporation of feedback [15]. Our study was not designed to directly compare the errorless and the errorful learning perspectives, since our focus was on conditions in which errors were present. However, our findings are in line with an errorful approach to learning, since retrieving errors followed by feedback improved later memory performance compared to when no responses were permitted. Although the errorless learning method seems to be helpful for individuals with learning disabilities or amnesia [56, but see 57], our findings build on work suggesting that this does not seem to be the case for healthy adults [58], not even when false memory errors are committed during practice as in the case of the current experiments.

Our results are consistent with previous studies demonstrating that attempting to retrieve information, by itself, enhances learning [19,26]. According to the reconsolidation theory, upon retrieval, memory traces become more labile and prone to modification. Thus, retrieval may facilitate memory updating, but in some cases, retrieval can come with the cost of impairing the learning of new information [59,60]. The Memory Updating After Retrieval (MUAR) framework accounts for the beneficial as well as impairing effects of retrieval on subsequent learning and provides a solid theoretical framework for the interpretation of our findings [22]. In the current studies, we have observed that when relevant information for retrieval (i.e., corrective feedback) is introduced just after a retrieval attempt, there is a ready updating of

knowledge. That is, memory malleability after retrieval allows for new related information to be incorporated into the retrieved traces [22]. A previous study using DRM lists has shown that new information presented after retrieval can increase (when false) or reduce (when correct) false memories, and may potentiate error correction, supporting the predictions of the MUAR framework [44]. Our results added evidence to this interpretation since retrieval, compared to the passive processing of information, promoted memory updating through the integration of corrective feedback, as new information aligned with the goals of the retrieval event. The false memory errors generated from pragmatic inferences share close semantic overlap with the correct information, which likely make these errors particularly difficult to notice. Given that pragmatic inference errors stem from semantic association, the benefit of wrongful retrieval to the incorporation of feedback could be explained by a better dissociation between semantic (wrong) and episodic (correct) components, that resulted in a higher rate of correct responses and a lower proportion of pragmatic inferences. Our findings add up to previous research showing that the learning benefit of error generation goes beyond the semantic mediation [28], and we provide additional evidence for active episodic recollection as the underlying mechanism that allows memory updating to learn from errors.

To deepen the study of the mechanisms for error correction, we introduced a manipulation in our second experiment that allowed the comparison between different sources of retrieval. Previous research has shown that being the agent of a retrieved error facilitates memory for the correct answer compared to witnessing other person's mistake or to a condition when no errors are made [61]. In the current experiments, we used two control conditions (restudy and passive recognition) where errors made by someone else were presented. The detection of an error made by someone else might trigger some degree of unintentional retrieval given that participants were not explicitly told whether or not the response was an error [62,63]. However, the recent study by Carneiro et al., [44]—using a similar restudy condition as our Experiment 1—did not observe differences in the subsequent retrieval of correct and incorrect sentences that were restudied. Moreover, previous studies have been able to distinguish the retrieval mode from involuntary uses of episodic memory using electrophysiological techniques [64–66], which in addition to Tulving's proposal that retrieval mode is only engaged through a conscious experience of recollection [41], led us to conclude that it is unlikely that our restudy and passive recognition conditions captured explicit retrieval. Importantly, Metcalfe & Xu suggest that it is the process of overt retrieval of a self-generated error what renders memory open to modification [61]. Our results provide further evidence to this hypothesis since we have observed that being an agent of a retrieved error and receiving feedback facilitated memory for the correct answer more than witnessing an error committed by someone else and receiving feedback afterwards. In our case this comparison was made between-groups, so it still remains unexplored whether the benefits of self-generating an error compared to seeing someone else's mistake and then receiving feedback are maintained within the same participants. The comparison between errors from different sources provides an additional argument in support of the benefits of retrieval for learning, since it shows that it is not only the mere exposure to an error followed by its correction that promotes learning. Instead, what underlies the full benefit of learning from errors is the retrieval of an own error which renders the exposed memory trace vulnerable to modification by corrective feedback [3,61].

The current study is not exempt of limitations. Importantly, the assessment of memory performance was done immediately after, representing a measure of short-term learning. Future studies should explore whether this benefit of retrieval for the correction of errors still holds after long delays and are beneficial for long-term learning. False memories represent a particularly interesting type of error for the study of error correction, since they are normally held with high levels of confidence [67–69]. Fazio & Marsh observed a hypercorrection effect for

pragmatic inferences so that errors made with higher levels of confidence were easier to correct than low confidence errors [70]. In the current study, the levels of confidence were not assessed, so it remains unexplored whether this hypercorrection effect is replicated across different learning conditions [69].

Caution is warranted when generalizing our findings to the classroom given that our materials were developed to generate a high proportion of errors and differ from educational environments where errors of different nature coexist (e.g., omissions in addition to retrieval failures). However, our results show that testing is a powerful mean to improve learning, not just assessing it. Making errors in tests and during the process of learning derives direct benefits to learners, when followed by feedback that offers the correct response. Thus, we support the use of learning strategies where errors are tolerated, and students engage in active, exploratory, and generative learning. Going a step forward, recent research suggests that learning can benefit from errors even when participants are aware of the correct answer and deliberately decide to commit an error, in contrast to avoiding giving incorrect answers, an effect called the *derring effect* with important implications for educational practice given that students often appear to be unaware of the benefits of deliberate erring [71]. In comparison with approaches that stress error avoidance, making training more challenging by allowing false starts and errors followed by feedback, discussion, and correction may ultimately lead to better and more flexible transfer of skills to later critical situations.

Conclusion

Our results suggest that the act of retrieval constitutes a valuable learning event. Even when the information is wrong, retrieval does more than correcting wrong information, and it is through the effortful engagement in intentional explicit retrieval what renders memory in a sufficient malleable state to allow the incorporation of feedback.

We propose that the benefits of retrieval go beyond the testing effect given that retrieval also facilitated false memory correction. Our results support the use of tests and active retrieval strategies in learning, showing we can learn from practice even when errors are generated.

Supporting information

S1 Dataset. The datasets presented in this study can be found in the following Open Science Framework project identifier: DOI [10.17605/OSF.IO/JVU4K](https://doi.org/10.17605/OSF.IO/JVU4K).
(DOCX)

Acknowledgments

We especially thank Nanci Lemos, André Vaz, Filipa Gonçalves and Joana Almeida for their help in data collection.

Author Contributions

Conceptualization: María J. Maraver, Ana Lapa, Leonel Garcia-Marques, Paula Carneiro, Ana Raposo.

Data curation: María J. Maraver, Ana Lapa.

Formal analysis: María J. Maraver, Ana Lapa, Paula Carneiro, Ana Raposo.

Funding acquisition: Ana Lapa, Paula Carneiro, Ana Raposo.

Investigation: María J. Maraver, Ana Lapa, Leonel Garcia-Marques, Paula Carneiro, Ana Raposo.

Methodology: María J. Maraver, Ana Lapa, Leonel Garcia-Marques, Paula Carneiro, Ana Raposo.

Project administration: María J. Maraver, Ana Lapa, Ana Raposo.

Resources: María J. Maraver, Paula Carneiro, Ana Raposo.

Software: María J. Maraver, Ana Lapa.

Supervision: María J. Maraver, Ana Lapa, Leonel Garcia-Marques, Ana Raposo.

Validation: María J. Maraver, Ana Lapa, Leonel Garcia-Marques, Ana Raposo.

Visualization: María J. Maraver, Ana Lapa, Leonel Garcia-Marques, Ana Raposo.

Writing – original draft: María J. Maraver.

Writing – review & editing: María J. Maraver, Ana Lapa, Leonel Garcia-Marques, Ana Raposo.

References

1. Nagai Y. Predictive learning: Its key role in early cognitive development. *Philos Trans R Soc B Biol Sci*. 2019;374. <https://doi.org/10.1098/rstb.2018.0030> PMID: 30852990
2. Rescorla R, Wagner A. A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In: Black A, Prokasy W. *Classical Conditioning II: Current Research and Theory*. New York: Appleton Century Crofts; 1972. pp. 64–99.
3. Metcalfe J. Learning from Errors. *Annu Rev Psychol*. 2017; 68: 465–489. <https://doi.org/10.1146/annurev-psych-010416-044022> PMID: 27648988
4. Karpicke JD, Roediger HL. The Critical Importance of Retrieval for Learning. *Science (80-)*. 2008; 319: 966–968. <https://doi.org/10.1126/science.1152408> PMID: 18276894
5. Roediger HL, Karpicke JD. Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychol Sci*. 2006; 17: 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x> PMID: 16507066
6. Pyc MA, Agarwal PK, Roediger HL III. Test-enhanced learning. Applying science of learning in education: Infusing psychological science into the curriculum. Washington, DC, US: Society for the Teaching of Psychology; 2014. pp. 78–90.
7. Pastötter B, Bäuml K-H. Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology*. 2014. p. 286. <https://doi.org/10.3389/fpsyg.2014.00286> PMID: 24772101
8. Kornell N, Metcalfe J. The effects of memory retrieval, errors and feedback on learning. Applying science of learning in education: Infusing psychological science into the curriculum. Society for the Teaching of Psychology; 2014. pp. 225–251.
9. Potts R, Shanks DR. The benefit of generating errors during learning. *Journal of Experimental Psychology: General*. 2014. pp. 644–667. <https://doi.org/10.1037/a0033194> PMID: 23815457
10. Barnes JM, Underwood BJ. “Fate” of first-list associations in transfer theory. *J Exp Psychol*. 1959; 58: 97–105. <https://doi.org/10.1037/h0047507> PMID: 13796886
11. Bertsch S, Pesta BJ, Wiscott R, McDaniel MA. The generation effect: A meta-analytic review. *Memory & Cognition*. 2007. pp. 201–210. <https://doi.org/10.3758/bf03193441> PMID: 17645161
12. Bandura A. *Social foundations of thought and action: A social cognitive theory*. Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ, US: Prentice-Hall, Inc; 1986.
13. Skinner BF. *The technology of teaching*. The technology of teaching. East Norwalk, CT, US: Appleton-Century-Crofts; 1968.
14. Terrace HS. Errorless transfer of a discrimination across two continua. *J Exp Anal Behav*. 1963; 6: 223–232. <https://doi.org/10.1901/jeab.1963.6-223> PMID: 13980669
15. Izawa C. Function of test trials in paired-associate learning. *J Exp Psychol*. 1967; 75: 194–209. <https://doi.org/10.1037/h0024971> PMID: 6062960

16. Izawa C. Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *J Exp Psychol.* 1970; 83: 340–344. <https://doi.org/10.1037/h0028541>
17. Anderson RC, Kulhavy RW, Andre T. Conditions under which feedback facilitates learning from programmed lessons. *J Educ Psychol.* 1972; 63: 186–188. <https://doi.org/10.1037/h0032653>
18. Pashler H, Cepeda NJ, Wixted JT, Rohrer D. When does feedback facilitate learning of words? *J Exp Psychol Learn Mem Cogn.* 2005; 31: 3–8. <https://doi.org/10.1037/0278-7393.31.1.3> PMID: 15641900
19. Richland LE, Kornell N, Kao LS. The Pretesting Effect: Do Unsuccessful Retrieval Attempts Enhance Learning? *J Exp Psychol Appl.* 2009; 15: 243–257. <https://doi.org/10.1037/a0016496> PMID: 19751074
20. Forcato C, Burgos VL, Argibay PF, Molina VA, Pedreira ME, Maldonado H. Reconsolidation of declarative memory in humans. *Learning & Memory.* 2007. pp. 295–303. <https://doi.org/10.1101/lm.486107> PMID: 17522018
21. Hubbach A, Gomez R, Hardt O, Nadel L. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learn Mem.* 2007; 14: 47–53. <https://doi.org/10.1101/lm.365707> PMID: 17202429
22. Finn B. A Framework of Episodic Updating: An Account of Memory Updating After Retrieval. In: Ross BHBT-P of L and M, editor. Academic Press; 2017. pp. 173–211. <https://doi.org/10.1016/bs.plm.2017.03.006>.
23. Kane JH, Anderson RC. Depth of processing and interference effects in the learning and remembering of sentences. *J Educ Psychol.* 1978; 70: 626–635. <https://doi.org/10.1037/0022-0663.70.4.626>
24. Potts R, Davies G, Shanks DR. The benefit of generating errors during learning: What is the locus of the effect? *J Exp Psychol Learn Mem Cogn.* 2019; 45: 1023–1041. <https://doi.org/10.1037/xlm0000637> PMID: 30024254
25. Slamecka NJ, Fevreiski J. The generation effect when generation fails. *J Verbal Learn Verbal Behav.* 1983; 22: 153–163. [https://doi.org/10.1016/S0022-5371\(83\)90112-3](https://doi.org/10.1016/S0022-5371(83)90112-3)
26. Kornell N, Hays MJ, Bjork RA. Unsuccessful Retrieval Attempts Enhance Subsequent Learning. *J Exp Psychol Learn Mem Cogn.* 2009; 35: 989–998. <https://doi.org/10.1037/a0015729> PMID: 19586265
27. Carneiro P, Lapa A, Finn B. The effect of unsuccessful retrieval on children's subsequent learning. *J Exp Child Psychol.* 2018; 166: 400–420. <https://doi.org/10.1016/j.jecp.2017.09.010> PMID: 29032207
28. Metcalfe J, Huelser BJ. Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia.* 2020; 138: 107296. <https://doi.org/10.1016/j.neuropsychologia.2019.107296> PMID: 31811845
29. Kornell N, Klein PJ, Rawson KA. Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 2015. pp. 283–294. <https://doi.org/10.1037/a0037850> PMID: 25329079
30. Huelser BJ, Metcalfe J. Making related errors facilitates learning, but learners do not know it. *Mem Cognit.* 2012; 40: 514–527. <https://doi.org/10.3758/s13421-011-0167-z> PMID: 22161209
31. Grimaldi PJ, Karpicke JD. When and why do retrieval attempts enhance subsequent encoding? *Mem Cognit.* 2012; 40: 505–513. <https://doi.org/10.3758/s13421-011-0174-0> PMID: 22238214
32. Knight JB, Hunter Ball B, Brewer GA, DeWitt MR, Marsh RL. Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *J Mem Lang.* 2012; 66: 731–746. <https://doi.org/10.1016/j.jml.2011.12.008>.
33. Kang SHK, Pashler H, Cepeda NJ, Rohrer D, Carpenter SK, Mozer MC. Does incorrect guessing impair fact learning? *J Educ Psychol.* 2011; 103: 48–59. <https://doi.org/10.1037/a0021977>
34. Pyc MA, Rawson KA. Why testing improves memory: Mediator effectiveness hypothesis. *Science.* 2010. p. 335. <https://doi.org/10.1126/science.1191465> PMID: 20947756
35. Cyr AA, Anderson ND. Learning from your mistakes: does it matter if you're out in left foot, I mean field? *Memory.* 2018; 26: 1281–1290. <https://doi.org/10.1080/09658211.2018.1464189> PMID: 29659332
36. Vaughn KE, Rawson KA. Diagnosing criterion-level effects on memory: what aspects of memory are enhanced by repeated retrieval? *Psychol Sci.* 2011; 22: 1127–1131. <https://doi.org/10.1177/0956797611417724> PMID: 21813798
37. Kornell N. Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 2014. pp. 106–114. <https://doi.org/10.1037/a0033699> PMID: 23855547
38. Bridger EK, Mecklinger A. Errorful and errorless learning: The impact of cue-target constraint in learning from errors. *Mem Cognit.* 2014; 42: 898–911. <https://doi.org/10.3758/s13421-014-0408-z> PMID: 24615717
39. Butowska E, Hanczakowski M, Zawadzka K. You won't guess that: On the limited benefits of guessing when learning a foreign language. *Mem Cogn.* 2021. <https://doi.org/10.3758/s13421-021-01254-2> PMID: 34913140

40. Seabrooke T, Mitchell CJ, Wills AJ, Inkster AB, Hollins TJ. The benefits of impossible tests: Assessing the role of error-correction in the pretesting effect. *Mem Cognit*. 2022; 50: 296–311. <https://doi.org/10.3758/s13421-021-01218-6> PMID: 34363196
41. Tulving E. Elements of episodic memory. 1983.
42. Tulving E. Episodic Memory and Autonoesis: Uniquely Human? The missing link in cognition: Origins of self-reflective consciousness. 2005. pp. 3–56. <https://doi.org/10.1093/acprof:oso/9780195161564.003.0001>
43. Tulving E, Hayman CA, Macdonald CA. Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. US: American Psychological Association; 1991. pp. 595–617. <https://doi.org/10.1037//0278-7393.17.4.595> PMID: 1832430
44. Carneiro P, Lapa A, Finn B. Memory updating after retrieval: when new information is false or correct. *Memory*. 2021;0: 1–20. <https://doi.org/10.1080/09658211.2021.1968438> PMID: 34412559
45. Bernstein DM, Scoboria A, Desjarlais L, Soucie K. “False memory” is a linguistic convenience. *Psychol Conscious Theory, Res Pract*. 2018; 5: 161–179. <https://doi.org/10.1037/cns0000148>
46. Brewer WF. Memory for the pragmatic implications of sentences. *Mem Cognit*. 1977; 5: 673–678. <https://doi.org/10.3758/BF03197414> PMID: 24203284
47. Carneiro P, Lapa A, Reis J, Ramos T. Testing pragmatic inferences: The impact of language and culture. *Psicológica J*. 2020; 41: 1–20. <https://doi.org/10.2478/psicolj-2020-0001>
48. Chan JCK, McDermott KB. Remembering pragmatic inferences. *Appl Cogn Psychol*. 2006; 20: 633–639. <https://doi.org/10.1002/acp.1215>
49. Mullet HG, Marsh EJ. Correcting false memories: Errors must be noticed and replaced. *Mem Cognit*. 2016; 44: 403–412. <https://doi.org/10.3758/s13421-015-0571-x> PMID: 26576564
50. Maraver MJ, Lapa A, Garcia-Marques L, Carneiro P, Raposo A. Imagination Reduces False Memories for Everyday Action Sentences: Evidence From Pragmatic Inferences. *Front Psychol*. 2021; 12: 3551. <https://doi.org/10.3389/fpsyg.2021.668899> PMID: 34489789
51. World Health Organisation. Declaration of Helsinki World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects. *J Am Med Assoc*. 2013; 310: 2191–2194. <https://doi.org/10.1001/jama.2013.281053>
52. McDermott KB. Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Mem Cogn*. 2006; 34: 261–267. <https://doi.org/10.3758/bf03193404> PMID: 16752590
53. Rowland CA. The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol Bull*. 2014; 140: 1432–1463. <https://doi.org/10.1037/a0037559> PMID: 25150680
54. Fazio LK, Marsh EJ. Surprising feedback improves later memory. *Psychon Bull Rev*. 2009; 16: 88–92. <https://doi.org/10.3758/PBR.16.1.88> PMID: 19145015
55. Bangert-Drowns RL, Kulik C-LC, Kulik JA, Morgan M. The Instructional Effect of Feedback in Test-Like Events. *Rev Educ Res*. 1991; 61: 213–238. <https://doi.org/10.3102/00346543061002213>
56. Clare L, Jones RSP. Errorless learning in the rehabilitation of memory impairment: a critical review. *Neuropsychol Rev*. 2008; 18: 1–23. <https://doi.org/10.1007/s11065-008-9051-4> PMID: 18247118
57. Dunn J, Clare L. Learning face-name associations in early-stage dementia: comparing the effects of errorless learning and effortful processing. *Neuropsychol Rehabil*. 2007; 17: 735–754. <https://doi.org/10.1080/09602010701218317> PMID: 17852759
58. Warmington M, Hitch GJ. Enhancing the learning of new words using an errorless learning procedure: evidence from typical adults. *Memory*. 2014; 22: 582–594. <https://doi.org/10.1080/09658211.2013.807841> PMID: 23799282
59. Davis SD, Chan JCK, Wilford MM. The Dark Side of Interpolated Testing: Frequent Switching Between Retrieval and Encoding Impairs New Learning. *J Appl Res Mem Cogn*. 2017; 6: 434–441. <https://doi.org/10.1016/j.jarmac.2017.07.002>
60. Finn B, Roediger HL III. Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2013. pp. 1665–1681. <https://doi.org/10.1037/a0032377> PMID: 23565780
61. Metcalfe J, Xu J. Learning from one’s own errors and those of others. *Psychon Bull Rev*. 2018; 25: 402–408. <https://doi.org/10.3758/s13423-017-1287-7> PMID: 28429177
62. Wahlheim CN, Jacoby LL. Remembering change: The critical role of recursive reminders in proactive effects of memory. *Mem Cogn*. 2013; 41: 1–15. <https://doi.org/10.3758/s13421-012-0246-9> PMID: 22918874
63. Gallo DA, Roediger HL, McDermott KB. Associative false recognition occurs without strategic criterion shifts. *Psychon Bull Rev*. 2001; 8: 579–586. <https://doi.org/10.3758/bf03196194> PMID: 11700910

64. Herron JE, Wilding EL. An electrophysiological dissociation of retrieval mode and retrieval orientation. *Neuroimage*. 2004; 22: 1554–1562. <https://doi.org/10.1016/j.neuroimage.2004.04.011> PMID: [15275912](https://pubmed.ncbi.nlm.nih.gov/15275912/)
65. Düzel E, Picton TW, Cabeza R, Yonelinas AP, Scheich H, Heinze HJ, et al. Comparative electrophysiological and hemodynamic measures of neural activation during memory-retrieval. *Hum Brain Mapp*. 2001; 13: 104–123. <https://doi.org/10.1002/hbm.1028> PMID: [11346889](https://pubmed.ncbi.nlm.nih.gov/11346889/)
66. Butterfield B, Mangels JA. Neural correlates of error detection and correction in a semantic retrieval task. *Cogn Brain Res*. 2003; 17: 793–817. [https://doi.org/10.1016/S0926-6410\(03\)00203-9](https://doi.org/10.1016/S0926-6410(03)00203-9)
67. Brewer WF, Sampaio C, Barlow MR. Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *J Mem Lang*. 2005; 52: 618–627. <https://doi.org/10.1016/j.jml.2005.01.017>
68. Sampaio C, Brewer WF. The role of unconscious memory errors in judgments of confidence for sentence recognition. *Mem Cogn*. 2009; 37: 158–163. <https://doi.org/10.3758/MC.37.2.158> PMID: [19223565](https://pubmed.ncbi.nlm.nih.gov/19223565/)
69. Butterfield B, Metcalfe J. The correction of errors committed with high confidence. *Metacognition Learn*. 2006; 1: 69–84. <https://doi.org/10.1007/s11409-006-6894-z>
70. Fazio LK, Marsh EJ. Correcting false memories. *Psychol Sci a J Am Psychol Soc / APS*. 2010; 21: 801–803. <https://doi.org/10.1177/0956797610371341> PMID: [20483825](https://pubmed.ncbi.nlm.nih.gov/20483825/)
71. Wong SSH, Lim SWH. The derring effect: Deliberate errors enhance learning. *J Exp Psychol Gen*. 2021. <https://doi.org/10.1037/xge0001072> PMID: [34242048](https://pubmed.ncbi.nlm.nih.gov/34242048/)