Title
What Drives Learning During Practice and Testing? Evidence for Distinct Roles of Exposure and Retrieval

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Cao, Meng
Gold, Gillian
Carvalho, Paulo F.

Publication Date
2026

What Drives Learning During Practice and Testing? Evidence for Distinct Roles of Exposure and Retrieval

Meng Cao (mengc2@andrew.cmu.edu)[1], Gillian Gold[1], & Paulo F. Carvalho[1]
[1]Human-Computer Interaction Institute, Carnegie Mellon University

Abstract

Practice improves learning, yet it remains unclear which learning mechanism drives memory change across practice and testing. Competing theories attribute learning to repeated exposure, error correction, or successful retrieval, but these mechanisms are rarely evaluated in direct competition with one another. We propose a modeling approach that enables learning mechanisms to be evaluated during practice and during tests within a dynamic representation of memory change over time. Learning mechanisms are formalized as competing update rules and evaluated using nested models fit to spaced learning data with delayed tests. Results show that learning during practice is best explained by a combination of exposure-based learning and effortful successful retrieval. During posttests without feedback, learning, when present, is driven primarily by successful retrieval, with no additional benefit of retrieval effort. These findings suggest that practice supports learning through multiple mechanisms, but that the mechanisms depend on the constraints of the learning context.

Keywords: Retrieval practice; learning mechanisms; memory modeling; error correction; exposure; forgetting

Introduction

Repeated practice is a cornerstone of learning and memory research. Across decades of empirical work, practice has been shown to improve fluency and retention across a wide range of materials, learners, and educational contexts, from laboratory memory tasks to classroom instruction (e.g., Cepeda et al., 2006; Roediger & Butler, 2011; Agarwal et al., 2021). However, practice is not a unitary cognitive operation. Each encounter can involve exposure to information, attempts at retrieval, feedback, and forgetting as time passes. Not all components of practice are assumed to contribute equally to learning, and their influence may depend on task structure and timing.

How do these different components of practice shape changes in memory over time across different phases? During practice, learners typically receive repeated exposure, engage in retrieval attempts, and often receive feedback. During tests without feedback, by contrast, learning may arise from retrieval alone (Bjork, 1975; Roediger & Karpicke, 2006; Pyc & Rawson, 2009; Kornell et al., 2011). Because different mechanisms are likely to operate across these phases, we distinguish learning during practice from learning during tests and review the evidence relevant to each.

Mechanisms of Learning During Practice With Feedback

During practice, multiple mechanisms are theoretically available to drive changes in memory, described below.

Exposure-Based Learning. One class of accounts proposes that learning is driven by exposure, such that repeatedly encountering information increases familiarity and performance, although learning gains diminish as the number of repetitions increases (e.g., Heathcote et al., 2000; Karpicke & Roediger III., 2007; Thorndike, 1913). From this perspective, repeated encounters benefit memory by reactivating stored information and creating opportunities for incremental strengthening over time (Nadel & Moscovitch, 1997).

In computational models, exposure-based accounts are commonly implemented as encounter-based updating rules in which each presentation strengthens memory. For example, the ACT-R (Adaptive Control of Thought-Rational) declarative memory model formalizes how practice history and forgetting jointly determine performance through increments to memory strength that decay over time (Anderson & Lebiere, 1998; Pavlik & Anderson, 2005). The Relearning Attenuates Decay (RAD) model extends this framework by proposing that relearning attenuates subsequent forgetting as a function of an item's activation at practice (Rawson et al., 2018). Similarly, MINERVA II is an exemplar-based model in which each encounter encodes a new episodic trace, with performance reflecting the aggregated influence of stored traces (Hintzman, 1984).

Successful Retrieval. A second class of accounts emphasizes effortful yet successful retrieval as the primary mechanism of learning during practice (Appleton-Knapp et al. 2005; Thios & D'Agnostino, 1978; see also reminding theory, Benjamin & Tullis, 2010). According to these accounts, actively retrieving information leads to changes in memory traces and slows forgetting (Bjork, 1975; Rachatasumrit et al., 2023; Yan et al., 2024). Crucially, these benefits depend on retrieval being both effortful and successful (Pyc & Rawson, 2009). When retrieval is too easy, additional retrieval attempts provide little benefit (Minear et al., 2018; Pyc & Rawson, 2009). Conversely,

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

when learners fail to retrieve the correct response, and no corrective feedback is provided, continued retrieval attempts tend to yield little benefit and may even impair learning (Fiechter & Benjamin, 2018; Rowland, 2014).

In computational models, retrieval-based learning is implemented through event-type models that assign larger learning updates to successful retrieval events than to study events or failed retrieval attempts. For example, Pavlik (2008) extended the ACT-R framework by explicitly distinguishing learning from study versus retrieval, with successful retrieval producing larger increments to memory strength than study events. Similarly, the Performance Factors Analysis (PFA) model decomposes prior practice history into counts of successes and failures (Pavlik et al., 2009). Successful retrievals usually contribute positively to learning, whereas failed retrievals contribute little or may even impede subsequent performance. Extensions of the PFA model further incorporate item difficulty to capture how the difficulty of prior successful and unsuccessful retrievals modulates learning (Cao et al., 2019).

Error Correction. A third class of accounts emphasizes error correction, proposing that learning is driven by feedback following incorrect responses, which corrects faulty or incomplete representations. Errors are thought to play a functional role by revealing gaps in understanding that would otherwise remain unaddressed. This mechanism is central to theories such as impasse-driven learning, which posits that encountering failure or impasse triggers additional cognitive processing (VanLehn et al., 1999), and productive failure, which suggests that generating incorrect solutions prior to instruction primes learners to integrate subsequent explanations or feedback (Kapur, 2008; Kapur, 2015). Consistent with these accounts, empirical work shows that error generation followed by feedback enhances conceptual understanding and transfer, particularly when learners actively engage with their mistakes (Darabi et al., 2018; Tawfik et al., 2015; Jackson et al., 2022).

Computationally, the Delta-rule or Rescorla-Wagner-style learning models propose that the magnitude of learning is proportional to the difference between predicted and actual outcomes (Bush & Mosteller, 1955; Rescorla & Wagner, 1972). Correct responses produce little or no updating when expectations are already high, whereas errors generate larger updates that revise internal representations toward the correct response. Learning in these models, therefore, depends on failure signals and corrective information, rather than on exposure or retrieval success per se.

Mechanisms of Learning During Tests Without Feedback

In contrast to practice with feedback, tests typically involve retrieval in the absence of feedback, limiting the learning mechanisms that can operate. From exposure-based perspectives, little or no learning is expected during tests without feedback, because such tests do not provide the repeated exposure assumed to drive learning during restudy or feedback-based practice, nor do they support error correction following incorrect responses. Retrieval-based accounts, by contrast, predict that learning can occur during tests contingent on successful retrieval. Consistent with this view, empirical studies have shown that retrieval practice can improve later retention even when feedback is withheld (Karpicke & Roediger III, 2008; Roediger III & Butler, 2011). One influential explanation is the distribution-based bifurcation model (Kornell et al., 2011), which proposes that testing without feedback produces a divergence in memory strength, with successfully retrieved items strengthening while unretrieved items remain weak.

Comparing Mechanisms to Understand Learning Dynamics

Different theoretical accounts propose distinct learning mechanisms, raising the question of which operate under the constraints of practice and tests and best explain changes in memory over time. To address this question, we adopt a modeling approach that evaluates competing learning mechanisms using a dynamic model of memory change fit to behavioral data.

During practice, we evaluated a hierarchy of models reflecting alternative learning mechanisms. If learning is driven primarily by exposure, a model with exposure-based updating should outperform a no-learning model. If successful retrieval contributes beyond exposure, a model combining exposure-based and retrieval-based updating should provide a better fit. If error correction plays a substantive role, a model incorporating exposure-based and error-correction learning should outperform an exposure-only model. Finally, if exposure, retrieval, and error correction jointly contribute to learning, a full model including all three mechanisms should best account for the data.

During tests without feedback, learning opportunities are more limited. If learning occurs during tests, models in which successful retrieval produces learning should outperform a no-learning model. If retrieval effort further modulates learning, models allowing learning to scale with retrieval difficulty should provide a better fit.

A Dynamic Model of Learning and Forgetting

The model represents memory as a latent state that evolves continuously over time. Each item is associated with a memory strength that determines the probability of successful retrieval, decays as a function of elapsed time between encounters, and may be updated when the item is encountered. Learning updates depend on different learning mechanisms (see Figure 1). During practice, retrieval attempts are followed by feedback, allowing learning to arise from exposure, effortful yet successful retrieval, or error correction following failure. During tests, retrieval occurs without feedback, such that retrieval itself is the only potential mechanism for learning.

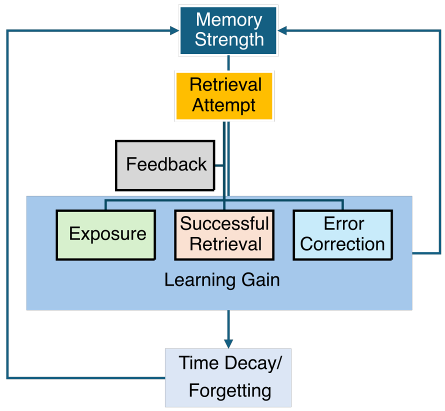

Figure 1: The Memory Dynamics During Practice and Retention

Model Specification

We next specify the learning and forgetting mechanisms that instantiate the model, defining how memory strength changes as a function of exposure, successful retrieval, error correction, and elapsed time.

Latent Memory State. For each learner *j* and item *i*, we assume there exists a latent memory strength $S_{ij}(t)$ that evolves continuously over time. This quantity represents the accessibility of the correct response in memory at time *t*. It is not directly observed but is inferred from the learner's responses across trials. At any moment, the current memory strength reflects the cumulative effects of prior learning events and ongoing forgetting. Learning increases $S_{ij}(t)$, whereas forgetting reduces it over time.

Mapping Memory to Performance. Observed responses are generated from the latent memory state through a probabilistic performance model. Specifically, we assume that the probability of a correct response on a trial is a monotonic function of memory strength:

$$p_{ij}(t) = \sigma(S_{ij}(t)) = \frac{1}{1+e^{-S_{ij}(t)}} \quad (1)$$

This logistic mapping produces diminishing returns of memory strength on accuracy, with strong memories approaching ceiling performance and weak memories remaining near chance, consistent with prior memory models (e.g., Anderson & Lebiere, 1998).

Learning Updates. When a learner encounters an item, memory changes in response to both retrieval and feedback. S_{ij}^- denotes the memory strength immediately before a trial, and S_{ij}^+ the strength immediately after the learning update. The transition from S_{ij}^- to S_{ij}^+ captures the overall change in memory strength induced by that trial:

$$S_{ij}^{+} = S_{ij}^{-} + \Delta S_{gain} \quad (2)$$

The term S_{gain} represents the amount of learning attributable to the trial and is defined differently across model variants to reflect alternative theoretical assumptions about what drives learning.

Exposure-Based Learning. If learning occurs whenever the learner sees the correct answer, regardless of retrieval success, there will be a constant increment in memory on every trial: $\Delta S_{exposure} = \alpha_0$.

Error-Driven Learning. If learning is driven by errors followed by corrective feedback, the memory strength updates occur when responses are incorrect: $\Delta S_{error} = \alpha_E(1 - y)$, $y \in \{0,1\}$ indicates the observed correctness of the response. When learners have a correct answer, ΔS_{error} equals 0.

Effortful and Successful Retrieval-Driven Learning. If learning is driven by successful retrieval, especially when retrieval is effortful, we formalize it as: $\Delta S_{retrieval} = \alpha_R y(1 - p)$.

In the formula, learning from retrieval occurs only when the learner successfully retrieves the correct information. If retrieval fails ($y = 0$), $\Delta S_{retrieval}$ equals 0. The factor $(1 - p)$ indexes retrieval effort, with p denoting the model's predicted probability of a correct response given the current memory strength. When retrieval is unlikely (p is small), a correct response indicates a more effortful retrieval and produces a larger learning gain. When retrieval is easy ($p \approx 1$), a correct response produces little learning.

If all three components drive learning during practice, they can be combined to produce a general learning rule: $S_{ij}^{+} = S_{ij}^{-} + \alpha_0 + \alpha_E(1 - y) + \alpha_R y(1 - p)$.

However, during tests without feedback, exposure and error correction are not plausible sources of learning, leaving retrieval as the only potential mechanism. If learning depends on effortful yet successful retrieval, $\Delta S_{retrieval}$ follows the formulation described above.

Successful Retrieval-Driven Learning (Retrieval Effort not Considered). If learning is driven solely by successful retrieval, independent of retrieval effort (e.g., Kornell et al., 2011): $\Delta S_{retrieval} = \alpha_R y$. Therefore, learning updates during tests compare alternative retrieval-based learning mechanisms.

Forgetting Across Time. Between learning events, memory decays as a function of elapsed time. Let Δt be the interval between two encounters with the same item. The memory strength just before the next encounter is determined by applying a forgetting function to the previous post-update strength $S_{ij}^{+}(t)$: $S_{ij}^{-}(t + \Delta t) = f(S_{ij}^{+}(t), \Delta t)$.

This function captures the loss of accessibility that occurs as time passes. Different theoretical accounts of forgetting correspond to different forms of the decay function *f*, including fixed versus item-specific decay rates, and alternative functional forms such as exponential, power-law, or logistic decay (e.g., Rubin & Wenzel, 1996; Wickelgren, 1974; Wixted & Ebbesen, 1997). Because the present work focuses on what drives learning during practice rather than on adjudicating theories of forgetting, we adopt a single, well-established exponential decay function and hold the forgetting process constant across models.

$$S^{-}_{ij}(t + \Delta t) = S^{+}_{ij}(t)e^{-\lambda\Delta t} \quad (3)$$

The parameter λ governs the rate of forgetting. This formulation captures the core intuition that memories weaken continuously over time after the retrieval or feedback. This $S^{-}_{ij}(t + \Delta t)$ determines the retrieval success on the next encounter of the same item.

Together, these components specify a state-space formulation of learning and forgetting in which performance is generated from a latent memory state, learning events update that state as a function of retrieval outcomes and feedback, and forgetting degrades it over time. In the next section, we fit and compare model variants to determine which learning mechanisms best account for observed patterns of learning and retention.

Applications

The goal of applying the model was to determine which learning mechanisms best account for learning during both practice and tests. We applied the model to data from a laboratory learning experiment that included a practice phase followed by three delayed posttests (Experiment 3, Cao et al., under review), enabling learning mechanisms during practice and tests to be evaluated within a single dataset. In this experiment, spacing and repetition were traded off under a fixed amount of practice time, creating systematic variation in retrieval difficulty, exposure, and feedback while holding overall practice time constant.

All model variants were fit to the same trial-level data using maximum likelihood estimation. For each learner, item, and trial, the model predicts the probability of a correct response from latent memory strength immediately before the trial, with responses modeled as Bernoulli outcomes. Model parameters were estimated by maximizing the log-likelihood as memory evolved according to each model's learning and forgetting equations. Competing models were compared using AIC (Akaike Information Criterion; Akaike, 1998) and BIC (Bayesian Information Criterion; Schwarz, 1978), with likelihood-ratio tests used for nested comparisons. Predictive accuracy was additionally assessed on held-out data using RMSE.

Dataset Description

Participants were randomly assigned to one of four spaced practice conditions that varied in the number of repetitions and spacing intervals (see Figure 2). In each condition, participants learned 12 Indonesian-English word pairs under a fixed practice schedule. For example, in the 2-repetition, 36-spacing-interval condition, all target words were practiced twice, with 24 filler items presented between repetitions to manipulate the spacing interval. The order of target words was randomized across participants, but the repetition order was held constant within each participant. The final data included 235 participants. All data and R code can be accessed at (https://osf.io/dgsr8/).

Condition	72 training trials							
2 Repetitions, 36 Spacing	12 target	24	12 target	24				
3 Repetitions, 24 Spacing	12 target	12	12 target	12	12 target	12		
4 Repetitions, 18 Spacing	12 target	6	12 target	6	12 target	6	12 target	6
6 Repetitions, 12 Spacing	12 target	12 target	12 target	12 target	12 target	12 target		

Figure 2: Number of Target and Filler Items in the Learning Session Across Conditions. *Note: the gray blocks represent the filler items*

Feedback was provided during the 72-trial learning session, with each feedback display lasting 2 seconds. After the learning session, participants completed three final tests: an immediate test, a 1-day delayed test, and a 3-day delayed test, testing their memory retention of the word pairs.

Model Comparison

We adopted a two-stage modeling approach to characterize learning mechanisms across the practice and posttest phases. In the first stage, we identified the learning mechanisms that best accounted for behavior during the practice phase, assuming that no learning occurred during posttests. In the second stage, we retained the practice learning mechanism identified in the first stage and evaluated whether allowing learning during posttests improved model fit.

Model Comparisons for the Practice Session. We compared a set of 5 nested models to identify the learning mechanisms that best accounted for behavior during the practice session. The no-learning model (M-1) served as a null model in which memory only decayed over time. The exposure-only model (M0) served as the baseline learning model, where learning updates during practice were driven solely by exposure, forgetting was modeled as an exponential decay process, and there were no learning updates during posttests. The exposure + error-correction model (M1) allowed additional learning following incorrect responses with feedback. The exposure + retrieval model (M2) allowed additional learning following successful retrieval, scaled by retrieval difficulty. Finally, the full model (M3) includes all three learning mechanisms.

Table 1 summarizes the fit of the competing models. The exposure-only model (M0) provided a better fit than the no-learning model (M-1), indicating that exposure contributed to learning during practice. Allowing learning to increase following errors with feedback (M1) substantially improved model fit, suggesting an important role for error correction beyond exposure alone. However, allowing learning to increase following successful retrieval, scaled by retrieval difficulty (M2), yielded the best overall performance across all fit metrics (lowest AIC, BIC, and RMSE). The full model (M3) did not further improve fit over M2 and was penalized by information criteria for its additional parameter.

Table 1: Model Fit Statistics for the Practice Learning Mechanisms

Model	k	Log-likelihood	AIC	BIC	RMSE
M-1	2	-10,158.79	20321.59	20337.00	0.47
M0	3	-9,307.66	18,621.32	18,644.43	0.44
M1	4	-7,999.85	16,007.71	16,038.53	0.40
M2	**4**	**-7,759.99**	**15,527.97**	**15,558.80**	**0.39**
M3	5	-7,759.90	15,529.81	15,568.34	0.39

We also compared nested models using likelihood-ratio chi-square tests. The results were consistent with the AIC and BIC comparisons, indicating that the model including exposure and effortful yet successful retrieval provided the best fit to the data (see Table 2). Model predictions closely matched observed performance during learning and posttests (Figure 3). Together, these results provide strong evidence that learning in this practice is primarily driven by exposure and effortful yet successful retrieval.

Table 2: Likelihood-ratio Chi-square Tests for Nested Model Comparisons

Comparison	$\chi^2(1)$	p
M-1 vs. M0 (exposure)	1702.27	< .001
M0 vs. M1 (exposure + error)	2615.61	< .001
M0 vs. M2 (exposure + retrieval)	3095.34	< .001
M1 vs. M3 (full)	479.90	< .001
M2 vs. M3 (full)	0.17	.684

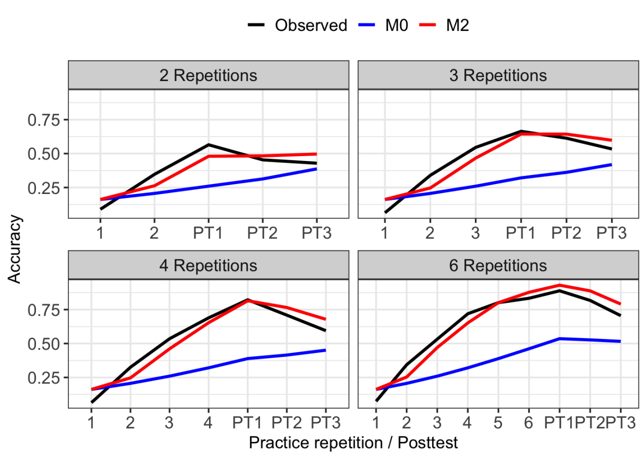

Figure 3: Observed and Model-Predicted Accuracy During the Learning Session and Posttests. *Note: M0 is the exposure-only model, and M2 is the exposure + retrieval model. PT1 represents posttest 1, PT2 represents posttest 2, and PT3 represents posttest 3.*

Model Comparisons for the Posttests. To examine whether learning occurs during tests without feedback, we used the best-fitting practice model (M2) as a baseline, which assumes learning during practice but none during tests. We compared this model to two extensions that differed only in their assumptions about test-phase learning. Model M2a allowed learning during tests contingent on successful retrieval, whereas Model M2b further allowed learning to scale with retrieval effort, paralleling the practice-phase assumption. Model comparisons based on AIC, BIC, and RMSE indicated that the success-only model provided the best fit, with no additional explanatory benefit of retrieval effort during tests (Table 3).

Table 3: Model Fit for Competing Posttest Learning Mechanisms

Model	Log-likelihood	AIC	BIC	RMSE
M2	-7,759.99	15,527.97	15,558.80	.391
M2a	**-7,566.25**	**15,142.49**	**15,181.02**	**.385**
M2b	-7,612.69	15,235.37	15,273.90	.386

Implications

The present work addressed a central question in the study of practice effects: what mechanisms drive learning during practice and during testing. We introduced a modeling approach that represents memory as a continuously evolving latent state and allows competing learning mechanisms to be evaluated across practice and posttest phases within the same behavioral data. Model comparisons revealed that the dominant learning mechanisms differed across phases. During practice, when feedback was available, learning was best explained by a combination of exposure-based learning and effortful yet successful retrieval. During posttests without feedback, learning following successful retrieval improved model fit, whereas incorporating retrieval effort provided no additional explanatory benefit.

Learning Mechanisms at Practice Session

During practice, exposure contributed to learning, consistent with prior research demonstrating that repetition is a fundamental component of skill and knowledge acquisition (Carpenter et al., 2022). As the number of repetitions increased, learning gains rose rapidly at first and then slowed over time, reflecting diminishing returns with continued practice (e.g., Thorndike, 1913; Heathcote et al., 2000). However, exposure-based learning alone was insufficient to explain the observed learning trajectories. Models in which learning was driven solely by exposure produced smaller learning gains than models that additionally incorporated effortful yet successful retrieval (see Figure 3). This pattern is consistent with the testing effect, which shows that actively retrieving information produces greater learning than passive exposure or rereading

(Roediger & Butler, 2011; see reviews by Adesope et al., 2017; Karpicke & Aue, 2015; Rowland, 2014).

Notably, models that emphasized effortful yet successful retrieval provided a better account of learning than models based on error-driven learning alone. This result aligns with theoretical perspectives that emphasize the role of retrieval difficulty in promoting learning, including retrieval-effort accounts (Appleton-Knapp et al., 2005; Thios & D'Agostino, 1978; see also reminding theory, Benjamin & Tullis, 2010) and desirable-difficulty frameworks (Bjork, 1994; Bjork & Bjork, 1992, 2011).

Importantly, the superior fit of the effortful-retrieval model does not imply that errors fail to contribute to learning. Effortful retrieval can arise from time-induced forgetting due to longer intervals between practice opportunities, as well as prior retrieval failures that leave memory strength relatively weak. These retrieval failures are associated with lower estimated memory strength, which increases the effort required for subsequent successful retrieval attempts. When learners later retrieve the correct response despite this reduced strength, retrieval is both successful and effortful, producing a larger learning update. In this way, learning that can be traced back to prior errors is captured through retrieval-based learning in the model, rather than appearing as a separate error-correction process.

Learning Mechanisms at Posttests

Posttests differ from the learning session in that they typically provide no feedback. As a result, any learning that occurs during this phase must be attributed to retrieval itself rather than to exposure or error correction. Model comparisons indicated that incorporating retrieval effort did not improve model fit beyond retrieval success, suggesting that learning during tests is driven primarily by retrieval success rather than by the degree of retrieval difficulty. One interpretation of this pattern is that successful retrieval during posttests reflects a process of memory stabilization or consolidation, rather than incremental strengthening through effortful search. By the time of the posttests, memory representations have already undergone substantial learning during the practice phase. Successful retrieval at this stage may therefore serve to maintain or reinforce existing representations, even in the absence of external feedback (Kornell et al., 2011). In contrast to the learning session, where retrieval difficulty varies widely and strongly modulates learning gains, posttest retrieval may play a qualitatively different role by consolidating memories that remain accessible after delays. Importantly, this result does not imply that retrieval effort is irrelevant during posttests. Rather, it suggests that, given the range of retrieval difficulty present at posttest, a simpler retrieval-based learning rule provides a sufficient account of the data.

Taken together, these findings refine theoretical accounts of the testing effect by showing that retrieval does not exert a uniform influence on learning across practice phases. Retrieval contributed to learning in a state-dependent manner, such that effortful yet successful retrieval drove learning during practice when feedback was available, whereas successful retrieval alone accounted for learning during posttests without feedback. This dissociation suggests that the testing effect reflects multiple retrieval-based mechanisms operating at different stages of learning, rather than a single process. By formalizing these mechanisms within a common latent-state model, the present work clarifies when and how retrieval improves learning.

Future Work

In the present model, spacing influences learning by increasing the interval over which forgetting occurs, thereby raising retrieval difficulty without eliminating retrieval success. Spacing does not directly alter the rate of forgetting in the model. However, prior work suggests that spaced retrieval may slow forgetting and produce more durable memory representations (Wollstein & Jabbour, 2022). Incorporating mechanisms through which spacing affects forgetting rates, such as retrieval-dependent stabilization or consolidation, is an important direction for future research.

The model also assumes a fixed exponential function to describe forgetting over time. Although this assumption is well established and parsimonious, forgetting dynamics may vary across tasks, time scales, or stages of learning (e.g., Ebbinghaus, 1885). Future work could compare exponential decay with alternative formulations, including power-law or stage-dependent forgetting functions, to better characterize how memory changes across learning and retention.

Conclusion

The present work identifies the mechanisms that drive learning during practice and testing by showing how retrieval outcomes interact with memory strength over time. Learning during practice is best explained by the combined influence of exposure and effortful yet successful retrieval, whereas learning during tests without feedback is driven primarily by retrieval success, with little additional contribution from retrieval effort. These findings indicate that retrieval supports learning through different mechanisms depending on the constraints of the learning context. More broadly, this mechanism-based account moves beyond categorical distinctions such as study versus test or massed versus spaced practice (for reviews, see Cepeda et al., 2006), and instead highlights the conditions under which learning updates occur. By identifying which learning mechanisms are primary under different conditions, the present work helps clarify how exposure-, retrieval-, and error-related accounts relate to one another, advancing a more precise understanding of how experience shapes memory over time.

Acknowledgments

This research was supported by the National Science Foundation under Grant No. 2301130 to Paulo F. Carvalho.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The Value of Applied Research: Retrieval Practice Improves Classroom Learning and Recommendations from a Teacher, a Principal, and a Scientist. *Educational Psychology Review*, *24*(3), 437–448. https://doi.org/10.1007/s10648-012-9210-2

Akaike, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer New York. https://doi.org/10.1007/978-1-4612-1694-0_15

Anderson, J. R., & Lebiere, C. J. (2014). *The Atomic Components of Thought* (0 ed.). Psychology Press. https://doi.org/10.4324/9781315805696

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the Spacing Effect in Advertising: Encoding Variability, Retrieval Processes, and Their Interaction. *Journal of Consumer Research*, *32*(2), 266–276. https://doi.org/10.1086/432236

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Bjork, R. A. (1975). Retrieval as a Memory Modifier: An Interpretation of Negative Recency and Related Phenomena. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123–144). Lawrence Erlbaum.

Bjork, R. A. (1994). Memory and Metamemory Considerations in the Training of Human Beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition* (pp. 185–206). The MIT Press. https://doi.org/10.7551/mitpress/4561.003.0011

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes, Vol. 1. From learning theory to connectionist theory; Vol. 2. From learning processes to cognitive processes* (pp. 35–67). Lawrence Erlbaum Associates, Inc.

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.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning.* John Wiley & Sons, Inc.

Cao, M., Pavlik, P., & Bidelman, G. (2019). Incorporating prior practice difficulty into performance factors analysis to model Mandarin tone learning. *EDM 2019 - Proceedings of the 12th International Conference on Educational Data Mining*, 516-519. Retrieved from https://digitalcommons.memphis.edu/facpubs/8060

Cao, M., Yan, V. X., Sana, F., & Carvalho, P. F. (under review). Balancing Spacing and Repetition for Time-Constrained Learning. https://doi.org/10.31219/osf.io/f6xu2_v3

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, *1*(9), 496–511. https://doi.org/10.1038/s44159-022-00089-1

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction*, *4*(4), 253–278. https://doi.org/10.1007/BF01099821

Darabi, A., Arrington, T. L., & Sayilir, E. (2018). Learning from failure: A meta-analysis of the empirical studies. *Educational Technology Research and Development*, *66*(5), 1101–1118. https://doi.org/10.1007/s11423-018-9579-9

Ebbinghaus, H. (1948). Concerning memory, 1885. In W. Dennis (Ed.), *Readings in the history of psychology.* (pp. 304–313). Appleton-Century-Crofts. https://doi.org/10.1037/11304-034

Fiechter, J. L., & Benjamin, A. S. (2018). Diminishing-cues retrieval practice: A memory-enhancing technique that works when regular testing doesn't. *Psychonomic Bulletin & Review*, *25*(5), 1868–1876. https://doi.org/10.3758/s13423-017-1366-9

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207. https://doi.org/10.3758/BF03212979

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, &amp Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

Jackson, A., Godwin, A., Bartholomew, S., & Mentzer, N. (2022). Learning from failure: A systematized review. *International Journal of Technology and Design Education*, *32*(3), 1853–1873. https://doi.org/10.1007/s10798-021-09661-x

Kapur, M. (2008). Productive Failure. *Cognition and Instruction*, *26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kapur, M. (2015). Learning from productive failure. *Learning: Research and Practice*, *1*(1), 51–65. https://doi.org/10.1080/23735082.2015.1002195

Karpicke, J. D., & Aue, W. R. (2015). The Testing Effect Is Alive and Well with Complex Materials. *Educational Psychology Review*, *27*(2), 317–326. https://doi.org/10.1007/s10648-015-9309-3

Karpicke, J. D., & Roediger, H. L. (2008). The Critical

Importance of Retrieval for Learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Karpicke, J., & Roediger III, H. (2007). Repeated retrieval during learning is the key to long-term retention☆. *Journal of Memory and Language*, *57*(2), 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1474–1486. https://doi.org/10.1037/xlm0000486

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227. https://doi.org/10.1016/S0959-4388(97)80010-4

Pavlik, 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. https://doi.org/10.1207/s15516709cog0000_14

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101–117. https://doi.org/10.1037/1076-898X.14.2.101

Pavlik Philip I., Cen Hao, & Koedinger Kenneth R. (2009). Performance Factors Analysis: A New Alternative to Knowledge Tracing. In *Frontiers in Artificial Intelligence and Applications*. IOS Press. https://doi.org/10.3233/978-1-60750-028-5-531

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Rachatasumrit, N., Carvalho, P. F., Li, S., & Koedinger, K. R. (2023). Content Matters: A Computational Investigation into the Effectiveness of Retrieval Practice and Worked Examples. In N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (Vol. 13916, pp. 54–65). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_5

Rawson, K. A., Vaughn, K. E., Walsh, M., & Dunlosky, J. (2018). Investigating and explaining the effects of successive relearning on long-term retention. *Journal of Experimental Psychology: Applied*, *24*(1), 57–71. https://doi.org/10.1037/xap0000146

Recorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.

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). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734–760. https://doi.org/10.1037/0033-295X.103.4.734

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2). https://doi.org/10.1214/aos/1176344136

Tawfik, A. A., Rong, H., & Choi, I. (2015). Failing to learn: Towards a unified design approach for failure-based learning. *Educational Technology Research and Development*, *63*(6), 975–994. https://doi.org/10.1007/s11423-015-9399-0

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 529–536. https://doi.org/10.1016/0022-5371(76)90047-5

Thorndike, E. L. (1913). *Educational psychology, Vol 2: The psychology of learning.* Teachers College. https://doi.org/10.1037/13051-000

VanLehn, K. (1999). Rule-Learning Events in the Acquisition of a Complex Skill: An Evaluation of Cascade. *Journal of the Learning Sciences*, *8*(1), 71–125. https://doi.org/10.1207/s15327809jls0801_3

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition*, *2*(4), 775–780. https://doi.org/10.3758/BF03198154

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, *25*(5), 731–739. https://doi.org/10.3758/BF03211316

Wollstein, Y., & Jabbour, N. (2022). Spaced Effect Learning and Blunting the Forgetfulness Curve. *Ear, Nose & Throat Journal*, *101*(9_suppl), 42S-46S. https://doi.org/10.1177/01455613231163726

Yan, V. X., Sana, F., & Carvalho, P. F. (2024). No Simple Solutions to Complex Problems: Cognitive Science Principles Can Guide but Not Prescribe Educational Decisions. *Policy Insights from the Behavioral and Brain Sciences*, *11*(1), 59–66. https://doi.org/10.1177/23727322231218906