From: *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*

Predicting Learning and Retention of a Complex Task Using a Cognitive Architecture

Farnaz Tehranchi[#] (farnaz.tehranchi@psu.edu)
Jacob D. Oury[†] (ouryjacob@psu.edu)
Frank E. Ritter[†] (frank.ritter@psu.edu)
Department of Computer Science and Engineering
† College of Information Sciences and Technology
Penn State, University Park, PA 16802 USA

Abstract

We use a model to explore the implications of ACT-R's learning and forgetting mechanisms to understand learning and retention on a complex task. The model performs a spreadsheet task that has 14 non-iterated subtasks. The model predicts a learning curve and knowledge decay for different learning stages. The model's learning curve fits the human data well for the first four trials without decay. When decay is examined, however, we have to make modifications to the retention equation for the model's predictions to match data and the shapes predicted by the other learning theories. To fix this anomaly, we modified the effect of time on decay (adjusting time outside the experiment to less than the effect of time in the experiment) and the strength of newly learned memories (less well known than the previous default value). From these results, we learn that training and testing have been confounded in many studies.

Keywords: learning; decay; ACT-R; cognitive architecture; procedural knowledge

Introduction

The study of learning theories often presents a three-stage framework to describe the progression from novice to expert. Kim, Ritter, and Koubek (2013) provided a review of learning theories presented as a summary theory. Their theory is primarily based on learning theories by Fitts (1964), Rasmussen (1986), and VanLehn (1996). Their theory is also consistent with further work in that review, as well as other theories of learning (e.g., Posner, 1973) and other data on learning (e.g., Seibel, 1963). Figure 1 shows a diagrammatic representation of this theory.

Here, we will test this theory. We use a task model in the ACT-R cognitive architecture (Byrne, 2012; Newell, 1990; Ritter, Tehranchi, & Oury, 2018) to make predictions about learning and forgetting, including where the learning stages might appear in short and long decay. We examine and illustrate this theory's predictions using existing data (Kim & Ritter, 2015) and modify an existing model (Paik, Kim, Ritter, & Reitter, 2015; Tehranchi & Ritter, 2018a, 2018b). The paper concludes with insights and new predictions derived from incorporating new schedules and memory types to ACT-R.

Review of Related Work

We start with the KRK theory (Kim, 2008; Kim et al., 2013) and its predictions. Table 1 summarizes predictions from the theory. Figure 1 presents further predictions that can be derived. The six predictions are supported by the theories and data noted in the Kim et al. (2013) review. We review some of the data and theories that support the KRK theory and discuss some further empirical and theoretical work that may provide further support and limitations for the theory.

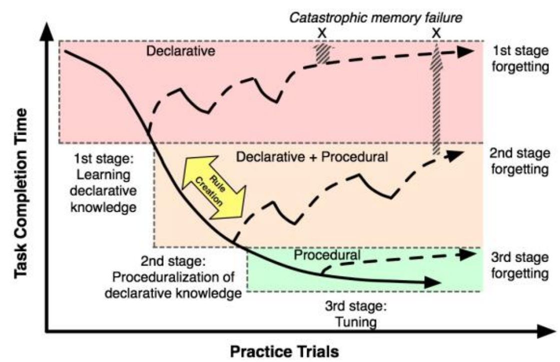

Figure 1: The KRK theory and some of its implications. Taken from Kim et al. (Kim & Ritter, 2016; 2013).

Table 1: Predictions for human performance based on the KRK theory, including established predictions derived from Kim et al. (2013) and Kim and Ritter (2015).

Predictions (Kim et al., 2013; Kim & Ritter, 2015)	
(1)	Learning follows the power-law curve of learning Time = A + BN^{-C} (A, B, C are constants)
(2)	Three stages of knowledge: Acquiring declarative and procedural knowledge Consolidating the acquired knowledge Tuning the knowledge towards overlearning
(3)	Retention of declarative knowledge decays quickly and catastrophically
(4)	Retention of mixed declarative and procedural knowledge decays moderately
(5)	Retention of proceduralized knowledge has the least decay
(6)	Recognition and perceptual-motor knowledge have different learning curves than procedural or declarative.

The KRK theory's predictions

In this paper, we propose two new predictions that are developed by considering the inclusion of perceptual-motor and recognition memory within the KRK theory: (a) Ideal training schedules will vary by knowledge-type; perceptual-motor may require minimum training block size and (b) Retention of perceptual-motor knowledge appears to decay little. We review each of the predictions for human performance.

A major prediction of the KRK theory is the rapid degradation of new declarative memories. This idea has been proposed in several theories, such as the ACT-R architecture and the levels of processing framework (Craik & Lockhart, 1972).

The forgetting curve and catastrophic decay have been demonstrated in several experiments and models. For example, the effects of massed or distributed practice on simple declarative memory tasks (*e.g.,* word-pair memorization, typing) have been modeled using ACT-R in several experiments (Anderson, 1993; Anderson, Fincham, & Douglass, 1999; Pavlik, 2007; Pavlik & Anderson, 2005).

Simple declarative tasks will show simple outcomes for forgetting (*e.g.,* lack of recall leads to no performance). For example, the optimal practice schedule for learning complex subjects implies a similar catastrophic memory failure for low-practice memories (Pavlik, 2007; Pavlik & Anderson, 2008). Complex tasks, with multiple knowledge sources (Kim et al., 2013), will have more complex outcomes (Halverson, Gunzelmann, Moore, & Van Dongen, 2010; McKenna & Glendon, 1985).

While this theory has been proposed within the context of ACT-R and Soar (Laird, Newell, & Rosenbloom, 1987), the research mostly falls short in reporting and modeling catastrophic forgetting for complex, multi-step tasks. In a complex task, the results where the task or subtask is forgotten are more complex. In the case of memory failure, a performer might have several strategies to recover and measure small failures, and it may be difficult to understand how behavior changes. With decay, the performance may slow down, may shift to be environmentally led (rather than recall-led), and might skip, invert, or invent steps rather than completely fail (Brown & Burton, 1980; Brown & VanLehn, 1980; Burton, 1982; VanLehn, 1982). These effects will make computing a summary score of performance more difficult.

Anderson (1993) studied learning, retention, and transfer of programming algorithms. They assumed that the knowledge was (essentially) declarative. They found that time on task (thinking time) for the second trial was about 46% of trial 1 when done immediately, but showed significant increases after even a 20-minute delay (56% of trial 1 with a 20-minute delay) and further performance loss after 24 hours delay (67% of trial 1). Compared to performing the second trial immediately, the delay causes statistically significant performance loss after only 20 minutes.

Paik and Ritter (Paik, 2011; Paik & Ritter, 2016) showed the effects of training schedules and training strategies on learning varies by knowledge type. Their study investigated three knowledge types: declarative, procedural, and perceptual-motor. They introduced a hybrid practice that is a mixed training schedule that blends distributed and massed practice. Then, hybrid-distributed, hybrid-massed, distributed, and massed training schedules were compared. This study's results show hybrid training schedules were able to predict and produce better performance than purely distributed or massed training schedules. In other words, the results indicate training schedules with some spacing and some intensiveness may lead to better performance. Unlike other studies and theories, perceptual-motor and recognition were considered in the Paik study.

Therefore, the types of knowledge used during tasks are expected to influence the optimal training schedule appropriate for any given task. This differentiation by knowledge type is proposed in several theories (e.g., Paik & Ritter, 2016).

For declarative memories (e.g., learning new foreign language nouns), Pavlik (2007; Pavlik & Anderson, 2008) used experimental data and later ACT-R modeling to show that the ideal practice schedule is distributed and follows a non-linear approach. It is determined according to each declarative item's expected memory strength on each trial.

For procedural tasks like solving math problems, the ACT-R theory predicts that the distributed practice schedules become superior in the long-term (Anderson et al., 1999; Pavlik, 2007; Pavlik & Anderson, 2005).

Work by Anderson, Fincham, and Douglass (1999) collected experimental data on the long-term retention of a rule-based task (up to 400 days between trials) following delays between training sessions. As subjects completed upwards of 240 trials, the performance was compared within-days and between sessions. They found that ACT-R's activation equations could account for performance changes within the experimental periods. Still, the scaling of time outside of the task was necessary to account for the asymptotic forgetting occurring over the significant time between trials (Pavlik & Anderson, 2005).

Overall, there are several problems with the empirical support for the complete set of hypotheses in Table 1. The forgetting or retention curves often are from single points of learning rather than at different levels of learning (e.g., Kim, 2008). In addition, most of the studies that are used to derive and support these hypotheses consider only simple tasks, such as choice reaction times (e.g., Pashler and Baylis, 1991; Seibel, 1963), word association and vocabulary (e.g., Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Taatgen and Anderson, 2002), mental arithmetic (Tenison & Anderson, 2016), and most of the tasks in Newell and Rosenbloom (1981).

Studying and validating the KRK theory predictions could be best accomplished with an empirical study of a complex, multi-step task with multiple training sessions and longer retention intervals. This study would need to explore the full set of predictions and would be a significant commitment.

To prepare for such a study, we use a cognitive model of a complex task built-in architecture with learning and forgetting and multiple skill representations. With this model, we explore the study and assess the model's predictions for learning and retention under a broader array of situations, including variations in practice frequency and schedule.

Method

We next describe the theoretical components used to generate quantitative predictions as an example of a complex task. We first describe the Dismal task. Dismal is used for three reasons: (a) it is complex and has been previously used to study learning, (b) we already have some human performance data on the task for different schedules, and (c) there is a functional running model that shows learning.

We then describe the modeling architecture, a model of a person performing the task, and the human data. The data provides insights regarding running the model with decay-scale and predicting learning and retention in a complex task using this model.

The Dismal task

The Dismal task is a complex spreadsheet task (Table 2) that can be used to measure procedural knowledge and skills learning and decay (Kim & Ritter, 2015). The revised and interactive Dismal model performs the task in the task environment using a new tool (JSegMan) that allows cognitive models to interact with the same, uninstrumented interfaces (Tehranchi & Ritter, 2018a, 2018b). The revised Dismal model is used here. The Dismal task consists of fourteen unique subtasks such as opening a file in Emacs that require attention shifts, encoding information, attending to information, key presses, and mouse moves/clicks. This range of actions allow us to study different types of knowledge: recall of keystroke commands and recognition of menu-based commands. Participants completed the task each day for 4 days. Then, participants were divided in three separate delay groups and completed the task after 6, 12, or 18 days.

Table 2: The 14 Dismal spreadsheet subtasks and sequence.

Dismal Procedure	
(1)	Open a file, named normalization.dis under the "experiment" folder
(2)	Save as the file with your initials
(3)	Calculate and fill in the frequency column (B6 to B10)
(4)	Calculate the total frequency in B13
(5)	Calculate and fill in the normalization column (C1 to C5)
(6)	Calculate the total normalization in C13
(7)	Calculate the length column (D1 to D10)
(8)	Calculate the total of the "Length" column in D13
(9)	Calculate the Typed Characters column (E1 to E10)
(10)	Calculate the total of the "Typed Characters" column in E13
(11)	Insert two rows at A0 cell
(12)	Type in your name in A0
(13)	Fill in the current date in A1 using a command
(14)	Save your work as a printable format

ACT-R

ACT-R is a theory of the mechanisms that make up cognition. It is an example of a unified theory of cognition (Byrne, 2012; Newell, 1990; Ritter et al., 2018) that intends to predict and explain human behavior by simulating the steps of cognition with a fixed set of mechanisms. ACT-R predicts behavior and activation of brain regions by using mechanisms, including procedural and declarative knowledge, and working memory as activation, to perform tasks.

The ACT-R Memory equation to calculate learning for chunks during the task procedure that is used in this work is Optimized Learning (OL) is shown in Equations 1. Equation 1 uses variables n, d, L, and β_i.

Equation 1: The Optimized Learning Equation (OL)

$$B_i = \ln\left(\frac{n}{1-d}\right) - d * \ln(L) + \beta_i$$

***n*:** The number of presentations for chunk *i*
***d*:** The decay parameter set using the :bll parameter
***L*:** The lifetime of chunk *i* (the time since its creation)
β_i: A constant offset set using the :blc parameter

We previously (Oury, Tehranchi, & Ritter, 2018) compared the performance of memory equations in their accuracy and computational cost. The OL equation simplifies the equation to primarily rely on the number of presentations and performance more accurately; a paired-sample t-test between human task time (N=30) and model (N=5) predictions of task time was significant (t=-2.538, p<.05). This model has 29 production rules and 1,159 declarative memory elements. The decay was set to d = 0.25, and the noise was 0.15 in ACT-R 7.5.

The Dismal spreadsheet task demonstrated that the learning and task knowledge from four practice trials did not decay catastrophically after six, twelve, or eighteen days without completing the task.

This paper extends previous work on how time-based decay can be more accurately implemented in ACT-R (Oury et al., 2018; Tehranchi & Ritter, 2018a). In addition to the previous work, we analyze the Dismal Task model under different conditions (i.e., training schedules and ACT-R parameters). The results below (a) present solutions that can make ACT-R more robust and accurate for modeling tasks over the course of days or weeks, and (b) extend the theory on the cognitive implications of procedural learning and the KRK theory.

Results

A series of models are run to explore how the ACT-R model of the Dismal task predicts learning and decay. The model runs using ACT-R's optimized learning (Equation 1).

We use a two-part naming convention that describes the decay conditions on the model. The first part is a proportion of the model's decay-time (i.e., internal ACT-R time) to real-time (as experienced by participants). For example, a condition labeled 1:4 would have 1 hour of ACT-R time decay for every 4 hours of real-time passage (i.e., 1:4 means ACT-R time is ¼ of real-time). The second part of the name indicates whether decay is applied between the first four practice trials. A model labeled "incorporated decay" indicates a 23-hour decay begins after the first trial, and a model labeled "delayed decay" indicates decay begins only after the fourth training trial.

Overview of results

Figure 2 shows that the human data (black X's) for days 1-4 match the Dismal model predictions (blue triangle) for trials 1-4 without daily decay. This is how models have typically been run for repeated trials (e.g., (Paik et al., 2015); Tehranchi and Ritter, 2018a), even when the training sessions are on separate days.

To explore retention, we adjusted the model's knowledge by adding corresponding decay periods for days 10, 16, and 22 to the model and running it again. Figure 2 shows that these human data (black X's) at retention are not particularly close to the predictions in blue triangles. For example, on day 16, the model predicts that performance will be essentially indistinguishable from a novice learning the task for the first time (Trial 1 ≈1300s on day 1 vs. Trial 5 ≈1600s on day 16). Yet, the human data clearly suggests that more knowledge is retained. Thus, the model appears to need some correction to how quickly knowledge decays.

To further understand how realistic decay affects the model predictions, we also tested the model when decay was added between every single trial, not just after trial 4 (as shown by the red line with circle icons in Figure 2).

We added this because the participants did not, in fact, perform the task four times in a row but performed the task, and then waited a day and performed the task again. This curve shows trial 2 for the model—the 1:1 Incorporated Decay condition—taking nearly twice as long as trial 1, even though the task has already been practiced once.

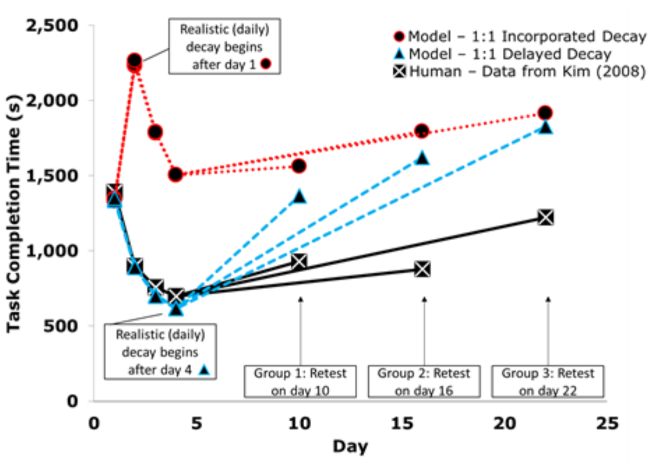

Figure 2: A comparison between human performance (N=30) and Dismal model (N=5) predictions when decay occurs between trials.

These large divergences from the human data suggest that the memories and learning modeled in ACT-R are being overpowered by time-based decay (red circles), even for established memories (blue triangle).

The human data shows that repeated daily practice leads to lower task times and some retention of the task knowledge, even after a delay of up to 12 days (on day 16). Yet, the model is unable to make useful predictions for that type of extended practice schedule. This issue is explored further below.

What happens when the decay length is adjusted?

We explore the effects of decay by including a 24-hour decay period after each practice to simulate the passage of real-time as experienced by the participants. The addition of time-based decay to the model is intended to provide greater accuracy for the simulated performance, and further analysis of different time scales may provide a solution.

We test the model predictions with different time scaling factors. Figure 3 shows how different time scales for inter-trial time affects the model's predictions. Different scales for the decay period length are based on the psychological time parameter estimates from previous ACT-R research (Anderson et al., 1999; Pavlik & Anderson, 2005).

The addition of decay periods to the model leads to unrealistic predictions in later days. Model–1:1 Incorporated Decay—shows how 24 hours of decay causes Day 2's time to be ≈900s slower than the initial trial (≈1300s to ≈2200s). Further tests with adjusted time scaling factors of 1:2 and 1:4 brought the results closer in line with the human data. Still, they did not adequately model the expected results as demonstrated by the large spike on Day 2.

We next adjust the decay time scale even further, as seen in some other ACT-R studies (e.g., Anderson et al., 1999; Pavlik and Anderson, 2005). Figure 4 expands upon Figure 3 by including additional time scales, 1:10 and 1:20, alongside the 1:1 decay time, 0:1 decay time, and the human data.

The model predictions with the 1:10 and 1:20 time scales provide stronger correlations with the human data than lower time scale adjustments. It takes reducing the time scale from 1:1 to 1:20 to finally model a learning and retention curve for this task that does not spike upwards on the second trial (i.e., day 2 being slower than day 1). RMSE between the human and model data showed that the 1:20 time scale was closest to the human data for OL (Human vs. 1:20 Incorporated Decay, OL: *RMSE* = 178.64).

These findings are consistent with Pavlik and Anderson's (2005) ideas that memory decays more slowly between trials with regards to time-based decay. After testing up to a 1:20 time scale adjustment, we find that including a time-scale adjustment improved the RMSE and correlations.

We pushed this modification further to assess where the optimal time-scale adjustment may be. Table 3 shows the additional time scale adjustments that were computed and showed that 1:25 incorporated decay had the highest correlation and lowest RMSE. Figure 5 shows the model predictions for 1:25 incorporated decay, the best fit of the model. In all figures, error bars are smaller than the symbol size.

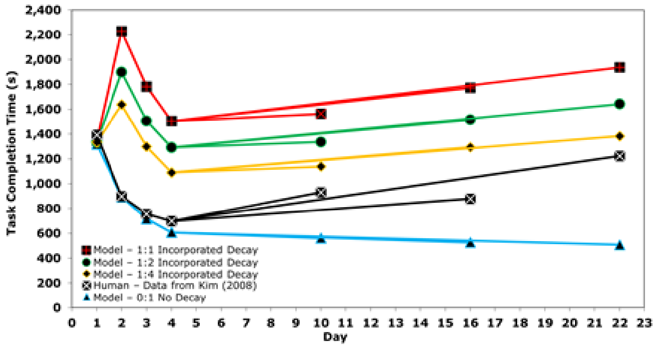

Figure 3: Predictions for task completion time when adjusting the length of decay periods. The model simulates four consecutive days of trials followed by either 6, 12, or 18 days without practice.

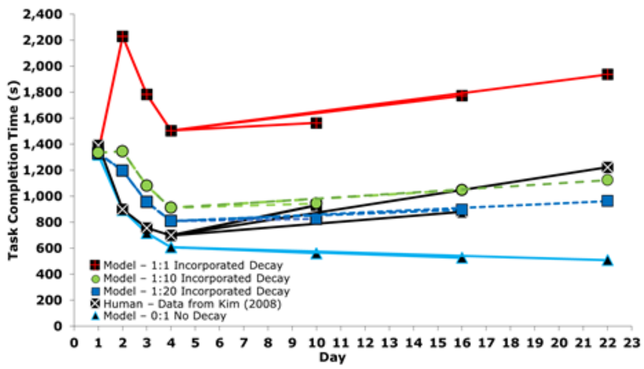

Figure 4: Predictions for task completion time when adjusting the decay periods' length, using optimized learning equation. The model simulates four consecutive days of trials followed by either 6, 12, or 18 days without practice.

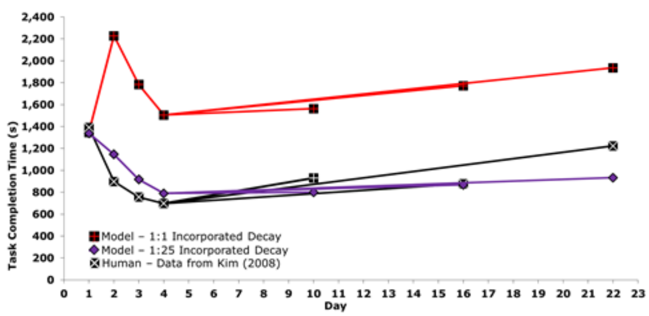

Figure 5: Predictions of task time compared to human data for the 1:25 adjusted length of decay period. The 1:1 decay condition is also shown for comparison.

What and how the model learns

To further our analysis, we look into the models' performance, and we describe how they learn based on new metrics: productions fired, declarative memories (DMs) retrieved, and mean chunk activation. Figure 6, in addition to response time, shows the number of productions that were fired and used in trials, the number of DMs that were retrieved, and the average activation value for all chunks in trials. The results illustrate the number of fired productions only dropped after the first trial, but with a longer decay, fewer productions were fired.

Again, we notice the same pattern for the number of declarative memories retrieved. The model gets faster while with lower activation values. For instance, in the second trial, the model uses the trial one activation values and new production rules that are generated by the first trial.

So, this model suggests that after proceduralization, the model gets faster in later trials through increased declarative memory strengthening (tuning in DM), not further rule learning. This is a single model of a single task, so we can only speculate about learning in other tasks, which we suspect may differ. The model's learning is almost entirely ascribable to declarative memory decay is apparent in Figure 6, showing the number of fired production rules and chunks being used by the model do not change with practice. Chunk's activation and the associate time to retrieve chunks mostly changed.

Table 3: Comparing the fit of learning and retention curves to human data for the Dismal task with time scale adjustments and the OL equation. The best fit is in bold.

Model Condition	RMSE (Human vs. Model)	Correlation (Human vs. Model)
1:10 Incorporated Decay	236.45	.60
1:20 Incorporated Decay	178.64	.66
1:25 Incorporated Decay	**168.73**	**.70**
1:30 Incorporated Decay	177.17	.68
1:40 Incorporated Decay	185.94	.68

Discussion and Conclusions

This model provides several insights about learning and the ACT-R memory procedure that helps to understand gradually how to model memory and what to model in memory during learning and decay.

There are many models of memory that do not include default decay outside of experiments or across days between sessions. Here, we show at a ratio of 1:25, the effect of outside-lab time on decay is small. ACT-R's decay function does not appear to predict cognition and decay when the model is applied across multiple days of a task. While realistic memory decay is expected to affect task performance, the model predictions demonstrate that ACT-R's decay model exaggerates the performance loss found in human data for tasks trained over consecutive days.

These effects may be caused by ACT-R's linkage between activation of a memory and its associated decay rate. Because activation drives both recall probability and decay rate for a memory chunk, ACT-R's current memory equations fail to account for spacing effects on relearning during multi-day learning. An alternate model of learning and relearning called the predictive performance equation (PPE) separates storage strength from retrieval strength in determining activation, and this may account for ACT-R's

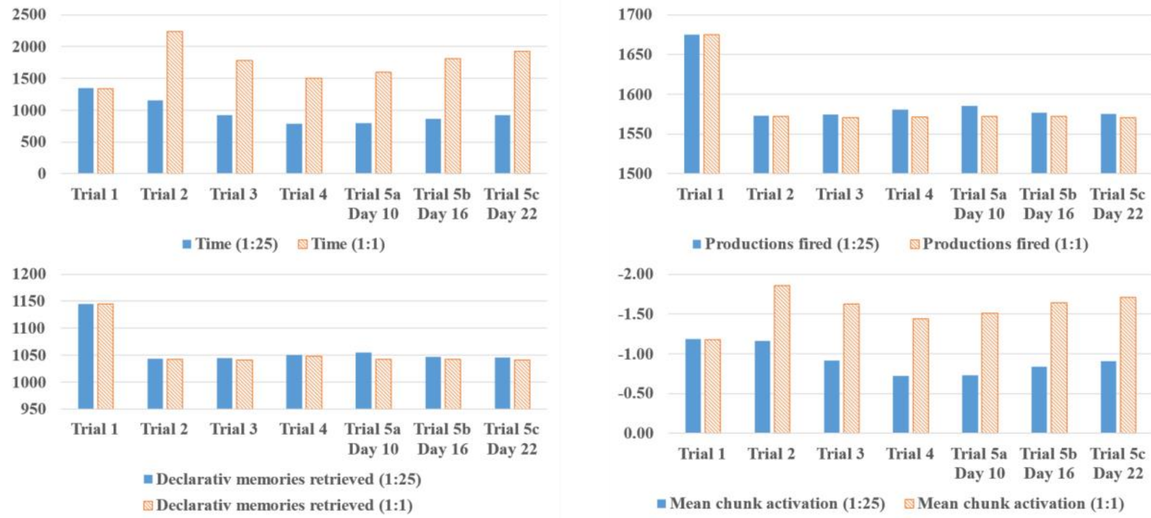

Figure 6: Time, productions fired, declarative memories retrieved, and mean chunk activation plotted for the 1:1 model and 1:25 model. Trial 5 occurs on either day 10 (5a), 16 (5b), or 22 (5c) following the standard 4-day training period.

deficiencies described here (Walsh, Gluck, Gunzelmann, Jastrzembski, & Krusmark, 2018; Walsh, Gluck, Gunzelmann, Jastrzembski, Krusmark, et al., 2018). Future work should consider whether ACT-R's memory equations would be improved by separating recall probability and decay similar to Bjork and Bjork's new theory of disuse (Bjork & Bjork, 1992). This effect might also be caused by interference of other memories and processing, which these models do not have.

As a final exploratory analysis on improving the Dismal Model's predictions, a pattern of parameter manipulations for parameters in ACT-R was generated that affect the declarative module and is compared with the 1:25 model predictions. Manipulating these parameters individually did not lead to a better fit than the default parameters that were used in the Dismal model. These manipulations were done one at a time. It is possible that combinations would lead to better fits or would be more appropriate, and there are tools that could be used to do this fit. Finally, Figure 6 raises interesting questions about how the model learns. Participants must be learning continually, but the model changes procedural knowledge only on the first day without any major strategy or knowledge representation changes in later days. These results do not illustrate the clear division in the learning stages shown in the KRK theory. However, it suggests that this task can be proceduralized at trial 1.

Looking at the subtask level model can provide more insights about where and when participants are learning complex tasks. Participants are not mastering the same sub-tasks each session. Still, they are often learning different subtasks each session or a different mix, but this suggests that perhaps individual subtasks follow the KRK theory better than the full task. Also, this work does not address retention decay and the task environment decay problem. New experiments are needed to examine this phenomenon, and more complex and additional tasks are required to test the KRK theory.

References

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(5), 1120-1136.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4*(5), 316-321.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes, 2*, 35-67.

Brown, J. S., & Burton, R. B. (1980). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2*, 155-192.

Brown, J. S., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science, 4*(4), 379-426.

Burton, R. (1982). Debuggy: Diagnosis of errors in basic mathematical skills. *Intelligent Tutoring Systems. London: Academic Press*.

Byrne, M. D. (2012). Unified theories of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(4), 431-438.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (Vol. 47, pp. 381-391). New York: Academic Press.

Halverson, T., Gunzelmann, G., Moore, L. R., & Van Dongen, H. P. A. (2010). Modeling the effects of work shift on learning in a mental orientation and rotation task. In *Proceedings of the 10th International Conference on Cognitive Modeling* (pp. 79-84). Philadelphia, PA.

Kim, J. W. (2008). *Procedural skills: From learning to forgetting.* Unpublished PhD thesis, Department of Industrial and Manufacturing Engineering, The Pennsylvania State University. University Park, PA.

Kim, J. W., & Ritter, F. E. (2015). Learning, forgetting, and relearning for keystroke- and mouse-driven tasks: Relearning is important. *Human-Computer Interaction, 30*(1), 1-33.

Kim, J. W., & Ritter, F. E. (2016). Microgenetic analysis of learning a task: its implications to cognitive modeling. In *Proceedings of the 14th International Conference on Cognitive Modeling* (pp. 21-26).

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science, 14*(1), 22-37.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33*(1), 1-64.

McKenna, S. P., & Glendon, A. L. (1985). Occupational first aid training: Decay in cardiopulmonary resuscitation (CPR) skills. *Journal of Occupational and Organizational Psychology, 58*(2), 109-117.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-51). Hillsdale, NJ: Erlbaum.

Oury, J. D., Tehranchi, F., & Ritter, F. E. (2018). Predicting learning and retention of a complex task. In *The 16th International Conference on Cognitive Modeling (ICCM 2018)* (pp. 90-95). Madison, WI.

Paik, J. (2011). *A novel training paradigm for knowledge and skills acquisition: Hybrid schedules lead to better learning for some but not all tasks.* Unpublished PhD thesis, Industrial Engineering, Penn State University. University Park, PA.

Paik, J., Kim, J. W., Ritter, F. E., & Reitter, D. (2015). Predicting user performance and learning in human-computer interaction with the Herbal compiler. *ACM Transactions on Computer-Human Interaction, 22*(5), 25.

Paik, J., & Ritter, F. E. (2016). Evaluating a range of learning schedules: Hybrid training schedules may be as good or better than distributed practice for some tasks. *Ergonomics, 59*(2), 276-290.

Pashler, H., & Baylis, G. C. (1991). Procedural learning: II. Intertrial repetition effects in speeded-choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(1), 33-48.

Pavlik, P. I. (2007). Timing is in order: Modeling order effects in the learning of information. In F. E. Ritter, J. Nerb, E. Lehtinen, & T. O'Shea (Eds.), *In order to learn: How the sequences of topics affect learning* (pp. 137-150). New York, NY: Oxford.

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.

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.

Posner, M. I. (1973). *Cognition: An introduction*. Glenview, IL: Scott Foresman.

Rasmussen, J. (1986). *Information processing and human-machine interaction. An approach to cognitive engineering*. New York, NY: Elsevier.

Ritter, F. E., Tehranchi, F., & Oury, J. D. (2018). ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1488.

Seibel, R. (1963). Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology, 66*(3), 215-226.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "broke"? A model of learning the past tense without feedback. *Cognition, 86*(2), 123-155.

Tehranchi, F., & Ritter, F. E. (2018a). Modeling visual search in interactive graphic interfaces: Adding visual pattern matching algorithms to ACT-R. In *Proceedings of ICCM - 2018-16th International Conference on Cognitive Modeling* (pp. 162-167). University of Wisconsin, Madison, WI.

Tehranchi, F., & Ritter, F. E. (2018b). Using java to provide cognitive models with a universal way to interact with graphic interfaces. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation.* (paper 50). Washington DC, USA.

Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 749.

VanLehn, K. (1982). Bugs are not enough: Empirical studies of bugs, impasses and repairs in procedural skills. *The Journal of Mathematical Behavior, 3*(2), 3-71.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*(1), 513-539.

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., & Krusmark, M. (2018). Evaluating the theoretic adequacy and applied potential of computational models of the spacing effect. *Cognitive Science, 42*, 644-691.

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., Krusmark, M., Myung, J. I., . . . Zhou, R. (2018). Mechanisms underlying the spacing effect in learning: A comparison of three computational models. *Journal of Experimental Psychology: General, 147*(9), 1325.