Beyond Mediator Retrievals: Charting the Path by Which Errors Lead to Better Memory Consolidation

Bridget Leonard (bll313@uw.edu)
Department of Psychology, University of Washington, Seattle, WA 98195 USA

Holly Sue Hake (hakehs@uw.edu)
Department of Psychology, University of Washington, Seattle, WA 98195 USA

Andrea Stocco (stocco@uw.edu)
Department of Psychology, University of Washington, Seattle, WA 98195 USA

Abstract

Expanding on previous research highlighting the learning benefits of errors, this study explores the enduring effects of error-induced learning. Using an adaptive fact-learning system, 23 participants engaged in recognition, recall, and error tasks, with repeated testing for memory assessment. Initial findings echoed previous results: items learned through errors initially took longer to retrieve. However, a significant shift occurred over time; error items demonstrated faster retrieval speeds compared to study items, and, most notably, they exhibited greater resilience against forgetting. This study reaffirms the positive role of errors in learning and uncovers their contribution to enhanced long-term memory retention. These insights challenge traditional learning paradigms, advocating for an educational approach that recognizes and leverages the value of errors in learning processes.

Keywords: errors; forgetting; memory; retrieval; mediation; computational modeling

Introduction

Previous studies have conclusively shown that errors lead to better memory. Error learning or improvements in recall on subsequent fact presentations after an incorrect answer or memory error. has been demonstrated across types of memory (recall vs recognition: Greving & Richter, 2018) and different types of facts (math: Kapur & Bielaczyc, 2012; trivia: Kornell, 2014). The foundation of this increased learning appears to stem from additional processing that occurs between the onset of a question and the presentation of the answer (i.e., a retrieval attempt) that leads to better answer processing, regardless of whether the attempt was successful (Kornell & Vaughn, 2016).

Two main theories of the mechanics underlying this extra processing exist. The elaborative theory of error learning states that generating an answer activates a semantic network of items related to the prompt-answer pairing. After receiving feedback, the initial (erroneous) response and the correct one are merged to form a more substantial memory trace that has more connections to the original probe and is this more likely to be retrieved in the future after prompt presentations (Huelser & Metcalfe, 2012; Karpicke, 2017). For example, in a simple word association task where participants must learn cue-target pairs, one may generate the word “tail” as a free associate in response to the cue word “whale” when the target word is actually “swims.” Instead of simply encoding the pair “whale” and “swims,” the individual may use the error to create a more robust network between the two words, perhaps thinking of a *whale* using its *tail* for *swimming*.

The *mediator* theory, on the other hand, argues for a more episodic account of error learning. Errors are specifically encoded and used as a secondary cue to the prompt in future presentations. This helps learners recall facts since they can use both the prompt and error to retrieve an answer (Huelser & Metcalfe, 2012; Mera et al., 2021). Referring to the previous example, at subsequent presentations of the word “whale,” one may recall their previous error, “tail,” and from it, attempt a second retrieval and recall the correct target word, “swims.”

In a recent study, Leonard et al. (2023) provided strong evidence in favor of the mediator theory. To investigate the cognitive mechanisms behind the well-documented theories, they developed models within a formal computational framework, specifically Anderson and Schooler’s (1991) model, which is now a part of the Adaptive Control of Thought–Rational (ACT-R) cognitive architecture (Anderson, 2007). In these models, memories in declarative memory are represented as “chunks” in a semantic network. Memory retrieval is a dynamic process in which memories matching a retrieval request compete for selection based on their activation, a combination of the recency and frequency of that memory’s presentation as well as its relation to the current context, or spreading activation.

The elaborative hypothesis was modeled in this architecture by creating elaborative error chunks. When an error is committed, and feedback is provided, chunks linking the cue and target words merge with chunks containing the cue and error words to form one elaborative chunk. In the previous paired-associate example, *whale-tail* would be merged with *whale-swims* to create a chunk: *whale-tail-whale-swims*. This chunk could represent the previously discussed meaningful links between cue and target words (i.e., the whale swims with its tail) or simply *whale-swims*, not *whale-tail*. Subsequent presentations of the cue spread more activation to this elaborative chunk;

multiple references of the cue word within the chunk increase their strength of association. This leads to both enhanced fact recall and faster response times than study items (Figure 1A).

The mediator hypothesis was modeled by adding an extra cognitive step in the model that specifically seeks out if a past error has been made. If this retrieval request is successful, it then uses the error to retrieve the correct answer. For example, *tail* is retrieved when the cue *whale* is presented, the learner remembers that *tail* is an error and uses it to help retrieve the correct target *swims*. This enhances the specificity and success of fact recall and importantly, increases response times due to extra time spent retrieving memories compared to study items (Figure 1B).

Leonard and colleagues used the models to mathematically prove that, under reasonable assumptions, the elaborative hypothesis would lead to shorter RTs for error items, while the mediator hypothesis would lead to longer RTs. To test their prediction, they also collected data from 61 participants; by analyzing the data with a random-slope mixed linear model, they were able to show that no single participant adopted the elaborative strategy. They also used maximum likelihood to fit the original ACT-R models to individual participant data on a trial-by-trial basis, showing again that all participants were more likely to be fit by the mediator model.

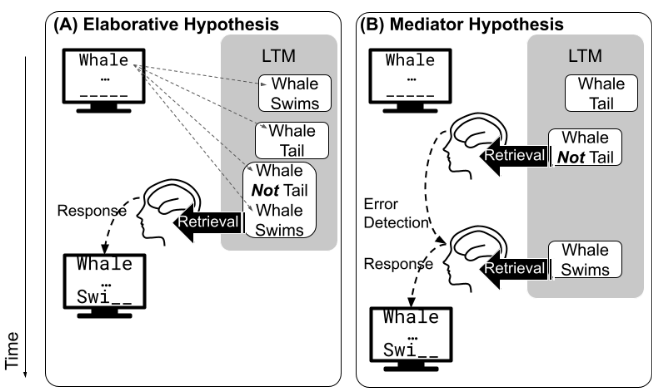

Figure 1: Mechanisms of error learning. (A) Errors provide increased learning by adding extra spreading activation during retrieval due to the memory's encoding in a deep semantic network. (B) Errors provide increased learning by acting as a secondary cue, adding an extra cognitive step that enhances retrieval specificity.

Limits of Existing Theories

One important limitation of previous work is that it does not explore the dynamics of trial learning, that is, they do not consider the response times and accuracies of items beyond the first trial after an error is made.

This leaves several possibilities open. It is possible that, after an error is made, mediator retrievals are consistently used ever after in the process of retrieving a memory. Alternatively, the mediator hypothesis might better explain the first few trials after a mistake was made, but, with repeated experience, these additional retrievals might lead to greater associations between the cue and the probe, thus leading to a long-term result that is compatible with the elaborative hypothesis. Or, it is possible that, while mediator retrievals are dropped after the first few trials, their effects persist over time in terms of reduced forgetting. This could be due, for example, to the increased depth of processing (Craik & Tulving, 1975) or to the greater motivational engagement associated with generating an answer (Shohamy & Adcok, 2010).

To distinguish between these alternative scenarios, it is necessary to run a different type of experiment, in which initial errors are followed for multiple trials and quantitative measures of memorability of each fact are tracked over multiple trials. Evidence for the mediator effect would manifest as persistently longer response times to error items after the first test.

Shorter response times after the first error, however, could be explained by either the creation of additional associative links, which would facilitate retrieval as posited by the elaborative hypothesis, by the creation of longer-lasting traces, which would lead to more resistant traces due to additional processing or arousal. To distinguish between these two explanations, it is necessary to estimate a memory's intrinsic tendency to be forgotten over time from the ease with which it can be retrieved when probed.

To dissociate the two, we used an adaptive fact-learning system (AFLS) testing paradigm. The AFLS presents testing probes adaptively, at a rate that is person- and item-specific, so that the presentation of each item is spaced to maintain a near-constant accuracy rate. Specifically, we employed a model-based AFLS, in which the spacing between different items is computed based on the parameters of an underlying computational model of memory, which is iteratively refined to match the participant's response and works as their "cognitive twin". Importantly, this approach allows us to analyze not only response times but also the latent model parameters for each item.

Model-Based Memory Assessment

To evaluate memory, we adopt a unique approach that integrates Anderson and Schooler's (1991) episodic memory model into the Adaptive Control of Thought–Rational (ACT-R) cognitive framework. Grounded in the Multiple Trace Theory (Nadel et al., 2000), which posits the creation of distinct memory traces with each encounter of new information, our method aligns with the power law of forgetting (Newell & Rosenbloom, 1982). This law suggests that these traces gradually diminish over time.

The model calculates the probability of recalling a memory m at a time t, based on its activation level $A(m, t)$. This activation level is the logarithmic sum of probabilities of accessing each trace, as expressed in the formula:

$$A(m, t) = \log \sum_i (t - t(i))^{-d(i)} \qquad (1)$$

In this equation, $t(i)$ signifies when the i-th trace was formed, and $d(i)$ is its decay rate. The decay rate is determined by the memory's residual activation at the time the trace was formed, linking it to the memory retention spacing effect (Cepeda et al., 2008):

$$d(i) = e^{A(m,\, t = t(i))} + \varphi \tag{2}$$

Equation (2) reveals the connection between the decay rate of each trace and the memory's activation level at its creation. Traces encountered closer in time have higher decay rates due to greater activation at the time of formation (Pavlik & Anderson, 2005; Sense et al., 2016).

At the core of this model is the *Speed of Forgetting* (SoF; φ). *SoF* is crucial in understanding how a memory's past influences its future recall potential. It underscores that a memory's likelihood of being recalled is primarily governed by how quickly it is forgotten, which is influenced by the frequency and context of its recall.

Educationally, this model has demonstrated effectiveness in enhancing student learning (Sense & Rijn, 2022; Sense, Velde, & Rijn, 2021; van Rijn et al., 2009; Wilschut, van der Velde, Sense, Fountas, & van Rijn, 2021). The software tailors the timing and frequency of fact presentation based on the *SoF* values derived from student responses, maximizing retention efficiency. Van Rijn et al. (2009) refined this approach by incorporating both error rates and response times to more accurately determine *SoF.*

Supporting the robustness of this model, Sense et al. (2016) demonstrated that *SoF* is a stable trait unique to individuals. In the realm of clinical application, Hake et al. (2023) have demonstrated the potential of the *SoF* parameter to classify memory impairments with high accuracy. Furthermore, neuroimaging studies by Zhou et al. (2021) and Xu et al. (2021) have linked *SoF* to individual differences in long-term memory function and revealed its correlation with spontaneous brain activity during rest.

Materials and Methods

Participants

Undergraduate students attending the local university (N = 23; 16 female) were recruited on a rolling basis over two quarters and provided with payment or course credit for their participation. Participants were originally included in a larger study that included other conditions and data collection. None of the manipulations, however, affected the data from this experiment in any way and can be ignored.

Memory Task

Four in-person assessments were completed using the online adaptive fact learning system (AFLS) described in Sense et al. (2016) and found online at https://www.memorylab.nl/en/. This system continuously estimates the individualized *Speed of Forgetting* values in real-time as the participant works through the lesson. The AFLS presents new study pairs (e.g., "France / Paris") and schedules repeated tests (e.g., "France / ?") at strategic points based on the online estimates of a user's *Speed of Forgetting*. The system also provides a lesson-level setting to turn off study trials, enabling error learning. Figure 2 provides an example of the software interface.

Each participant performed the task in three different modalities. Four lessons of each modality were administered over four consecutive days. Fact lessons were specifically designed with material that is likely unfamiliar to undergraduate students like Caribbean flags and Swahili words. Lessons were balanced by difficulty such that each modality's lessons contained many facts that could all be learned around the task's 8-minute end time.

Recognition mode. In the recognition mode, participants were presented with study items and then tested on those items. Participants selected the correct response by clicking with a mouse on the button corresponding to the option they thought was correct (Figure 2A).

Recall mode. In the recall mode, study items were presented in the same way but responses were given verbally, by saying out loud the name of the option corresponding to the response or by simply typing in the response. During the response phase, only the cue was presented (Figure 2B).

Error generation mode. The error generation mode was similar to the recall mode except participants typed in their answers and no study item was presented to promote error commission (Figure 2C). To follow a similar protocol as Huelser and Metcalfe (2012) and Leonard et al (2023), 30 weakly associated word pairs were selected for each lesson using Nelson, McEvoy, and Schreiber's (1998) norms.

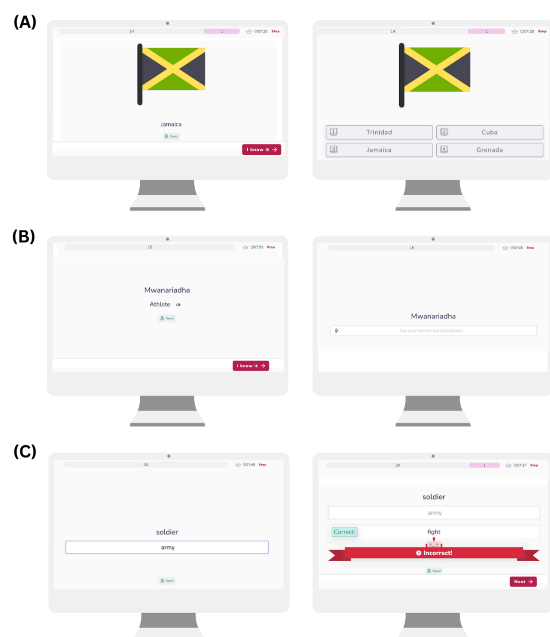

Figure 2: MemoryLab task modalities either with a study opportunity (A and B) or without (C). (A) Recognition mode: multiple choice learning of flag and country pairs. (B) Recall mode: speak or type in response to Swahili word and English translation pairs. (C) Error generation mode: type-in response of word pairs with no study trial.

Measures

The primary measure we are interested in is response time. This will allow us to see if mediation continues to be the mechanism underlying increased error performance. Accuracy will mostly be analyzed in the scope of *Speed of Forgetting* since the modalities contain different material that could differ in fact difficulty. However, incorporating *Speed of Forgetting* and accuracy will allow us to continue to distinguish between error and study learning. Do the benefits of errors remain over multiple trials (more accurate, slower forgetting)? Or does the benefit of multi-trial testing lead to a balance between error and study item performance?

To answer the question of how mechanisms of error learning change over multi-trial testing, we must look at the characteristics of each fact over time, specifically response time, accuracy, and *Speed of Forgetting*. This will allow us to form a comprehensive story for facts learned in different modalities and how their first learning instance can change their retrievability in future trials.

Results

Response Times

To remove extreme values from our data, we used a maximum cutoff point of 15,000ms and a minimum cutoff point of 200ms. Only correct trials were included.

Supporting past findings of mediator retrievals of error learning, on the first fact presentation, participants had longer response times on error items (M = 2536.33 +/- 1519.55ms) than both recall items (M = 2070.07 +/- 1765.97ms) and recognition items (M = 1852.98 +/- 996.42ms), as seen in Figure 3. However, for each subsequent presentation, response times got faster and faster for error items (M = 1868.11 +/- 1455.49ms) such that they were retrieved, on average, faster than both recall (M = 2625.02 +/- 2324.58ms) and recognition items (M = 2101.49 +/- 1403.52ms). Response times for recall and recognition items did not differ significantly over time (see Figure 3 *predicted*).

To analyze response times, mixed linear models were used to account for variability and individual differences. Specifically, we fitted a mixed model to all of the experimental trials, including a fixed effect as the interaction between trial condition (Recognition vs. Recall vs. Error) and fact presentation (repetition) two random effects: a participant-level intercept and a participant-level slope to account for individual differences in response latencies and slopes between fact presentations, respectively. This complex interaction model fit the data better than both a simple model with fixed effects of condition and fact presentation and a participant-level intercept. The interaction model also fit the data better than the simple model with a random intercept and slope for each participant in each condition. The model only captured the mediator effect between error items and recognition items. That is, at presentation 0, error items were estimated to have a longer response time than recognition items (β = -193.62, SE = 37.32, t = -5.19, $p < 0.0001$) but shorter response times than recall items (β = 314.84, SE = 39.57, t = 7.96, $p < 0.0001$), likely due to the large increase in response times for recall items on presentation 1 (See Figure 3 *observed* vs *predicted*). Additionally, the model uncovered a large interaction effect between fact presentation and condition. On average, error item response times decreased over the number of presentations (β = -157.49, SE = 21.32, t = -7.39, $p < 0.0001$) compared to the response times of recall items (β = 156.71, SE = 14.25, t = 11.00, $p < 0.0001$) and recognition items (β = 161.56, SE = 13.21, t = 12.23, $p < 0.0001$). The complete results of the model are shown in Table 1.

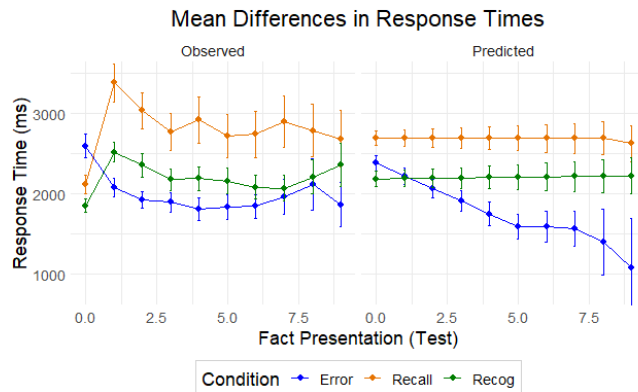

Figure 3: Differences in response times for different task modalities over the first ten fact presentations. Colored lines, dots, and error bars represent means +/- SE for the Error (blue), Recall (orange), and Recognition (green) conditions. Left: Empirical data; Right: Fitted model.

Table 1: Results of the Mixed-Level Model for Response Times

Predictor	β estimate	*SE*	*t*	*p*
(Intercept)	2381.636***	96.156	24.769	0.000
Condition (Recall)	314.842***	39.574	7.956	0.000
Condition (Recog)	-193.621***	37.316	-5.189	2e-07
Fact Presentation	-157.486**	21.324	-7.385	0.000
Recall x Presentation	156.709	14.247	11.000	0.000
Recog x Presentation	161.560	13.205	12.235	0.000

$^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$

Speed of Forgetting

On average, participants had lower *Speed of Forgetting* (SoF) on error items (M = 0.24 +/- 0.04) than both recognition items (M = 0.30 +/- 0.0) and recall items (M = 0.30 +/- 0.07), as seen in Figure 4. To analyze *SoF* values, mixed linear models were used to account for variability and individual differences. Specifically, we fit a mixed model to all of the experimental trials, including the particular trial condition (Recognition vs. Recall vs. Error) as a fixed effect and a participant-level intercept as a random effect to account for individual differences in speeds of forgetting. We also included a participant-level slope as a random effect to account for individual differences in slopes between lesson modalities. This more complex model fit the data better than both a simple model (fixed effect only) and a fixed effect model with just a participant-level random effect. The model uncovered a large main effect of condition such that the error lessons had significantly lower *SoF*s compared to both the recall ($\beta = -0.06$, $SE = 0.01$, $t = 6.10$, $p < 0.0001$) and recognition lessons ($\beta = -0.06$, $SE = 0.007$, $t = 8.53$, $p < 0.0001$). The complete results of the model are shown in Table 2. These results demonstrate the existence of error learning with an alternate measure rather than accuracy, one that is more reflective of the dynamics of memory over time. Furthermore, these results show that error trials are not only accessed more easily, but they do seem to be intrinsically more resistant to forgetting than other memories.

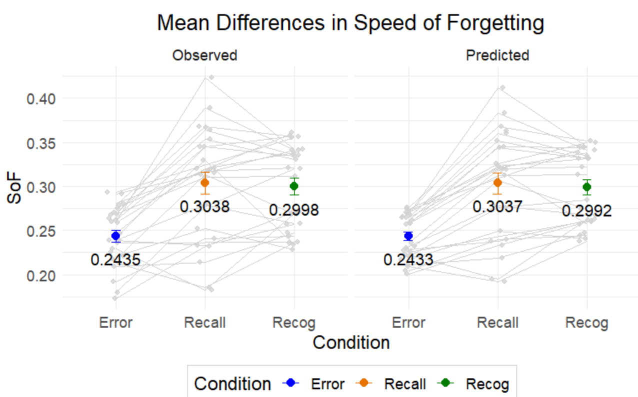

Figure 4: Differences in *SoF* in the three task modalities. Gray dots and lines represent data for individual participants; colored dots and error bars represent means +/- SE for the Error (blue), Recall (orange), and Recognition (green) conditions. *Left*: Empirical data. *Right*; Fitted linear model.

Table 2: Results of the Mixed-Level Model for *Speed of Forgetting*

Predictor	**β estimate**	***SE***	***t***	***p***
(Intercept)	0.25***	0.006	40.04	0e+00
Condition (Recall)	0.06***	0.010	6.0790	2.4e-06
Condition (Recog)	0.06***	0.0070	8.5330	0e+00

**p < 0.05 **p < 0.01 *** p < 0.001*

Errors or Levels of Processing?

The previous results have shown that mediator retrievals play a significant role only in the very first trial after the error is made and corrective feedback is given and that subsequent trials show, on average, faster response times in the error-generation condition than in the other modalities. There is no difference between the trials except how many times the fact has been presented beforehand and how long it has been since the first error was committed.

This prompts the question of what causes items in the error-generating condition to be more memorable and more resistant to decay. To address this question, we looked at the error-generating condition more closely. Although the condition was designed to elicit errors in most trials, about 95% of our participants correctly guessed the answer on at least one item. Thus, we directly compared correctly answered vs incorrectly answered items in the error-generating condition. If the superiority of error trials is exists because they were incorrectly responded to (e.g., because of the additional encoding of error feedback), we would expect to find that only items that were initially generated incorrectly would exhibit better recall.

Surprisingly, correctly and incorrectly answered items in the error-generating condition showed substantially similar patterns in terms of both *SoF* and RTs. This was further confirmed when fitting random-slope mixed linear models to the data: no significant difference was found between these two conditions in terms of RTs, while correctly answered items showed an even smaller *SoF* than incorrectly answered ones ($\beta = -0.01$, $p < .001$), as seen in Figure 5 and Figure 6, respectively.

These findings show that it is *not* the generation of an error *per se* that leads to better accuracy and recall. Rather, it is the additional effort of generating a response when no answer is readily available that leads to sustained improvements—an account that is compatible with the idea that error-generating trials are associated with, and benefit from, greater depth of processing (Craik & Tulving, 1975).

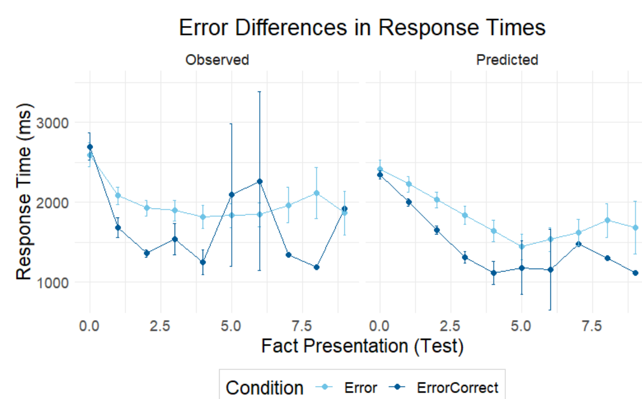

Figure 5: Differences in response times between correct guesses and error items in the error generation condition only. Colored dots and error bars represent means +/- SE for true errors (light blue) and correct errors (dark blue). Left: Empirical data. Right; Fitted linear model.

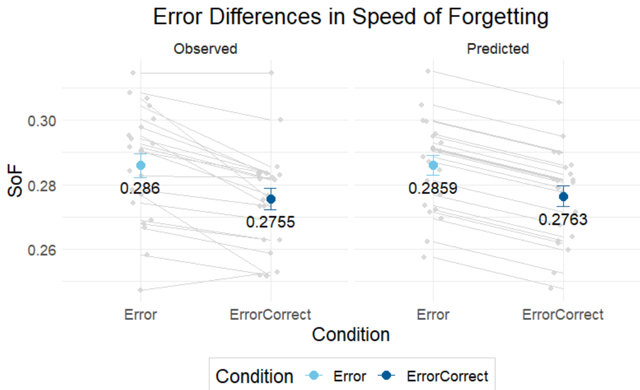

Figure 6: Differences in *SoF* between correct guesses and error items in the error generation condition only. Gray dots and lines represent data for individual participants; colored dots and error bars represent means +/- SE for true errors (light blue) and correct errors (dark blue). *Left*: Empirical data. *Right*; Fitted linear model.

Discussion

In this paper, we investigated the nature of the error learning time course with a multi-trial testing paradigm and built-in fact learning optimization. This enhanced past findings in two primary ways. First, despite findings of strong evidence for mediator retrievals of errors during testing, a single test trial paradigm restricts investigating if there's prolonged use of mediator retrievals past the first test trial or even a continued enhanced performance on error items compared to study items. Secondly, this single-trial learning environment does not benefit all learner types equally. Relying solely on accuracy as a measure of learning effectiveness may inadvertently amplify disparities among different learner types. For these reasons, we reimplemented the error-learning task in an adaptive fact-learning system. Not only does this paradigm offer many test opportunities after learning, but it also uses the testing and spacing effect to optimize learning for all individuals. This allowed us to investigate the multi-trial effects of error learning and ensure that all participants were at maximal performance. The software also provided us with a highly stable and individually specific measure on each trial, *Speed of Forgetting*, which reflects how quickly the fact memory is estimated to decay.

Overall, we found a significantly lower *Speed of Forgetting* for error items compared to both recall and recognition items. This further supports the existence of error learning, going beyond the typical accuracy measure to establish error benefits with a more stable measure and in a more individualized task.

Importantly, we were able to answer our question about the relationship between mediator retrieval and repeated fact testing. On the first fact presentation, error items had longer response times on average compared to both recall and recognition items. This finding is consistent with past work that found increased response times on error items due to the additional cognitive processing occurring with mediator retrievals of error memories (Kornell et al., 2009; Huelser & Metcalfe, 2012; Leonard et al., 2023). However, for subsequent fact presentations, error response times got faster and faster, while recall and recognition response times stayed relatively stable over time. This demonstrates that after a single mediator retrieval, the association between cue and error items becomes so strong that it is compatible with an elaborative hypothesis where error response times are faster than study items.

The most notable limitation of this study is the use of different fact materials for lesson types. As mentioned, this data results from a larger study that aimed to assess multiple fact materials and presentations. Thus, our comparison between conditions (Recognition vs. Recall vs. Error) is not solely clear-cut and designed for error analysis. However, with these promising results and the combined success and availability of the task software, one could easily redesign the lessons to make them equal in material and differ in specific facts. Although previous studies have repeatedly found no differences in *SoF* values between different types of materials (Sense et al., 2016; Zhou et al., 2021; Hake et al., 2023), these cannot be ruled out entirely in this experiment. Furthermore, there are likely material-specific differences in response times that we cannot fully accommodate with the use of mixed linear models, so such a design would likely ensure clearer comparisons. Additional work could use this condition-based paradigm alongside the MemoryLab software to test the error learning effect across different materials like trivia, math, language learning, etc. to further understand the interaction between material, learning condition, and *Speed of Forgetting*. Our final analysis involving levels of processing is also restricted by the task design where correct guesses are limited to occur ~5% of the time. Controlling correct guesses and errors in a modified paradigm would give more power to this error-specific analysis to further investigate how feedback influences response times and *Speed of Forgetting* after the first retrieval.

Nevertheless, this work marks an important step in using modeling approaches to capture learning phenomena. Applying models that capture the dynamic of memory in favor of simpler paradigms can help apply experimental findings to real-world situations. In this case, we were able to look at the history of different memory types and establish how they differ in resistance to forgetting. We were also able to gain insight into how mechanisms of error memory retrieval change over time, that with repeated presentations associations between mediator memories increase such that they are retrieved faster and faster. Further investigation of error learning with this paradigm can lead to advancements in individual-based learning methods, where each learner receives optimized tasks that fit their specific learning style.

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological science, 2(6), 396-408.

Anderson, J. R. (2007). How can the human mind exist in the physical universe? New York, NY: Oxford University Press.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological science*, *19*(11), 1095-1102.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of learning and motivation*, *53*, 63-147.

Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, *9*.

Hake, H. S., Leonard, B., Ulibarri, S., Grabowski, T., Van Rijn, H., & Stocco, A. (2023). Breaking New Ground in Computational Psychiatry: Model-Based Characterization of Forgetting in Healthy Aging and Mild Cognitive Impairment. *medRxiv*, 2023-05.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*(4), 514–527.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In *Learning and Memory: A Comprehensive Reference* (pp. 487–514). Elsevier.

Kim, A. S. N., Wong-Kee-You, A. M. B., Wiseheart, M., & Rosenbaum, R. S. (2019). The spacing effect stands up to big data. *Behavior Research Methods*, *51*, 1485-1497.

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.

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 106–114.

Kornell, N., & Vaughn, K. E. (2016). Chapter five - How retrieval attempts affect learning: A review and synthesis. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 65, pp. 183–215). Academic Press.

Leonard, B., & Stocco, A. (2023). Faulty Memories, Favored Outcomes: How Errors Impact Learning Processes. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Mera, Y., Rodríguez, G., & Marin-Garcia, E. (2022). Unraveling the benefits of experiencing errors during learning: Definition, modulating factors, and explanatory theories. *Psychonomic bulletin & review*, *29*(3), 753–765.

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus*, *10*(4), 352-368.

Nelson, D., McEvoy, C., & Schreiber, T. (1998). *University of South Florida Free Association Norms*.

Rosenbloom, P. S., & Newell, A. (1982). Learning by chunking: A production-system model of practice.

Pavlik Jr, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation based model of the spacing effect. *Cognitive science*, *29*(4), 559-586.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science, 8*(into, 305-321.

Sense, F., & van Rijn, H. (2022, January 27). Optimizing Fact-Learning with a Response-Latency-Based Adaptive System.

Sense, F., van der Velde, M., & van Rijn, H. (2021). Predicting university students' exam performance using a model-based adaptive fact-learning system. *Journal of Learning Analytics*, *8*(3), 155-169.

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. Trends in cognitive sciences, 14(10), 464-472.

van Rijn, D., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving Learning Gains by Balancing Spacing and Testing Effects. *Proceedings of the 9th International Conference on Cognitive Modeling*, 108–114.

Wilschut, T., van der Velde, M., Sense, F., Fountas, Z., & van Rijn, H. (2021). Translating a Typing-Based Adaptive Learning Model to Speech-Based L2 Vocabulary Learning. *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 245–250. Utrecht Netherlands: ACM.

Xu, Y., Prat, C., Sense, F., Rijn, H. van, & Stocco, A. (2021, August 5). *Distributed Patterns of Functional Connectivity Underlie Individual Differences in Long-Term Memory Forgetting* (p. 2021.08.04.455133). p. 2021.08.04.455133. bioRxiv.

Zhou, P., Sense, F., van Rijn, H., & Stocco, A. (2021). Reflections of idiographic long-term memory characteristics in resting-state neuroimaging data. *Cognition*, *212*, 104660.