UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

THE EFFECT OF RETRIEVAL ON THE CORRECTION OF MEMORY ERRORS STEMMING FROM PRAGMATIC INFERENCES

Filipa Margarida Rodrigues Bastos Gonçalves

MESTRADO INTERUNIVERSITÁRIO EM NEUROPSICOLOGIA CLÍNICA E EXPERIMENTAL

2021

UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

THE EFFECT OF RETRIEVAL ON THE CORRECTION OF MEMORY ERRORS STEMMING FROM PRAGMATIC INFERENCES

Filipa Margarida Rodrigues Bastos Gonçalves

Dissertação orientada pela Doutora María Jesús Maraver e pela Professora Doutora Ana Luísa Raposo

MESTRADO INTERUNIVERSITÁRIO EM NEUROPSICOLOGIA CLÍNICA E EXPERIMENTAL

2021

Acknowledgments

I would not be where I am today, finishing my Master’s degree, without the help of a whole array of people. They deserve all my gratitude, and more.

To my supervisor, María J. Maraver, PhD, for her motivation, never-ending patience, and kindness, always available to teach me whatever, whenever, as many times as necessary. To my supervisor, Professor Ana Raposo, PhD, for her invaluable feedback and for always having a kind word to say, even when I was lacking. To both my supervisors, for all their support and for the incomparable knowledge they selflessly shared with me.

To Ana Lapa, MSc, for all her work in the data collection process, and for taking the time to patiently and kindly explain everything to me.

To my parents, for their unwavering support and for always believing in me and my capabilities, even when I do not. To my family and friends, for being there for me every single time. Although indirectly, you were all important for the completion of this thesis.

It is not easy to translate feelings into words, but I hope it was clear how fortunate and deeply thankful I am for having crossed paths with all of you.

Thank you.

This work was supported by Fundação para a Ciência e a Tecnologia through a Project Grant (PTDC/PSI-ESP/28414/2017).

Index

Abstract

A current debate concerns whether error production benefits or hinders learning. According to the Memory Updating After Retrieval framework (Finn, 2017), the act of retrieval makes memory more malleable to incorporate new information, and thus, if error retrieval is followed by corrective feedback, memory should be updated and benefit learning.

In the present study, to elicit errors similar to everyday memory errors, we chose the Pragmatic Inferences Paradigm (Brewer, 1977), comprising sentences such as “The baby stayed awake all night” which often leads individuals to pragmatically imply and later remember that “The baby cried all night”. To investigate if these retrieval errors, followed by feedback, benefit learning, we manipulated retrieval (active vs. passive recognition) and feedback (with vs. without). Therefore, participants (n=120) were randomly assigned to one of these four groups.

In line with Finn’s (2017) framework, we hypothesized that participants in the active recognition condition would produce more correct responses and less pragmatic inference errors than participants in the passive condition, and that this benefit would be greater in the feedback condition, since feedback represents the new information that is more easily incorporated during retrieval of an error.

Overall, our results showed an interaction between type of retrieval and feedback, such that participants who engaged in active recognition produced more correct responses and generated less pragmatic inference errors, when compared to those in the passive condition, and these differences were greater in the feedback than no feedback condition. Importantly, feedback promoted error correction in the active recognition condition in a greater extent than error correction in the passive condition. These results support Finn’s framework, by showing that the active retrieval of information (even if it contains errors) promotes learning (in a greater extent than passive recognition) as long as it is followed by corrective feedback.

Keywords: retrieval; errorful learning; pragmatic inferences; error correction; feedback

Resumo Alargado

Em contexto educativo, os testes são habitualmente utilizados meramente como forma de os professores avaliarem os conhecimentos dos alunos, para perceberem o que já foi aprendido e o que precisa de ser revisto. No entanto, vários estudos têm vindo a demonstrar que os testes podem ser uma ferramenta importante para melhorar e otimizar a aprendizagem (para uma visão global, ver Pyc et al., 2014 e Roediger & Butler, 2011). Este benefício para a aprendizagem é designado de Efeito de Teste, e acontece quando informação previamente apresentada é recuperada através de um teste, em vez de ser simplesmente estudada novamente (Roediger & Karpicke, 2006). Este efeito de recuperação tem maioritariamente sido investigado com tentativas de recuperação bem-sucedidas, ou seja, quando as pessoas recuperam a informação correta. Mas o que acontece quando é cometido um erro, ou seja, quando é recuperada informação incorreta?

A literatura atual tem procurado perceber que papel tem o tipo de recuperação no Efeito de Teste e, consequentemente, no seu benefício para a aprendizagem: será uma recuperação bem-sucedida (quando é recuperada a resposta correta) mais benéfica, ou, pelo contrário, trará uma recuperação mal-sucedida (quando é recuperado um erro) mais vantagens? Existem duas principais perspetivas com respostas opostas a esta questão.

A perspetiva da Aprendizagem Sem Erros defende que cometer erros durante a aprendizagem os torna mais fortes e salientes, aumentando a probabilidade de que ocorram novamente (Metcalfe, 2017). Os autores propõem que, ao recuperar informação incorreta, estão a ser praticadas estratégias erradas e ineficientes, tornando mais difícil a correção dos erros e a aprendizagem, mais tarde, de estratégias eficazes (Ausubel, 1968; citado por Metcalfe, 2017). De acordo com esta perspetiva, o *feedback* deve ser utilizado apenas como reforço positivo, ignorando os erros e focando-se em como executar as tarefas de forma correta (Metcalfe, 2017).

Por outro lado, a perspetiva da Aprendizagem Com Erros, proposta pela primeira vez por Izawa (1970), demonstrou que tentativas de recuperação mal-sucedidas conduziram a uma melhoria da aprendizagem em ensaios subsequentes. Recuperar erros, seguidos pela apresentação de *feedback* corretivo, resulta numa melhor aprendizagem da resposta correta do que simplesmente voltar a estudar a informação correta (e.g., Kornell et al., 2009). No entanto, para que este benefício aconteça, os erros produzidos devem ser detetados (Mullet & Marsh, 2016) e devem estar relacionados com o alvo da aprendizagem (Grimaldi & Karpicke, 2012; Hulser & Metcalfe, 2012). Este efeito benéfico tem vindo a ser estudado não só com materiais educativos (Richland et al., 2009), mas também em contexto de sala de aula (Kapur & Bielaczyc, 2012).

Partindo da proposta de que a recuperação mal-sucedida beneficia a aprendizagem, quais poderão ser os mecanismos subjacentes a este efeito? Existem várias propostas para a explicação deste fenómeno, mas Finn (2017), com a sua abordagem sobre atualização da memória após a recuperação, apresenta uma proposta que merece ser destacada. Segundo Finn (2017), a nova informação aprendida é guardada em memória através de um processo de consolidação, podendo ser mais tarde recuperada. Este processo de recuperação torna a memória mais maleável e propensa à incorporação de nova informação, que após sofrer um processo de reconsolidação é atualizada na nossa memória. Este efeito de recuperação pode tanto beneficiar como prejudicar a aprendizagem, dependendo se a nova informação incorporada é correta ou incorreta. É necessário que esta nova informação esteja bem alinhada com os objetivos e com o alvo da tentativa de recuperação para que beneficie, em vez de prejudicar, a aprendizagem. No que se refere especificamente à Aprendizagem Com Erros, quando alguém recupera um erro está a tornar a sua memória mais moldável e recetiva à incorporação de nova informação, que neste caso será o *feedback* corretivo. De seguida, através da reconsolidação, a memória é atualizada com a resposta correta, providenciada pelo *feedback* corretivo.

O presente estudo tem como principal objetivo investigar se a recuperação de erros, seguida pela apresentação de *feedback* corretivo, beneficia ou não a aprendizagem. Com esse intuito, foi utilizado o Paradigma das Inferências Pragmáticas (Brewer, 1977) em que são apresentadas frases que criam expectativas sobre algo que não foi explicitamente declarado (Carneiro, Lapa, Reis, & Ramos, 2020). Um bom exemplo é a frase “O bebé ficou acordado toda a noite”, que conduz à implicação pragmática de que “O bebé chorou toda a noite”. Frases como esta são responsáveis por originar inúmeras memórias falsas no nosso dia-a-dia, particularmente nas nossas interações sociais, em que recordamos informação que não foi apresentada (i.e., que o bebé chorou toda a noite; McDermott & Chan, 2006). No presente estudo, os participantes (n=120) leram um conjunto de frases que incluíam inferências pragmáticas. Seguidamente, durante a recuperação, na condição de Reconhecimento Ativo, foi-lhes pedido que decidissem se a frase apresentada estava correta ou incorreta (antiga ou nova) relativamente ao que tinha sido apresentado durante a codificação. Na condição de Reconhecimento Passivo, foi pedido aos participantes que lessem não só a frase apresentada, mas também a resposta que tinha sido dada por outro participante da condição de Reconhecimento Ativo (*Yoked design*). Além disso, foi manipulada a apresentação de *feedback*, com uma condição de *Feedback* (sob a forma da frase correta) vs. Sem *Feedback* (operações matemáticas simples para resolver). Os participantes foram distribuídos de forma aleatória por um de quatro grupos: reconhecimento ativo + *feedback*; reconhecimento passivo + *feedback*; reconhecimento ativo + sem *feedback*; reconhecimento passivo + sem *feedback*. De acordo com a literatura sobre o Efeito de Teste, temos como hipótese que os participantes na condição de reconhecimento ativo terão um melhor desempenho (mais respostas corretas e menos inferências pragmáticas) do que os participantes da condição de reconhecimento passivo. Esperamos encontrar o mesmo padrão em relação aos participantes da condição de *feedback* relativamente à condição sem *feedback*. Antecipamos também que o efeito benéfico da condição de reconhecimento ativo, quando comparada com a de reconhecimento passivo, será maior na condição de *feedback* do que na condição sem *feedback*. Estas hipóteses vão ao encontro e são apoiadas pela abordagem de Finn (2017), uma vez que o *feedback* representa a nova informação mais facilmente incorporada devido à recuperação de um erro, recuperação essa que só ocorre na condição de reconhecimento ativo.

Os nossos resultados demonstraram que os participantes da condição de reconhecimento ativo produziram mais e persistiram mais nas respostas corretas, bem como geraram menos erros de inferências pragmáticas, do que os da condição de reconhecimento passivo. Ademais, na condição de *feedback*, os participantes do grupo de reconhecimento ativo persistiram menos e corrigiram mais erros do que os do grupo de reconhecimento passivo. No entanto, verificou-se o oposto quando não era fornecido *feedback* - os participantes da condição de reconhecimento ativo persistiram mais e corrigiram menos erros do que os da condição de reconhecimento passivo. Este padrão de resultados constitui evidência tanto a favor da proposta de Finn (2017), como da perspetiva da Aprendizagem Com Erros. Por um lado, confirmam que a recuperação ativa (vs. passiva) torna a informação em memória mais maleável e suscetível à incorporação de nova informação. Por outro lado, demonstra a relevância da inclusão de feedback corretivo durante a recuperação, por forma a que a nova informação incorporada seja correta.

Os resultados sugerem que os estudantes beneficiariam em ter um papel mais ativo no seu processo de aprendizagem, através da adoção de estratégias de recuperação ativa (por exemplo, através da realização de testes), mesmo quando são evocados erros. Neste contexto, o fornecimento de *feedback* por parte do professor é também fundamental para a aprendizagem da informação correta.

Palavras-chave: recuperação; aprendizagem com erros; inferências pragmáticas; correção de erros; *feedback*

Introduction

False Memories, Pragmatic Inferences, and the Misinformation Effect

Every day of our lives we experience a multitude of events, which are stored in our memory, so we are able to later on recall them. However, we do not always remember the events exactly the way we lived them.

Memory has a malleable and reconstructive nature. This is highlighted by the mental constructions we develop of our understanding of everyday events (i.e., schemas), of which we retain only the gist of the information, a general impression of the whole (Bartlett, 1932). People are not always able to remember a certain event exactly as they have experienced it. For example, when someone witnesses a car accident and is later on asked about it, often they will provide details that are not accurate, such as claiming that one of the drivers had a very deep head wound when, in fact, it was a small cut. The information available during memory reconstruction is often unintentionally distorted for various reasons, such as to achieve coherence or be consistent with one's current knowledge, to align it with one's expectations, or merely because people may be deceived by their associative memory (Carneiro, Lapa, Reis, & Ramos, 2020). These reconstructive memory mechanisms underlie the development of false memories.

When experiencing false memories, individuals often feel like they are remembering specific details of the target event in a very vivid way (Chan & McDermott, 2006; Roediger & McDermott, 1995), which might lead to reporting memory errors with high confidence (Sampaio & Brewer, 2009). Furthermore, unlike when simply experiencing false beliefs, those who report false memories provide additional evidence by, for example, elaborating on the event, talking about their relating emotions, or explicitly expressing their confidence in their memory of the event (Wade et al., 2018). It is seen as unlikely, but nonetheless possible, for anyone to be found resistant to the development of false memories, as not even individuals with highly superior autobiographical memory are immune to episodic memory distortions (Patihis et al., 2013).

There are currently several methodologies available to study false memories, such as the very commonly used DRM (Deese-Roediger-McDermott) Paradigm (Deese, 1959; Roediger & McDermott, 1995) In this paradigm, participants are asked to study lists of semantically related words (e.g., bed, rest, awake) associated with a non-presented critical word (e.g., sleep) and are later on asked to complete a recall task, in order to see how often the non-presented critical word is reported and with how much confidence. The DRM is a lab-based paradigm that only partially represents how false memories are presented in everyday life events. Although there are studies defending otherwise (e.g., Burns et al., 2006; Burns et al., 2007), it is generally accepted that, in the DRM, false memories arise from the relational processing of the non-presented critical word, which is believed to enhance both veridical and false recognition (Gunter et al., 2005). This relational processing is thought to occur due to either the non-presented critical words receiving semantic activation from the related word list items (spreading activation approach), or because a substantial amount of gist information relating to the non-presented critical words is encoded (Burns et al., 2007).

Other paradigms have been employed for the study of false memories, such as the Pragmatic Inference Paradigm, which encompasses both semantic and episodic components processing, and is a better option when the goal is to have false memories that participants feel are common in their everyday lives. When undergoing everyday social interactions, we, as listeners, are required to engage in active inferential processes, which can affect our memory even when no manipulation is used and no erroneous information is supplied, leading to inaccuracies in recall (McDermott & Chan, 2006). Pragmatic inferences and pragmatic implications are a common example of such inaccuracies, occurring when a listener, led by the information in a sentence, expects something not explicitly stated nor necessarily implied by it (Carneiro, Lapa, Reis, & Ramos, 2020). For instance, the sentence “She lost her balance on the surfboard” pragmatically implies that “She fell from her surfboard”. According to the literature (Brewer, 1977; Carneiro, Lapa, Reis, and Ramos 2020), for the purpose of evaluating if a sentence could be categorized as having a pragmatic implication, the *but not* test should be used, which is able to rule out semantically synonymous sentences, logically implied sentences, or sentence pairs with no relation to each other. Using the previous example, there is a pragmatic implication because both sentences can be joined with the conjunction *but not* and form an acceptable coherent sentence: “She lost her balance on the surfboard *but* did *not* fall”. Despite this, pragmatic implications are not fully determined by the original sentence, as it has been shown by Brewer (1977) through the qualitative analysis of his study, in which he found ambiguous implications that originated different inferences from participants (e.g., “The absent-minded Professor didn’t have his car keys” was recalled by participants as the Professor “forgot his car keys”, “lost his car keys” or “left his car keys in the car”). Pragmatic inferences are a useful tool to investigate false memories as it has been consistently shown that participants tend to recall the pragmatic implications of the sentences rather than what was actually stated (Brewer, 1977; Carneiro, Lapa, Reis, & Ramos, 2020; Chan & McDermott, 2006). In contrast to the DRM paradigm that elicits false memories through inter-item (relational) processing, pragmatic inferences induce false memories at an intra-item level and allow dissociating between semantic and episodic memory components. In the sentence “The baby stayed awake all night”, it is our general semantic knowledge that babies cry during the night that makes us infer that “The baby cried all night”. On the other hand, the episodic component is the unique event that the baby stayed awake and it is this that participants need to remember in order to generate a correct answer, rather than what people commonly guess, infer, or what they know usually happens in situations like these.

Once experienced, false memories are hard to correct, because it is not enough that the individual realizes that she/he has made a mistake; she/he also needs to be aware of the correct information. Moreover, while in a laboratory setting these conditions might be easier to fulfil, in the real world it is less likely that one can find objective information that could be compared to one’s personal memories in order to correct them, especially since these may elicit a much higher level of confidence and a much lower level of willingness to accept corrections, when compared to those prompted in research experiments (Mullet & Marsh, 2016).

Due to the malleable and reconstructive nature of our memory, memory distortions can arise without explicit external influence or explicit misinformation being provided, as it is the case with pragmatic inferences. However, when erroneous information is supplied and an individual is exposed to it, the change in the initial reporting that follows is called a Misinformation Effect (Loftus, 2005).

According to Loftus (2005), there are certain conditions that, when met, make people more susceptible to the negative impact of misinformation. The Discrepancy Detection Principle (Tousignant et al., 1986) states that one’s recollections are more likely to change if discrepancies between the memory of an original event and the misinformation are not detected right away, even though it is still possible to have false memories when these discrepancies are detected, since when confronted with misinformation one can believe that information is right and they are wrong. Furthermore, the mere passage of time can affect the memory of the original event, allowing it to fade and thus be more prone to be changed by misinformation (Loftus et al., 1978), because due to its weakened state, it is less probable that any discrepancies are perceived (Loftus, 2005). Additionally, temporarily changing someone's state or even simply suggesting it, such as making them believe they have drunk alcohol (Assefi & Gary, 2003), can increase the misinformation effect, leaving the individuals more susceptible to it and disrupting their ability to detect discrepancies (Loftus, 2005).

Consonant with the Discrepancy Detection Principle (Tousignant et al., 1986), warning a person about the presence of misinformation after it has already been processed does not enhance their ability to resist it (Greene et al., 1982), likely because the misleading information has already been incorporated into the original memory and altered it (Loftus, 2005). However, there are limited circumstances in which these post-misinformation warnings, regardless if they are general or item-specific, might be successful in reducing the misinformation effect, such as when they immediately follow the encoding of the misleading information that must be in a low state of accessibility, otherwise they will not have the desired effect (Eakin et al., 2003).

Testing Effect, Successful/Unsuccessful Retrieval, and Hypercorrection Effect

Throughout the years, a lot of research has been conducted to understand the role of retrieval (usually in the form of tests) in learning. Tests are usually thought of as purely assessment tools, to appraise what has been successfully learned and what needs to be further studied or corrected. However, tests can be a powerful tool to improve learning, as well as offer metacognitive benefits by helping the learner to identify whether or not the information has been understood or learned (Bjork & Bjork, 2011), as has been described across the literature (overview in Pyc et al., 2014, and Roediger & Butler, 2011). More specifically, this improvement in long-term retention is called testing effect (Roediger & Karpicke, 2006) and consists of a greater benefit for future learning when previously learned information is retrieved in a test, rather than simply restudied. One mechanism that is thought to underlie this effect is the activation of elaborative information related to the target answer, which increases the probability that said information will facilitate later retrieval of the correct answer (Carpenter, 2009). A variation of this effect is the forward testing effect, in which retrieval of previously learned information enhances the learning of subsequent new information (review in Pastötter & Bäuml, 2014). Furthermore, subsequent learning of the tested information can also be potentiated by retrieval or attempted retrieval, and this is called test potentiated learning (Carneiro, Lapa, & Finn, 2020).

However, there is still some debate about what role the type of retrieval (successful, when the correct answer is produced, or unsuccessful, when there is the generation of an error) plays in future learning. Does unsuccessful retrieval enhance, hinder, or simply has no effect on subsequent learning? Two main perspectives have been proposed.

On the one hand, it has been argued that the errorless learning procedure is the most beneficial, which is consistent with several well-established theories of learning and memory (e.g., Bandura, 1986; Skinner, 1953). Authors following this view believe that committing errors can make them more prominent and stronger, increasing their likelihood of recurrence and, therefore, it is suggested that they should be completely avoided (Metcalfe, 2017). In congruence with this view, a study by McDermott (2006), using semantic associates (e.g., hill, valley, climb), found that when individuals took three initial tests before the final free recall test, forgetting of the studied words was observed. However, across the three initial retrieval attempts, non-studied semantic associates (e.g., mountain) were recalled with increasing frequency. Furthermore, in line with this account, Ausubel (1968; cited in Metcalfe, 2017) warned about the dangers of using an exploratory learning strategy, defending that letting people make errors during the learning process not only encourages them to practice incorrect and inefficient approaches, but also makes it harder for them to learn the correct procedures later on, because those errors are difficult to overwrite. Contrastingly, using an errorless procedure has been found to improve the learning of vocabulary in healthy adults (Warmington & Hitch, 2013). In their study, Warmington and Hitch (2013) demonstrated that the ability to learn a set of novel names for novel objects was significantly superior when individuals followed an errorless learning procedure than when they followed an errorful one, and this beneficial effect persisted over a 3–4-day delay. Moreover, the authors replicated these results in a second experiment, in which they used a more naturalist task (i.e., learning rare English words and their meaning). According to their interpretation of these results, errorless learning elicits the creation of more specific and more retrievable memory representations of novel items and decreases interference between errors and correct responses during the retrieval process. Additionally, it is recommended by the authors advocating for the errorless learning approach that feedback should be seen and used as a mere form of positive social reinforcement, focusing on how to correctly execute the tasks, and ignoring any errors that might arise (Metcalfe, 2017).

On the other hand, it has been proposed that an errorful learning approach is the most beneficial, and there are many studies (e.g., Kornell et al., 2009; Potts & Shanks, 2014) showing that generating errors - as long as they are followed by corrective feedback - results in better memory for the correct response than simply studying the correct information (Metcalfe, 2017). This idea was first introduced by Izawa (1970), who suggested that unsuccessful retrieval attempts led to improved learning in subsequent trials, when compared to a procedure generating fewer incorrect responses. Moreover, across the years, these findings have been thoroughly replicated and extended, and it has been shown that, in order for this beneficial effect to take place, the errors produced must be detected (Mullet & Marsh, 2016) and cannot be unrelated to the target (Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012). Additionally, a study by Kang et al. (2011) showed that when participants are forced to produce a completely uninformed guess, the beneficial effect to learning does not take place. In terms of how this effect might occur in educational settings, a study by Richland et al. (2009) has shown, using educationally relevant materials (e.g., an essay about vision), that taking a test before being presented with the materials enhances memory to a higher degree than simply spending an equivalent amount of time studying them. Another study by Kapur and Bielaczyc (2012), using complex mathematics materials in a real classroom, found that students in the group that was unsuccessful at the pretest were the ones that had a better performance on the posttest. These benefits, stemming from error generation during the learning process, have been suggested to have their locus on episodic/explicit memory, rather than semantic/implicit memory (Metcalfe & Huelser, 2020).

According to Metcalfe (2017), the hypercorrection effect consists of the increased probability of high-confidence errors to be corrected at retest, immediate or delayed (e.g., Butler et al., 2011), when compared to errors made with lower confidence. People rarely make errors when they are highly confident on their responses, but when they do, there are two main factors that may be responsible for the hypercorrection effect in young adults. First, the surprise they feel by making a mistake when they were certain they were right, which might make them allocate their attentional resources to better remember the correct response. This has been supported by brain activity responses using EEG, with younger adults presenting a strong attention-related P3a to high-confidence error feedback, but not to low-confidence error feedback (Metcalfe et al., 2015). Also, there is a greater semantic familiarity with the domain of high-confidence errors versus low-confidence errors, which has been shown through the finding that young adults have a higher likelihood of producing a correct second guess, as well as choosing the correct answer in a multiple-choice test that excludes their first answer, following the production of a high-confidence error (Metcalfe & Finn, 2011). Furthermore, Metcalfe (2017) suggests that the familiarity effects might be combined with the effect of surprise, since a person's surprise at producing an error in a certain domain is more prominent when they are very familiar with the domain of knowledge.

How Error Production Benefits Learning

Nowadays, it is generally agreed that producing errors can help learning in certain circumstances. Several theories have been put forward to explain the mechanisms behind the benefit of producing errors during learning.

According to the Mediation Theory (Carpenter, 2011; Pyc & Rawson, 2010), one is likely to generate effective mediational retrieval cues when tested and, along these lines, errors act as stepping stones to reach the correct responses, instead of being seen as competitors and interferences. Moreover, this idea is congruent with what other research has previously suggested (e.g., Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012), that errors must be self-generated and related to the target to enhance learning. However, it is still not clear whether this mediation is entirely semantic or if it also relies on an episodic-memory component (Metcalfe, 2017).

The Recursive Reminding Theory (Jacoby & Wahlheim, 2013) postulates that errors, as well as other contextual aspects at encoding, might facilitate learning, in the sense that they are related to the retrieval of the initial episodic event, where both the correct response and the error were incorporated. The authors call recursive reminding to the process, relying on one's episodic memory, through which one remembers the context of an error they made, and by doing so they not only bring the error back to mind, but they also think how it had been clear said error had been made and that the expected response was the provided correction for it. However, while this theory seems to be consistent with research reporting that people with impaired episodic memory see their learning process jeopardized by error generation, there are studies, such as Butterfield and Metcalfe's (2001) and Metcalfe and Miele's (2014), showing that correct recall is independent of error generation during retrieval, which goes against the assumption this theory seems to propose of error re-generation at retrieval having an enhancing effect on the memory for the correct answer. Considering these contradictory findings, further research is required to clarify and expand the information available for this theory.

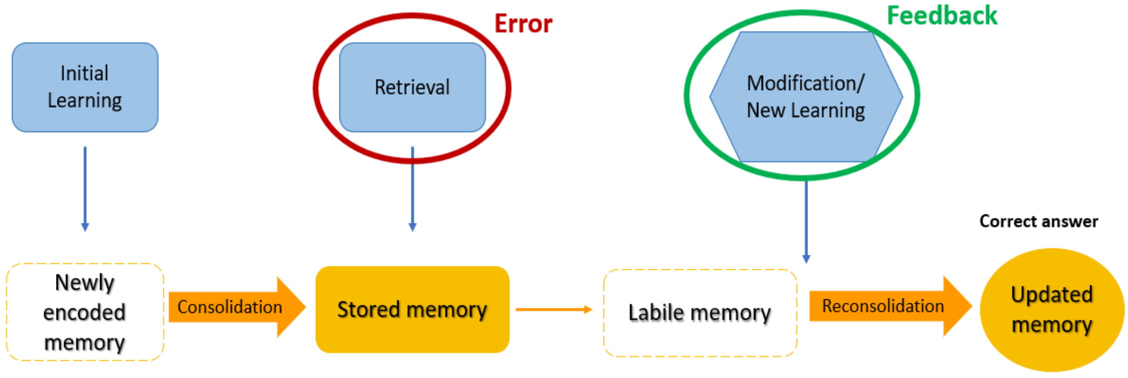
The Prediction Error Theory is based on computational and machine models of learning, which posit that learning happens when a person's expectations and the outcome are discrepant, or when a prediction error exists, because these unexpected results force the network to change in order to accommodate them. In terms of applicability to the error-correction paradigm, these models seem to be particularly suitable for the hypercorrection effect, since it appears that the extent of learning is determined by the size of the prediction error, and high-confidence errors seem to be substantial and surprising. Notwithstanding, if one thinks about the assessment of the prediction error with regard to the comparison between the correct answer and the representational characteristics of the retrieved item, then less learning should take place with high-confidence errors, because their magnitude should be smaller considering that they were found to be more similar to the correct answers than low-confidence errors, and the participants who produced them were more likely to be correct on a second try, as well as completing the correct response when given fragmentary cues (Metcalfe & Finn, 2011).

Despite being conceived within a fear-conditioning paradigm, aiming to address, for example, the return of fear responses (Schiller et al., 2010), the Reconsolidation Theory has evident similarities with the error-correction paradigm, and is consistent with the core results of studies exploring learning from errors with feedback. The main postulation of this theory is that to change or eliminate a conditioned fear, it is essential that the unwanted response is evoked first, since it is then, within a short time window and in a transient state, that said response can be eliminated or altered and reconsolidated. Thinking about the error-correction paradigm, the need for an error to be retrieved in order to be susceptible to change is one of the similarities that it has with this framework. Furthermore, as reported by Lee (2008), the likelihood of a fear response being malleable increases the more strongly it is evoked, which might be comparable to not only parts of the error-correction paradigm, but also the hypercorrection effect. Yet, once learned, is the correct response permanent or can the incorrect response make a spontaneous recovery? This is, undeniably, a question of the utmost importance in educational settings, but research has not yet found a clear answer for it (Metcalfe, 2017).

Along these lines, Finn (2017), with the Memory Updating After Retrieval framework, makes a proposal of the potential mechanisms behind this beneficial effect of error production (see Figure 1). According to this perspective, when someone learns new information, that information is encoded and then, through the process of consolidation, is stored in memory. It is from here that the information can be later retrieved. This process of retrieval makes memory more malleable and prone to the incorporation of new information, which, when presented, is reconsolidated into an updated memory. However, retrieval can both enhance and impair learning of new information. If the new information presented is not well-aligned with the goals and the target of the retrieval attempt, impairment may occur. Thinking specifically about error correction, by retrieving an error, one is making their memory more malleable and prone to the incorporation of new information, which in this case is the corrective feedback. Then, through reconsolidation, the memory is updated with the correct answer. A study by Carneiro, Lapa, and Finn (2020), using a DRM Paradigm, has presented results favorable to this framework. When compared to restudy, retrieval enhanced the integration of subsequent false information. Furthermore, retrieval also enhanced the integration of subsequent correct information (feedback) when that information was presented immediately after each error. Overall, the results of this study suggest that retrieval facilitates the incorporation of new and related information, regardless of whether it is correct or false, which is in line with the Memory Updating After Retrieval Framework. However, to the best of our knowledge, no previous study has tested this framework using experimental paradigms that generate errors more representative of real-world situations.

Figure 1.

The Memory Updating After Retrieval Framework Applied to Error Correction



The Importance of Corrective Feedback

Throughout the years, across many studies, corrective feedback has been shown to be of crucial importance to the errorful learning process (Metcalfe, 2017), including in real educational settings (Kornell & Metcalfe, 2014). Indeed, learners get almost no benefit from making errors unless the feedback provided is well-aligned with the goals of retrieval and contains the correct answer, instead of merely informing if their answer is correct or not (Pashler et al., 2005). It is essential to make sure that learners pay attention and understand the feedback presented, otherwise it will not produce beneficial effects (Metcalfe, 2017). In addition, a study by Finn and Metcalfe (2010) demonstrated amplified beneficial effects when the feedback was elaborative and scaffolded.

In terms of timing of the feedback presentation, a study by Metcalfe and colleagues (2009) found that college students performed as well when receiving immediate feedback, as they did when receiving delayed feedback (up to one week). However, a more recent study by Mullet and Marsh (2016), using a paradigm with false memories stemming from pragmatic inferences, found that while both immediate and delayed feedback highly reduced the proportion of errors produced across the tests, immediate feedback was more effective. Furthermore, the same study showed that participants who did not receive any kind of feedback produced about as many pragmatic inferences on the final test as they did in the beginning, which greatly contrasts with the participants who were provided with feedback containing the correct answer. The latter rarely produced any pragmatic inferences in the final test, even though they had produced them about half the time during the initial test.

Error Correction Paradigms

Across the literature, many studies have found beneficial effects to the errorful learning approach, using different methodologies (e.g., reading an essay, Richland et al., 2009). One of the most widely used is the Weak Associate Paradigm (e.g., Kornell et al., 2009) that promotes the generation of a high number of guessing errors. In Kornell and colleagues' study (2009), participants underwent test trials and read-only trials across four different experiments using the Weak Associate Paradigm (Experiments 3 to 6). In the test trials, they were presented, for 8 seconds, with a word (cue) (i.e., garden) and were asked to guess the target, which was always a weak associate (i.e., playground), although participants' guess response was often a strong associate (i.e., flowers). Participants were afterwards presented, for 5 seconds, with feedback that included the "correct" answer, i.e., the weak associate. In the read-only trials, participants were presented, for 13 seconds, with both the cue and the target and were only asked to read them (i.e., garden - playground). After a distractor task, there was a cued-recall test, in which participants were asked to complete the blank space with the correct answer (e.g., garden - ____). There were some differences in the design of the four experiments. In experiment 3, time of presentation was manipulated in the read-only trial, with the cue and target words being presented for only 5 seconds, matching the feedback presentation in the test trials. In experiment 4, the delay between studying the words and performing the test was 5 minutes, while in experiment 5 the delay was an average of 38 hours. In experiment 6, the type of trial (test vs. read-only) was manipulated between-participants, unlike all the other experiments that used a within-participants design. This modification aimed to account for the possibility that the mix between the read-only items and the tested items was responsible for the testing advantage, since not only the tested items might have been rehearsed during the presentation of the read-only items, but they might also have been encoded more distinctly. The results showed that, after the test trials, participants had a better performance in the cued-recall test than after the read-only trials. This was true across all four experiments conducted by the authors using this paradigm, despite the different manipulations that were made in each experiment, suggesting that this beneficial effect is resistant to different time delays between retrieval and the final memory test (i.e., 5 minutes and 38 hours) and persists in both within and between-participants manipulations.

Even though this methodology allows participants to generate numerous guessing errors, these are not very representative of “real world” errors, because they are not correct or incorrect beyond the context of the experiments. If the goal is to study errors as close to what participants experience in their daily lives as possible, the Pragmatic Inferences Paradigm (Brewer, 1977) appears to be more adequate, since pragmatic inferences are likely to be responsible for originating many everyday false memories, with the additional advantage of possibly providing a bridge between word-list studies and studies of discourse comprehension (McDermott & Chan, 2006). Unlike guessing errors, and as discussed earlier, errors stemming from pragmatic inferences allow researchers to tap into both the semantic and the episodic components of memory. Importantly, when individuals generate this type of error, they do so with more confidence than guessing errors, and as such pragmatic inference errors are more similar to one’s personal everyday memory errors (Mullet & Marsh, 2016).

Current Study

The current study lies within the scope of the errorful learning perspective. Our goal is to understand if the retrieval of errors stemming from pragmatic inferences, followed by corrective feedback, benefits learning. For that purpose, the Pragmatic Inferences Paradigm (Brewer, 1977) is used, with 32 Portuguese sentences chosen from Carneiro, Lapa, Reis, and Ramos’ (2020) adaptation to the language. By using this paradigm, we increase ecological validity as the errors generated are more representative of everyday experiences and interactions.

Two factors were manipulated. First, we manipulated the type of retrieval, with two conditions: Active Recognition (which corresponds to retrieval in previous studies) vs. Passive Recognition (which corresponds to the mere presentation of information in previous studies). In the active condition, participants are actively involved in the retrieval process and are asked to decide if the sentence presented is correct or incorrect (old or new), based on whether or not it was shown during encoding. In the passive condition, participants read the responses of another participant (i.e., the answer provided by another participant in the active condition). In addition, we manipulated feedback, with two conditions: Feedback vs. No Feedback. In the feedback condition, participants are provided with the full correct sentence after they make their old/new decision for each sentence (active recognition condition) or after they are presented with the sentence and response of another participant (passive recognition condition). In the no feedback condition, instead of being provided with the correct answer, participants are asked to solve simple math operations.

We anticipate that participants in the active recognition condition will perform better (i.e., more correct answers and less pragmatic inferences) on the final test than those in the passive recognition condition, which is in line with the literature on the testing effect. Furthermore, we expect the same for participants in the feedback relative to the no feedback condition, and hypothesize that the beneficial effect for the active recognition condition, relative to the passive recognition condition, will be greater in the feedback condition, when compared to the no feedback condition. These hypotheses are in line with and are supported by Finn's (2017) framework, since feedback represents the new information more easily incorporated due to the retrieval of an error, which only happens in the active recognition condition.

Method

Participants

A total of 128 people from a database of volunteers of the Faculty of Psychology, University of Lisbon participated in this online study. From these, 8 participants were excluded due to several reasons: having invalid data (n=2), reporting having been interrupted during the study (n=2), having no yoked pair (n=2), or having performed the tasks twice or contrary to the instructions (n=2). Thus, only the data of 120 valid participants (65 female; $M_{age} = 27.47 \pm 8.92$) was analyzed. Participants were contacted via e-mail and were sent a link to complete the experiment. After verification that they had, indeed, completed the experiment, they were awarded 10€ vouchers for their participation. All participants gave written informed consent, and the study was approved by the local Ethics Committee of the Faculty of Psychology at the University of Lisbon.

Participants were randomly assigned to one of four group conditions (passive recognition + feedback; passive recognition + math operations; active recognition + feedback; active recognition + math operations).

Design

The current study followed a 2 (recognition: active vs. passive) x 2 (feedback: corrective feedback vs. no feedback, i.e., math operations) design, with a between-subjects manipulation for recognition and feedback. The dependent variables were the correct answers, pragmatic inference errors, intrusions, and omissions produced by the participants.

A yoked control design was used in order to increase methodological control and guarantee that active and passive groups were similar and had a matching proportion of errors in both conditions. In this design, each participant in the passive recognition condition was paired with a participant in the active recognition condition and saw the latter's exact responses during the recognition task in the intermediate phase.

Materials

Thirty-two pragmatic implication sentences in Portuguese were selected from those adapted by Carneiro, Lapa, Reis, and Ramos (2020) (see Appendix A). To improve experimental control of the materials used, pragmatic inference sentences were counterbalanced as a function of both the format of the sentence presented at encoding (critical vs. filler) and the format of the sentence presented in the intermediate phase (match vs. mismatch). Specifically, 16 critical sentences were presented at encoding in their original form, to elicit a pragmatic inference (e.g., "The baby stayed awake all night"), while 16 filler sentences were adapted from the original sentences and presented in the pragmatic inference format (e.g., "The baby cried all night"). Filler sentences were very similar to the critical ones, not only to ensure that the effects found were not due to the format of the sentences, but also to make sure that the participants did not become aware or suspicious of the manipulations. Of note, during encoding, each participant saw the same sentence frame (e.g., "The baby ____ all night") only once, in either the original ("stayed awake") or the adapted format ("cried"). Subsequently, in the intermediate phase, the presented sentences either matched or mismatched the sentences presented at encoding. More specifically, a sentence was considered a match when it was presented at the intermediate phase in the same format as it was presented in the encoding phase (e.g., "The karate champion hit the cinder block"), and a mismatch when it was presented differently from the encoding phase (e.g., "The karate champion broke the cinder block"). Examples of how counterbalanced sentences were presented are provided in Table 1.

Table 1

Examples of How Counterbalanced Sentences Were Presented During the Encoding Phase (Critical or Filler) and the Intermediate Phase (Match or Mismatch)

Encoding Phase "The baby _____all night"		**Intermediate Phase**	
IF:		**Match (16)**	**Mismatch (16)**
Critical (16)	"stayed awake"	*If critical presented at encoding:* "The baby stayed all night"	*If critical presented at encoding:* "The baby cried all night"
IF:		**Match (16)**	**Mismatch (16)**
Filler (16)	"cried"	*If filler presented at encoding:* "The baby cried all night"	*If filler presented at encoding:* "The baby stayed awake all night"

Procedure

Due to the covid-19 pandemic, the experiment was run online, and participants completed the experiment in their own computers, at home. The software used for programming and data collection was the online platform Qualtrics (Qualtrics, Provo, UT) and a link to access the experiment was shared with the participants via e-mail. To guarantee the yoked design, a first set of participants randomly selected to the active recognition group were sent the links. After completing the experiment, each one of them was manually paired with a participant in the passive recognition group, who was then sent a specific link containing the yoked responses.

The experimental procedure (see Figure 2) was divided in three phases: Encoding Phase, Intermediate Phase, and Final Memory Test.

During the Encoding Phase, participants were presented with the instructions, which asked them to read and memorize a series of sentences and solve simple mathematical operations. They underwent a short practice session first, and afterwards, before starting the actual encoding trials, they were once again presented with the instructions. Each of the 32 sentences was presented for a total of 4.5 seconds and participants had 5 seconds to solve a math operation presented after each sentence.

Before beginning the Intermediate Phase, participants were asked to complete a distractor task, consisting of a "spot the differences" activity, for 5 minutes.

During the Intermediate Phase, participants were presented with new instructions, which differed depending on which conditions they were in (active vs. passive recognition; feedback vs. no feedback). In the active recognition group, participants were told they were going to see the same or similar sentences to the ones they had previously seen during encoding. Their task was to decide if the way the sentence was presented was correct (i.e., the exact same sentence as before) or incorrect (i.e., a similar but different sentence), by pressing one key for "correct" and another key for "incorrect". In the feedback condition, after responding, participants were provided with the correct answer in the form of the sentence they had seen during the encoding. In the no feedback condition, they were asked to solve a simple math operation. In the passive recognition group, participants saw the same sentences and answers of the participant in the active condition with whom they were paired with and were instructed to read them. Just like in the active recognition group, in the passive recognition group, those in the feedback condition were presented with the correct answer in the form of the original sentence, while those in the no feedback condition were asked to solve a simple mathematical operation. Participants in the active condition had a maximum of 10 seconds to provide their answer (but they could submit their answer after 6.5 seconds, if they wanted), while participants in the passive condition were presented with the sentences and answers of their yoked pairs for 6.5 seconds. Feedback, in the form of the correct sentence, was presented for 4.5 seconds for each sentence, and the participants in the mathematical operations condition also had 4.5 seconds to solve each problem.

Before beginning the Final Memory Test, participants were asked to complete another distractor task, consisting of a "spot the differences" activity, but this time they spent a total of 8 minutes on it.

Lastly, during the Final Memory Test, participants were presented with a cued-recall task and were instructed to correctly complete the missing information in the sentences presented, according to how they had first seen them during the Encoding Phase (e.g., "The baby ____ all night"). They were given 60 seconds to complete each sentence.

At the end, five self-report questions were asked about participants' attention and surroundings during the experimental process, to facilitate the exclusion of participants who might not have been paying attention or might have been completing the experiment in an inadequate place (e.g., very noisy and with many distractions). Participants were asked to answer the first ("Please rate your attention during the study") and the second ("Please rate the quality of your data") questions using a 7-point rating scale, where 1= "None" and 7= "Absolute", i.e., the participant performed the task with complete attention and considered their data to be of maximum quality. For the third question "Is there any reason why we should consider your data with caution? (For example, if you were listening to music during the study, interrupted to go to the bathroom, were simultaneously watching tv, or were on Facebook at the same time…)", participants had the option to answer affirmatively or negatively. If their answer was that we should consider their data with caution, they were presented with a box where they could write the reason why. Finally, in the last question, participants were asked if they had any comments they wanted to add and were provided with a box to do so.

Figure 2

Illustration of all Phases (Encoding, Intermediate, and Final Memory Test) of the Experimental Procedure

Condition	1st Encoding Phase	2nd Intermediate Phase Task	 Feedback	3rd Final Memory Test
Active recognition + Feedback	32 sentences of which 16 critical **" The baby stayed awake all night"** and 16 fillers **"The karate champion broke the cinder block"** Intercalated with math operations e.g., 25 + 4 =	**"The baby cried all night"** Correct or incorrect?	"The baby stayed awake all night"	32 cued-recall sentences **"The baby _____ all night"**
Active recognition + No Feedback			23 – 7 =	
Passive Recognition + Feedback		**"The baby cried all night"** Correct or incorrect. (Participants see the response from another participant in the active condition)	"The baby stayed awake all night"	
Passive Recognition + No Feedback			23 – 7 =	

Results

All statistical analysis were performed using IBM SPSS Statistics software version 26 (IBM Corp., Armonk, NY, USA). The dependent variable was participants' responses in the final cued-recall test for the 16 critical sentences. In order to decide if participants' responses were correct, pragmatic inference errors, intrusion errors or omissions, we used the standard scoring procedure proposed by Brewer (1977). For example, for the sentence "The baby stayed awake all night", the correct response would be "stayed awake", a pragmatic inference error would be "cried", an intrusion could be "slept", and no response would be considered an omission. A detailed explanation of what we considered a correct response and a pragmatic inference error for each sentence is available in Appendix B.

A one-way ANOVA revealed no differences in the overall math operations accuracy between the active recognition (*M* = .90, *SD* = .11) and passive recognition (*M* = .89, *SD* = .18) conditions, $F(1,56) = 0.06$, $p = .802$, $\eta^2_p = .001$.

Type of response

We performed a two-way ANOVA for each response type, using the mean of responses in the 16 critical sentences as the dependent variable, with the recognition group (active vs. passive) and feedback group (feedback vs. no feedback) as the two independent factors (for the descriptive statistics, see Table 2).

For the correct responses, we found a main effect of recognition type, $F(1,116) = 10.91$, $p = .001$, $\eta^2_p = .086$, revealing that participants in the active condition gave more correct responses (*M* = .50, *SD* = .28) than those in the passive condition (*M* = .37, *SD* = .27). We also found a main effect of feedback, $F(1,116) = 89.68$, $p < .001$, $\eta^2_p = .436$, showing that participants in the feedback condition (*M* = .61, *SD* = .26) presented more correct responses than those who did not receive feedback (*M* = .25, *SD* = .17). Furthermore, an effect of interaction between recognition and feedback was found, $F(1,116) = 4.42$, $p = .038$, $\eta^2_p = .037$. Post-hoc planned pairwise comparisons revealed that in the feedback condition, there was a significant difference between the active recognition and the passive recognition conditions ($F(1,60) = 11.32$, $p = .001$, $\eta^2_p = .159$), so participants in the active recognition plus feedback condition were significantly more accurate in their responses than participants in the passive recognition plus feedback condition. No significant difference between the active recognition and passive recognition conditions was observed for the no feedback condition ($F(1,56) = 1.09$, $p = .301$, $\eta^2_p = .019$).

For the pragmatic inference errors, we found a main effect of recognition, $F(1,116) = 7.89$, $p = .006$, $\eta^2_p = .064$, showing that participants in the active condition produced less pragmatic inference errors (*M* = .43, *SD* = .26) than those in the passive condition (*M* = .52, *SD* = .25). We also observed a main effect of feedback, $F(1,116) = 99.99$, $p < .001$, $\eta^2_p = .463$, revealing that participants in the feedback condition produced less pragmatic inference errors (*M* = .31, *SD* = .22) than those who did not receive feedback and were instead asked to solve mathematical operations (*M* = .65, *SD* = .15). Additionally, there was an interaction effect between recognition and feedback, $F(1,116) = 5.87$, $p = .017$, $\eta^2_p = .048$. Post-hoc pairwise planned comparisons revealed that, in the feedback condition, there was a significant difference between the active recognition and the passive recognition conditions ($F(1,60) = 11.08$, $p = .001$, $\eta^2_p = .156$), thus participants in the active recognition plus feedback condition produced less pragmatic inference errors than those in the passive recognition plus feedback condition. For the no feedback condition, no significant difference between active and passive recognition was observed ($F(1,56) = 0.10$, $p = .750$, $\eta^2_p = .002$).

No significant differences were found for intrusions (all *p*s > .279) and omissions (all *p*s > .052).

Table 2

Means and Standard Deviations of the Four Dependent Variables (Correct Responses, Pragmatic Inference Errors, Intrusions, and Omissions) Across the Four Experimental Conditions (Active and Passive Recognition With and Without Feedback)

	Active Recognition		**Passive Recognition**	
	Feedback	**No Feedback**	**Feedback**	**No Feedback**
Correct Responses	.71 (.19)	.27 (.15)	.51 (.28)	.23 (.18)
Pragmatic Inferences	.23 (.16)	.64 (.13)	.40 (.25)	.66 (.17)
Intrusions	.06 (.05)	.08 (.09)	.08 (.07)	.08 (.07)
Omissions	.01 (.04)	.00 (.02)	.01 (.03)	.04 (.09)

Persistency of Correct Responses, Persistency of Pragmatic Inference Errors and Error Correction Indexes

The analysis on the type of response only focused on the performance in the final test, regardless of what participants did or read during the intermediate phase. However, since our main interest is to investigate error correction, we created three indexes that take into account participants' performance during the intermediate phase.

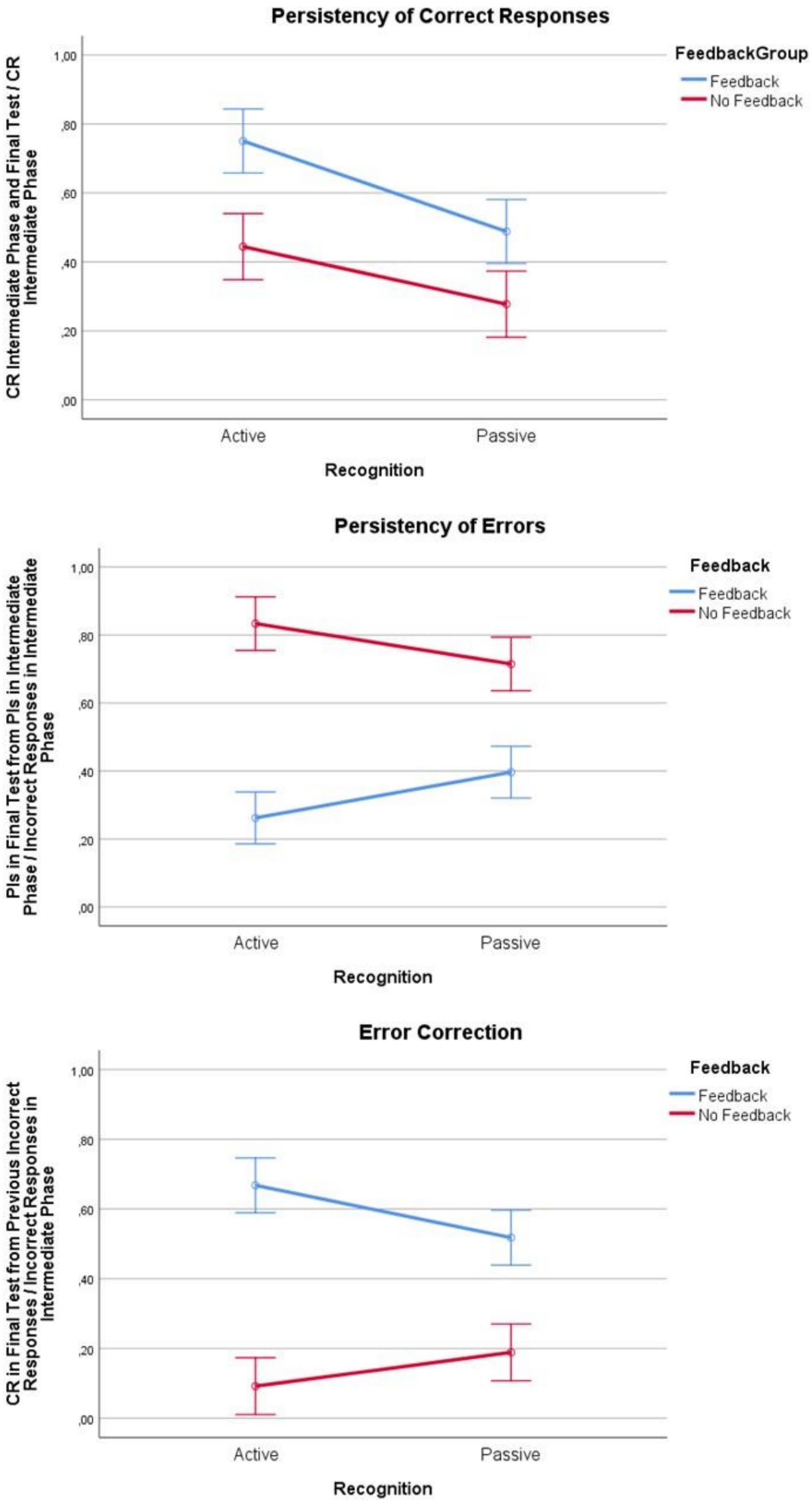
First, we derived an index of Persistency of Correct Responses, which stems from the sentences correctly recognized or presented in the intermediate phase that were also correctly recalled in the final memory test (see Figure 3, upper panel). This index was calculated by the sum of the correct responses in the final test correctly recognized or presented in the intermediate phase divided by the sum of the sentences correctly recognized or presented in the intermediate phase. We found a main effect of recognition, $F(1,116) = 20.66$, $p <.001$, $\eta^2_p = .151$, showing that participants in the active condition persisted more in their correct responses ($M = .60$, $SD = .28$) than participants in the passive condition ($M = .39$, $SD = .29$). We also found a main effect of feedback, $F(1,116) = 30.01$, $p <.001$, $\eta^2_p = .206$, revealing that participants in the feedback condition also persisted more in their correct responses ($M = .62$, $SD = .30$) than those in the no feedback condition ($M = .36$, $SD = .26$). We did not find an effect of interaction, $F(1,116) = 1.03$, $p =.313$, $\eta^2_p = .009$.

Second, we created an index that denotes the Persistency of Errors, which consisted of the pragmatic inference errors made or presented in the intermediate phase, that were again incorrectly recalled during the final memory test (see Figure 3, middle panel). This was calculated by the sum of pragmatic inferences recalled in the final memory test from pragmatic inferences made or presented in the intermediate phase divided by the sum of sentences incorrectly recognized or presented in the intermediate phase. There was no main effect of recognition, $F(1,116) = 0.04$, $p =.839$, $\eta^2_p <.001$, but we observed a main effect of feedback, $F(1,116) = 131.71$, $p <.001$, $\eta^2_p = .532$, revealing that participants in the no feedback condition ($M = .77$, $SD = .20$) persisted more in their pragmatic inference errors than participants in the feedback condition ($M = .33$, $SD = .24$). This effect was modulated by an interaction, $F(1,116) = 10.68$, $p = .001$, $\eta^2_p = .084$. Post-hoc pairwise planned comparisons indicated that in the feedback condition, there was a significant difference between the active and the passive recognition conditions ($F(1,60) = 5.37$, $p = .024$, $\eta^2_p = .082$), revealing that participants in the active recognition condition that received feedback persisted less in their errors than those in the passive recognition condition that also received feedback. A significant difference between the active recognition and passive recognition conditions was also observed in the no feedback condition ($F(1,56) = 5.48$, $p = .023$, $\eta^2_p = .089$), but in the opposite direction such that participants in the active recognition with no feedback persisted more in their errors than those in the passive no feedback condition.

Finally, we derived an index of Error Correction, which consisted of the sentences incorrectly recognized or presented in the intermediate phase, but correctly recalled in the final memory test (see Figure 3, lower panel). This was calculated by the sum of correct responses in the final test from incorrectly recognized or presented sentences in the intermediate phase divided by the sum of sentences incorrectly recognized or presented in the intermediate phase. No main effect of recognition was found, $F(1,116) = 0.44$, $p = .508$, $\eta^2_p = .004$, yet there was a main effect of feedback, $F(1,116) = 127.60$, $p < .001$, $\eta^2_p = .524$, such that participants in the feedback condition corrected their errors more ($M = .59$, $SD = .27$) than participants in the no feedback condition ($M = .14$, $SD = .17$). There was a significant interaction between recognition and feedback, $F(1,116) = 9.50$, $p = .003$, $\eta^2_p = .076$. Post-hoc pairwise planned comparisons revealed that, in the feedback condition, there was a significant difference between the active and passive recognition conditions ($F(1,60) = 5.06$, $p = .028$, $\eta^2_p = .078$), so participants in the active recognition condition corrected more errors than participants in the passive recognition condition. In contrast, in the no feedback condition, participants in the active recognition condition corrected less errors than those in the passive recognition condition, ($F(1,56) = 5.28$, $p = .025$, $\eta^2_p = .086$).

Figure 3

Mean Persistency of Correct Responses, Mean Persistency of Errors, and Mean Error Correction in the Active and Passive Recognition With and Without Feedback Conditions



Note. Error bars represent the Standard Error of the Mean.

Discussion

The purpose of this study was to investigate if the retrieval of errors stemming from pragmatic inferences, followed by corrective feedback, benefits learning.

Our results are in line with our hypotheses and with the recent literature on unsuccessful retrieval and error correction. We found main effects of retrieval condition (recognition) and feedback for both the correct responses and the pragmatic inference errors, revealing that participants in the active recognition condition (relative to those in the passive recognition condition) and participants in the feedback condition (compared to those in the no feedback condition) gave more correct responses and made fewer pragmatic inference errors on the final memory test. Of note, we also found interaction effects, indicating that in the feedback condition, participants in the active recognition group produced more correct responses and less pragmatic inference errors than those in the passive recognition group, whereas no such differences were observed in the no feedback condition.

Since the main goal was to understand if retrieving memory errors (pragmatic inference errors), followed by corrective feedback, could benefit learning, we have also analyzed participants' performance in terms of persistency of correct responses, persistency of pragmatic inference errors, and error correction. For the persistency of correct responses, our results revealed that participants in the active recognition condition were better at maintaining their correct responses in the final memory test than their counterparts in the passive recognition condition. The same pattern was observed for participants in the feedback condition relative to the no feedback condition. However, no interaction effect was found, suggesting that the retrieval effect is independent of the feedback condition for the persistency of correct responses.

For the purpose of establishing a comparison between the errorful learning and the errorless learning perspectives, we derived an index of persistency of pragmatic inference errors and found both an effect of feedback and a retrieval type by feedback interaction effect. Unlike what studies following an errorless learning approach have reported in the past (e.g., Warmington & Hitch, 2013), our results do not show that retrieving errors, followed by corrective feedback, makes them more prominent, persistent, and more difficult to correct. While it is true that in the no feedback condition, participants in the active recognition group persisted more in their pragmatic inference errors than those in the passive recognition group, in the feedback condition we found the opposite pattern. Participants in the active recognition plus feedback condition persisted less in their pragmatic inference errors than those in the passive recognition plus feedback condition. In terms of error correction, as expected, our results revealed that participants who received feedback were better at correcting their errors than those who did not. More interestingly, the interaction suggests that retrieval was only beneficial if followed by corrective feedback. In the no feedback condition, participants in the active recognition group corrected less pragmatic inference errors than those in the passive recognition group. The opposite happened in the feedback condition, in which participants in the active recognition group corrected more pragmatic inference errors than those in the passive recognition group. These findings support the benefit of retrieving errors followed by corrective feedback, proposed by the errorful learning perspective (overview in Metcalfe, 2017).

To the best of our knowledge, our study is the first to test the Memory Updating After Retrieval framework (Finn, 2017) using errors more representative of everyday experiences, and to present results that support that framework with a more ecological study design. Our data shows that participants who engaged in retrieval during the intermediate phase (active recognition condition) were more prone to correct their errors in the final memory test than those who were merely asked to read the answers from another participant (passive recognition condition), as long as corrective feedback was provided. This is in line with Finn's (2017) proposal that retrieval benefits learning because it makes memory more malleable to the incorporation of new information. Applying it to the results of this study, as well as error correction in general, retrieval of an error makes memory more pliable and prone to the incorporation of corrective feedback, which could explain why participants in the active recognition plus feedback condition persisted less and corrected more pragmatic inference errors than participants in the passive recognition plus feedback condition, but the same did not happen when no feedback was provided.

Furthermore, instead of using materials to elicit guessing errors, we had participants generate pragmatic inference errors, considering that this type of error is frequently committed in everyday situations, such as social interactions. This choice allowed us to increase the ecological validity of our results. Unlike previous studies (e.g., Kornell et al., 2009) that measured guessing errors, our study measured errors retrieved from memory, using a retrieval condition (active recognition). Even though it is harder to elicit a high number of errors when compared to a guessing methodology, our participants still frequently incorrectly remembered the sentences presented. This happened most likely because even though our participants were asked to retrieve unique episodic events, their semantic knowledge might have interfered in the retrieval process. For example, when they were asked to retrieve "The baby stayed awake all night", their semantic knowledge that babies cry during the night might have led them to infer that the baby was crying, despite that not being explicit in the sentence presented, leading them to produce an error.

We were very careful with our design and methodology, which guaranteed a high level of methodological control, by using a yoked design and the counterbalance of the materials as a function of both the format of the sentence at encoding and at the intermediate phase. Another strength of our study is the fact that, in the intermediate phase, we used a recognition task, which is different from our final cued-recall task, preventing any potential effects of practicing in our results. Although in the passive recognition condition our task was not truly a recognition task (the participants were not asked or tested on whether they recognized the sentences), we tried to match it as much as possible to the active recognition condition, asking our participants to read not only the sentences presented, but also the answers that their yoked pair in the active recognition condition gave to each sentence. Prior studies (e.g., Kornell et al., 2009) did not do this, as their participants in the restudy condition simply read the sentences presented, which were all correct sentences. For that reason, unlike in our study, there was not a match of correct and incorrect sentences across participants in the retrieval and restudy groups.

The current study has some limitations worth acknowledging. We did not ask our participants about their educational level (e.g., number of schooling years), so we had no information on this indicator. Since we are studying the effect of retrieval and error correction in learning, differences in formal education could potentially have implications on our results. Thus, it would be very interesting and relevant, in future studies, to explore the potential existence of differences between participants of different educational levels. Furthermore, it would also be relevant to investigate potential differences in developmental stages, such as between kindergarten-level children, older children, adolescents, and adults. Several cognitive abilities are acquired during childhood, but during adolescence some of those abilities suffer a process of refinement and maturation

(Luna, 2009). Organic changes and refinements across the brain (i.e., synaptic pruning and increase in myelination) during adolescence (Huttenlocher, 1990; Wozniak & Lim, 2006), support the integration of information and promote the improvement of high-order cognitive processes (Goldman-Rakic, 1988), such as executive functions. These executive functions enable the cognitive and voluntary control of behavior, including response planning and preparation, response inhibition, and working memory, which all support cognitive flexibility and abstract thought (Luna, 2009). Since during adolescence these abilities are still maturing, adolescents are limited when it comes to having a consistent performance, which might leave them vulnerable to making errors (Luna, 2009). This limitation in the efficacy of their executive processes could have implications in their error correction ability, considering it may require the engagement of these processes (e.g., inhibitory control and error monitoring) in order to be successful. Children, due to the higher immaturity of their inhibitory control capabilities, are more susceptible to interference from distractors than adults (Bjorklund & Harnishfeger, 1990; Luna, 2009), which could also have an impact on their error correction ability, especially in younger ages. There are already studies looking into error correction in children of different ages (e.g., Carneiro et al., 2018), but considering that the brain undergoes many important alterations and sophistications not only between childhood and adolescence, but also between adolescence and adulthood, it would be even more interesting if future studies could compare these pairs, using the same methodology. The comparison between adolescents and adults would be particularly interesting, not only because of the brain maturation processes underlying the gap between them, but also because that transition is thought to leave adolescents vulnerable to impaired development (Luna, 2009).

Recently, research into the possible application of the testing effect to clinical settings has been increasing, notably using retrieval practice in patients with memory impairments, such as patients with schizophrenia (Jantzi et al., 2019), multiple sclerosis (Sumowski et al., 2013), and children and adults recovering from a traumatic brain injury (Coyne et al., 2015; Sumowski et al., 2014). All these studies revealed long-term episodic memory improvements following retrieval practice. In patients with schizophrenia, retrieval (followed by feedback) has been found to efficiently improve episodic memory when compared to restudy, suggesting that this learning strategy should be considered for cognitive rehabilitation programs (Jantzi et al., 2019). Furthermore, the authors propose that this memory performance enhancement might occur because retrieval practice promotes semantic elaboration, and patients with schizophrenia present mostly with difficulty to self-initiate semantic encoding strategies. However, most studies have used retrieval practice associative paradigms, such as verbal paired weak associates (e.g., ground-cold; Sumowski et al., 2014), which are not very representative of the type of stimuli patients might frequently struggle to remember and learn in their daily lives. Even though our study was conducted with healthy participants, the results suggest that it could be interesting for future research to use a pragmatic inferences paradigm with patients with memory impairment, since pragmatic inferences are obstacles that they might frequently find in their everyday social interactions (McDermott & Chan, 2006). Our participants in the active recognition plus feedback condition (the most similar to previous studies) produced the highest number of correct responses on the final memory test. Thus, it would be useful to establish if active retrieval followed by feedback also promotes patient performance in future studies using pragmatic inferences.

Our findings have relevant implications to educational settings. They suggest that students should have an active role in their learning process, rather than being passively involved in the classroom, since it appears to be more beneficial for them. The testing effect, which has been demonstrated to benefit learning numerous times (e.g., Carpenter et al., 2008; overview in Pyc et al., 2014, and Roediger & Butler, 2011) has already clued us to the advantages of involving students in their learning process, through providing them with more opportunities to test their knowledge of the materials taught. However, the latest findings in the literature, including the ones from the current study, suggest that that involvement should be taken further. In class, if students are given the opportunity to discuss amongst themselves and explore a new topic before it is taught, they will most likely make errors along the way, which the teacher, while guiding their discussion, can correct. Since retrieval of errors, followed by corrective feedback, has been demonstrated in this study to be beneficial for learning, we suggest that more dynamic learning activities could be adopted in classrooms. With that in mind, it would be interesting for future research to study not only the applicability, but also the feasibility of applying teaching techniques, based on an errorful learning approach, to real classrooms of different educational levels, taking into consideration possible developmental changes.

Conclusion

The results of the current study are in line with recent research on the errorful learning approach. They indicate that participants present more correct responses and make less errors when they are given the opportunity to actively retrieve the information they encoded relative to a passive condition, as long as retrieval is followed by corrective feedback. Of note, retrieval followed by feedback allows participants to persist less in their errors and promotes error correction. These findings support Finn's (2017) Memory Updating After Retrieval proposal by revealing that after retrieval, memory is more permeable to incorporate new information, providing an opportunity for a correct update of the stored information. While these results are promising and suggest that students would benefit from having a more active role in their learning process, through the adoption of errorful learning strategies, future research should focus on assessing the feasibility of applying such strategies to real classrooms, while taking into consideration developmental factors.

References

Assefi, S. L., & Gary, M. (2003). Absolut memory distortions: Alcohol placebos influence the misinformation effect. *Psychological Science, 14*(1), 77-80. https://doi.org/10.1111/1467-9280.01422

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge University Press.

Bjork, E. L., & Bjork, R. A. (2011*).* Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, J. R. Pomerantz (Eds.) & FABBS Foundation, *Psychology and the real world: Essays illustrating fundamental contributions to society* (p. 56–64). Worth Publishers.

Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review, 10*(1), 48-71. https://doi.org/10.1016/02732297(90)90004-N

Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition, 5*(6), 673-678. https://doi.org/10.3758/BF03197414

Burns, D. J., Jenkins, C. L., & Dean, E. E. (2007). Falsely recalled items are rich in item-specific information. *Memory & Cognition, 35*(7), 1630-1640. https://doi.org/10.3758/BF03193497

Burns, D. J., Martens, N. J., Bertoni, A. A., Sweeney, E. J., & Lividini, M. D. (2006). An item gains and losses analysis of false memories suggests critical items receive more item-specific processing than list items. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*(2), 277-289. https://doi.org/10.1037/0278-7393.32.2.277

Butler, A. C., Fazio, L. K., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin & Review, 18*, 1238-1244. https://doi.org/10.3758/s13423-011-0173-y

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1491-1494. https://doi.org/10.1037//0278-7393.27.6.1491

Carneiro, P., Lapa, A., & Finn, B. (2018). The effect of unsuccessful retrieval on children's subsequent learning. *Journal of Experimental Child Psychology, 166*, 400-420. https://doi.org/10.1016/j.jecp.2017.09.010

Carneiro, P., Lapa, A., & Finn, B. (2020). *Memory updating after retrieval: When new information is false or correct.* [Manuscript submitted for publication]. Faculty of Psychology, University of Lisbon.

Carneiro, P., Lapa, A., Reis, J., & Ramos, T. (2020). Testing pragmatic inferences: The impact of language and culture. *Psicológica, 41*(1), 1-20. https://doi.org/10.2478/psicolj-2020-0001

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563-1569. https://doi.org/10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1547-1552. https://doi.org/10.1037/a0024140

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*, 438-448. https://doi.org/10.3758/MC.36.2.438

Chan, J. C. K., & McDermott, K. B. (2006). Remembering pragmatic inferences. *Applied Cognitive Psychology*, *20*(5), 633-639. https://doi.org/10.1002/acp.1215

Coyne, J. H., Borg, J. M., DeLuca, J., Glass, L., & Sumowski, J. F. (2015). Retrieval practice as an effective memory strategy in children and adolescents with traumatic brain injury. *Archives of Physical Medicine and Rehabililitation*, *96*(4), 742-745. https://doi.org/10.1016/j.apmr.2014.09.022

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17-22. https://doi.org/10.1037/h0046671

Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 813-825. https://doi.org/10.1037/0278-7393.29.5.813

Finn, B. (2017). A framework of episodic updating: An account of memory updating after retrieval. *Psychology of Learning and Motivation*, *67*, 173-211. https://doi.org/10.1016/bs.plm.2017.03.006

Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory & Cognition*, *38*(7), 951-961. https://doi.org/10.3758/MC.38.7.951

Goldman-Rakic, P. S. (1988). Topography of cognition: Parallel distributed networks in primate association cortex. *Annual Review of Neuroscience, 11*, 137-156. https://doi.org/10.1146/annurev.ne.11.030188.001033

Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 207-219. https://doi.org/10.1016/S0022-5371(82)90571-0

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, 40, 505-513. https://doi.org/10.3758/s13421-011-0174-0

Gunter, R. W., Ivanko, S. L., & Bodner, G. E. (2005). Can test list context manipulations improve recognition accuracy in the DRM paradigm? *Memory*, *13*(8), 862-873. https://doi.org/10.1080/09658210444000458

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*(4), 514-527. https://doi.org/10.3758/s13421-011-0167-z

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia*, *28*(6), 517-527. https://doi.org/ 10.1016/0028-3932(90)90031-i

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83*(2), 340-344. https://doi.org/10.1037/h0028541

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgements and cued recall. *Memory & Cognition*, *41*, 625-637. https://doi.org/10.3758/s13421-013-0298-5

Jantzi, C., Mengin, A. C., Serfaty, D., Bacon, E., Elowe, J., Severac, F., Meyer, N., Berna, F., & Vidailhet, P. (2019). Retrieval practice improves memory in patients with schizophrenia: New perspectives for cognitive remediation. *BMC Psychiatry, 19*(355), 1-12. https://doi.org/10.1186/s12888-019-2341-y

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., & Carpenter, S. K. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology, 103*(1), 48-59. https://doi.org/10.1037/a0021977

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45-83. https://doi.org/10.1080/10508406.2011.591717

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989-998. https://doi.org/10.1037/a0015729

Kornell, N., & Metcalfe, J. (2014). The effects of memory retrieval, errors and feedback on learning. In V. A. Benassi, C. E. Overson & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (225-251). Division 2, American Psychological Association.

Lee, J. L. C. (2008). Memory reconsolidation mediates the strengthening of memories by additional learning. *Nature Neuroscience, 11*(11), 1264-1266. https://doi.org/10.1038/nn.2205

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*(4), 361-366. https://doi.org/10.1101/lm.94705

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human, Learning and Memory, 4*(1), 19-31. https://doi.org/10.1037//0278-7393.4.1.19

Luna, B. (2009). Developmental changes in cognitive control through adolescence. *Advances in Child Development and Behavior*, *37*, 233-278. https://doi.org/10.1016/s0065-2407(09)03706-9

McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, *34*(2), 261-267. https://doi.org/10.3758/BF03193404

McDermott, K. B., & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition, 34*(6), 1273-1284. https://doi.org/10.3758/BF03193271

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology, 68*, 465-489. https://doi.org/10.1146/annurev-psych-010416-044022

Metcalfe, J., Casal-Roscum, L., Radin, A., & Friedman, D. (2015). On teaching old dogs new tricks. *Psychological Science, 26*(12), 1833-1842. https://doi.org/10.1177/0956797615597912

Metcalfe, J., & Finn, B. (2011). People's hypercorrection of high confidence errors: Did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 437-448. https://doi.org/10.1037/a0021962

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributed to episodic recollection rather than semantic mediation. *Neuropsychologia*, *138*, 1-8. https://doi.org/10.1016/j.neuropsychologia.2019.107296

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, *37*(8), 1077-1087. https://doi.org/10.3758/MC.37.8.1077

Metcalfe, J., & Miele, D. B. (2014). Hypercorrection of high confidence errors: Prior testing both enhances delayed performance and blocks the return of errors. *Journal of Applied Research in Memory and Cognition, 3,* 189-197. https://doi.org//10.1016/j.jarmac.2014.04.001

Mullet, H. G., & Marsh, E. J. (2016). Correcting false memories: Errors must be noticed and replaced. *Memory & Cognition*, *44*(3), 403-412. https://doi.org/10.3758/s13421-015-0571-x

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3-8. https://doi.org/10.1037/0278-7393.31.1.3

Pastötter, B., & Bäuml, K-H. T. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology, 5*(286), 1-5. https://doi.org/10.3389/fpsyg.2014.00286

Patihis, L., Frenda, S. J., LePort, A. K. R., Petersen, N., Nichols, R. M., Stark, C. E. L., McGaugh, J. L., & Loftus, E. F. (2013). False memories in highly superior autobiographical memory individuals. *PNAS, 110*(52), 20947-20952. https://doi.org/10.1073/pnas.1314373110

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644-667. https://doi.org/10.1037/a0033194

Pyc, M. A., Agarwal, P. K., & Roediger, H. L. (2014). Test-enhanced learning. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 78-90). American Psychological Association Society for the Teaching of Psychology.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*(6002), 335. https://doi.org/10.1126/science.1191465

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15*(3), 243-257. https://doi.org/10.1037/a0016496

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20-27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803-814. https://doi.org/10.1037/0278-7393.21.4.803

Sampaio, C., & Brewer, W. F. (2009). The role of unconscious memory errors in judgements of confidence for sentence recognition. *Memory & Cognition*, *37*(2), 158-163. https://doi.org/10.3758/MC.37.2.158

Schiller, D., Monfils, M-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, *463*(7277), 49-53. https://doi.org/10.1038/nature08637

Skinner, B. F. (1953). *Science and human behaviour.* Macmillan.

Sumowski, J. F., Coyne, J., Cohen, A., & DeLuca, J. (2014). Retrieval practice improves memory in survivors of severe traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *95*(2), 397-400. https://doi.org/10.1016/j.apmr.2013.10.021

Sumowski, J. F., Leavitt, V. M., Cohen, A., Paxton, J., Chiaravalloti, N. D., & DeLuca, J. (2013). Retrieval practice is a robust memory aid for memory-impaired patients with MS. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, *19*(14), 1943-1946. https://doi.org/10.1177/1352458513485980

Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition*, *14*(4), 329-338. https://doi.org/10.3758/BF03202511

Wade, K. A., Garry, M., & Pezdek, K. (2018). Deconstructing rich false memories of committing crime: Commentary on Shaw and Porter (2015). *Psychological Science, 29*(3), 471-476. https://doi.org/10.1177/0956797617703667

Warmington, M., & Hitch, G. J. (2013). Enhancing the learning of new words using an errorless learning procedure: Evidence from typical adults. *Memory, 22*(5), 582-594. https://doi.org/10.1080/09658211.2013.807841

Wozniak, J. R., & Lim, K. O. (2006). Advances in white matter imaging: A review of in vivo magnetic resonance methodologies and their applicability to the study of development and aging. *Neuroscience & Biobehavioral Reviews, 30*(6), 762-774. https://doi.org/10.1016/j.neubiorev.2006.06.003

Appendix A

Sentences Presented in the Experiment

This appendix consists of 32 sentences that were first presented during the Encoding Phase, and then presented again during the Intermediate Phase. The words in **bold** are part of the critical form of the sentences, while the words in brackets are part of the filler form of the sentences. The underlined words are the ones that were replaced by blank spaces during the final cued-recall test.

1. O ágil gato **alcançou** (apanhou) o peixe com as patas.
2. A meio da noite o marido sonolento foi buscar o jornal e **acertou no** (matou o) mosquito.
3. Depois de deixar os filhos na escola, a mãe foi **buscar** (comprar) pão.
4. O hipnotizador **juntou** (estalou) os dedos e acordou o cliente.
5. A corrida começou quando o árbitro **pressionou o gatilho** (apitou).
6. O barulhento cão de guarda **rosnou** (ladrou) ao homem que passava na rua.
7. O distraído professor **não tinha as** (esqueceu-se das) chaves do carro.
8. O cristão fecha os olhos e **fica uns minutos em silêncio** (reza) antes de cada refeição.
9. O campeão de karaté **bateu no** (partiu o) bloco de cimento.
10. Ela **pegou no** (calçou o) seu par de sapatos preferido e saiu.
11. O esquilo esfomeado estava a **roer** (comer) a relva.
12. Ele **pegou no** (atendeu o) telefone, ao ver que era a mãe quem lhe estava a ligar.
13. A jiboia esfomeada **apanhou** (comeu) o rato.
14. Tudo o que o Ricardo queria era poder **descansar** (dormir) uma noite inteira.
15. O ladrão **tirou** (acendeu) o isqueiro para fazer detonar o explosivo.
16. O vândalo **fez pontaria com** (atirou) a pedra à janela.
17. No dia da entrevista, ela **escolheu** (vestiu) a sua camisa da sorte.
18. O bebé **esteve acordado** (chorou) toda a noite.
19. A criança **picou** (rebentou) o balão com um alfinete.
20. O guarda-noturno **tirou** (bebeu) café do seu termo.
21. O King Kong **esteve no alto do** (subiu o) Empire State Building.
22. O príncipe encantado, gentilmente, **baixou-se em direção ao** (beijou o) rosto da Branca de Neve.

23. Dennis, o Pimentinha, sentou-se **na cadeira** (no colo) do Pai Natal e pediu um elefante.
24. Aquela estação de rádio só **gostava de** (passa) música hard rock.
25. O ladrão entrou numa loja e **pegou num** (roubou um) chocolate.
26. Depois de perseguir o ladrão por três quarteirões, o polícia finalmente **alcançou** (apanhou) o ladrão.
27. O boneco de neve **desapareceu** (derreteu) quando a temperatura atingiu os 26ºC.
28. O simpático empregado recebeu **umas moedas** (gorjeta) do cliente.
29. Assim que chegou à praia, a Marta estendeu a toalha, pôs **o chapéu de sol** (protetor solar) e deitou-se na areia.
30. O Capitão do submarino **disse** (gritou/ordenou): - "Submergir submarino!"
31. Ela **tropeçou** (caiu) ao descer as escadas.
32. O rato foi **atraído** (apanhado) pela ratoeira.

Appendix B

Responses That We Considered Correct Responses and Pragmatic Inference Errors For Each Sentence

Participants' responses that we considered correct responses and pragmatic inference errors for each of the 32 sentences, following Brewer's (1977) standard scoring procedure, on the final cued-recall memory test.

Sentence Number	Responses we assumed as **correct**		Responses we assumed as **pragmatic inference errors**	
	Critical	**Filler**	**Critical**	**Filler**
1	Alcançou	apanhou, agarrou, caçou, prendeu	apanhou, agarrou	alcançou
2	acertou no	matou o, matou, matar	matou o, matou, finalmente matou o	acertou no, bateu no
3	buscar, buscar o	comprar, comprar o	comprar, comprar o, à mercearia e comprou, comprar algum	buscar, buscar o, foi buscar o
4	juntou, juntou os	Estalou	estalou	juntou
5	pressionou o gatilho, disparou o gatilho, apertou o gatilho, puxou o gatilho, premiu o gatilho, carregou no botão	apitou, soprou o apito, apita, assobiou no apito, soou o apito	apitou, tocou o apito, soprou o apito	ressionou o gatilho, apertou o gatilho
6	Rosnou	ladrou, ladrava, latiu	ladrou, começou a ladrar	rosnou

7	não tinha as, não tinha, não trouxe as, não tem as	esqueceu-se das, esqueceu as, esqueceu-se, esqueceu das	esqueceu-se das, esqueceu as	não tinha as
8	fica uns minutos em silêncio, faz silêncio, faz silêncio uns minutos, fica em silêncio, fica em silêncio alguns minutos, fica em silêncio uns segundos, fica em silêncio um minuto, fica uns momentos em silêncio, fica um tempo em silêncio, fica calado, fica calado um tempo, espera uns momentos em silêncio	reza, rezou, reza em silêncio, ora por uns minutos,	reza, faz uma oração, reza uns minutos,	fica uns minutos em silêncio, fica em silêncio
9	bateu no, bateu num, bateu, atingiu, acertou no, bateu com força no	partiu o, partiu no, partiu um, partiu, quebrou o, quebrou um	partiu o, partiu um, partiu, partiu com as mãos um, quebrou, quebrou o	bateu no, bateu
10	pegou no, pegou, apanhou	calçou o, calçou os, calçou, usou, usou o, vestiu	calçou o, calçou, vestiu o, pôs, colocou o	pegou no

11	roer, roeu, morder	comer, a comer, devorar	comer, a comer, devorar	roer
12	pegou no, ela pegou no, pegou o, pegou, agarrou, Dennis pegou no	atendeu o, atendeu, atende o	atendeu o, atendeu	pegou no, pegou
13	Apanhou	comeu, comeu o, devorou, engoliu	comeu, engoliu	apanhou, caçou
14	Descansar	dormir, dormir uma	dormir, dormir sossegado, dormir descansado	descansar
15	tirou, pegou, pegou no, atirou	acendeu, acendeu o, acende	acendeu, ligou o	tirou
16	fez pontaria com, fez pontaria, apontou	atirou, mandou, jogou, lançou	atirou, atira, mandou, lançou	fez pontaria com
17	escolheu, escolheu vestir	vestiu, utilizou, usou	vestiu, usou, utilizou	escolheu
18	esteve acordado, ficou acordado, não dormiu	chorou, esteve a chorar, dormiu	chorou, chorou a, chorou durante, chora	esteve acordado, ficou acordado, não dormiu
19	picou, espetou	rebentou	rebentou, rebenta, arrebentou, explodiu, estourou	picou, espetou

20	tirou, tirou o, tirou um, retirou, serviu, dispensou, pegou	bebeu, bebeu o, bebia, tomou o	bebeu, bebeu o, bebeu todo o, bebe, tomou	tirou, tirou o
21	esteve no alto do, estava no alto do, esteve no cimo do, esteve no topo do, estava no cimo do, estava no topo do, estava no, esteve em cima do, esteve no, foi ao topo do, chegou ao cimo do, esteve no ponto mais alto do	Subiu o, subiu ao, subiu, escalou o, escalou, trepou o	subiu o, subiu ao, subiu ao topo do, subiu ao topo da, subiu ao cimo do, subiu ao ponto mais alto do, subiu, trepou o, escalou o	esteve no alto do, esteve em cima
22	baixou-se em direção ao, baixou a cabeça ao, baixou a cabeça em direção ao, baixou-se até ao, baixou-se sobre o, baixou-se para, baixou-se ao nível do, baixou o rosto até ao, abaixou-se até ao, aproximou-se do, aproximou-se no, aproximou o, se dirigiu em direção ao,	beijou o, beijou, deu um beijo no	beijou o, beijou, aproximou-se e beijou o, baixou-se e beijou	baixou-se em direção ao, aproximou-se do

	inclinou-se em direção ao, direcionou-se para o			
23	na cadeira, na poltrona	no colo, no colo do, ao colo, para o colo	no colo, no colo do, ao colo	na cadeira, cadeira
24	gostava de, gosta de, gosta, prefere	passa, passava, toca, transmite, dava	passa, passava, tocava, toca	gostava de, gosta de
25	pegou num, pegou no, pegou um, pegou, tirou o, tirou um, apanhou um	roubou um, roubo o, roubou	roubou um, roubou o, roubou	pegou num, pegou no, tirou um
26	alcançou, encontrou	apanhou, apanhou-o, agarrou, capturou	apanhou, apanhou o, capturou, pegou, conseguiu apanhar	alcançou, encontrou
27	desapareceu	derreteu, começou a derreter, descongelou	derreteu, derreteu-se, começou a derreter, desfez-se	Desapareceu
28	umas moedas, as moedas, algumas moedas, moedas, dinheiro	gorjeta, a gorjeta, uma gorjeta, uma boa gorjeta	gorjeta, uma gorjeta, a gorjeta	umas moedas, moedas, odinheiro, dinheiro

29	o chapéu de sol, o guarda-sol, guarda-sol, debaixo do chapéu de sol, o chapéu na areia	o protetor solar, protetor solar, protetor, creme de proteção solar, creme, creme solar	o protetor solar, o protetor, protetor, protetor solar, o creme	o chapéu de sol, o chapéu
30	disse	gritou, ordenou	gritou, ordenou	Disse
31	tropeçou	caiu	caiu	Tropeçou
32	atraído, foi atraído	apanhado, foi apanhado, caçado, entalado, preso	apanhado, caçado, capturado	Atraído