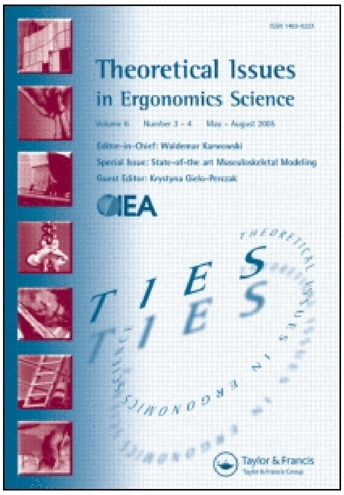

Theoretical Issues in Ergonomics Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713697886

An integrated theory for improved skill acquisition and retention in the three stages of learning

Jong W. Kim[a,b]; Frank E. Ritter[a]; Richard J. Koubek[c]
[a] College of Information Sciences and Technology, Pennsylvania State University, University Park, PA 16802, USA [b] Department of Psychology, University of Central Florida, Orlando, FL 32816, USA [c] College of Engineering, Louisiana State University, Baton Rouge, LA 70803, USA

First published on: 24 May 2011

Theoretical Issues in Ergonomics Science
2011, 1–16, iFirst

(2013). Theoretical Issues in Ergonomics Science. 14(1), 22-37.

An integrated theory for improved skill acquisition and retention in the three stages of learning

Jong W. Kim[ab]*, Frank E. Ritter[a] and Richard J. Koubek[c]

[a]*College of Information Sciences and Technology, Pennsylvania State University, University Park, PA 16802, USA;* [b]*Department of Psychology, University of Central Florida, Orlando, FL 32816, USA;* [c]*College of Engineering, Louisiana State University, Baton Rouge, LA 70803, USA*

(*Received 15 January 2010; final version received 22 February 2011*)

We introduce an integrated theory of learning and forgetting that has implications for training theory and practice. We begin with a review of skill acquisition research that argues that individuals employ different cognitive mechanisms when learning, which can be reliably associated with three stages of learning. This review leads to our proposed skill retention theory, which recommends a method to increase skill retention when designing systems, covering a range of system design issues, from interface design to training. We conclude with a discussion of how we might optimise skill retention based upon this approach. Specifically, we discuss how we might improve training by better spacing the iterations between training sessions to support proceduralisaion to improve skill retention.

Keywords: skill retention; cognitive architecture; learning; forgetting; training

1. Introduction

Training and education are designed to improve learning and produce qualified performance through retention of knowledge. However, individuals often forget important skills, which leads to decreased performance. Infrequent use of learned skills can cause skill decay. In this article, we discuss how facilitating the proceduralisation of these skill sets may strengthen skill retention.

Skill retention is especially significant in professions where individuals must successfully perform important skills that they rarely practice. For instance, non-medical trainees on a space flight who may have to rapidly perform advanced cardiac life support during the course of a flight, as well as their specific mission roles (Ramos and Chen 1994). This happens on earth as well. McKenna and Glendon (1985) found that only a quarter of 120 occupational first responders were still proficient 6 months after receiving cardiopulmonary resuscitation training.

The knowledge necessary for performing a task may be declarative, procedural or a mixture of declarative and procedural knowledge. Declarative knowledge represents factual information. Procedural knowledge indicates task knowledge, such as typing or

*Corresponding author. Email: jwkim@mail.ucf.edu

ISSN 1463-922X print/ISSN 1464-536X online

DOI: 10.1080/1464536X.2011.573008
http://www.informaworld.com

riding a bicycle. Management of both knowledge types is generally required for successful performance.

Cognitive architectures such as ACT-R (Anderson and Lebiere 1998, Anderson *et al.* 2004, Anderson 2007), Soar (Laird *et al.* 1987, Newell 1990) and EPIC (Kieras and Meyer 1997) represent each knowledge type and associated mechanisms. We chose to use ACT-R to describe our theory because it provides cognitive mechanisms representing procedural and declarative knowledge learning (Anderson 1982) and declarative forgetting (Pavlik and Anderson 2005). Other architectures are more restricted than ACT-R (Chong 2004, Kim *et al.* 2007). Soar provides a rule learning mechanism, chunking, but not forgetting. EPIC does not even provide a rule learning mechanism. ACT-R represents declarative knowledge as a propositional network of facts, called chunks, while representing procedural memory using production rules. Each production rule consists of a condition/action pair that is used in sequence with other productions to perform a task. The theory ascribes each rule to a cognitive step taken during the performance of a task (Anderson 1982).

In this article, we review the current understanding of learning and retention in terms of a unified theory of cognition – a theory like Newell's (1973, 1990) that proposes a theory of cognition for the complete processing of human performance models. Based on this understanding, we introduce a new theory of skill retention that integrates learning with forgetting into a broader view of improved skill retention. We then describe how we might implement this perspective as a cognitive model that is able to predict the learner's knowledge in relation to each learning stage, and to predict performance, specifically learning and forgetting rates.

2. Learning and forgetting

To better address the issues related to learning and long-term retention, it is necessary to predict the learner's future cognitive states. Cognitive modelling and simulation provide a way to do this. A cognitive architecture with a unified theory of cognition provides a platform where a cognitive model can be built and to provide accurate prediction of cognitive, perceptual and motor interactions of human performance (Kieras and Meyer 1997, Gray and Boehm-Davis 2000). This modelling and simulation thrust can reduce the cost of running experiments with participants to acquire forgetting data that might be expensive because it requires a study characterised by both training and forgetting intervals (obtained separately and sequentially) followed by a test. That is, once a cognitive model based on a cognitive architecture is validated with experimental data, the validated model can provide predictions of human performance, reducing the cost to evaluate systems and interfaces (Booher and Minninger 2003, Pew 2007, Pew and Mavor 2007, St. Amant *et al.* 2007, Bolton and Bass 2009, to note just a few who have argued for this approach and documented its potential savings).

2.1. *Earlier theories of learning*

Proctor and Dutta (1995) provide a good review of skill acquisition – including theories proposed by Fitts (1964), Anderson (1982) and Rasmussen (1986). Fitts posited three learning phases: *cognitive*, *associative* and *autonomous*. Drawing from Fitts, Anderson's (1982) theory of cognitive skill acquisition developed three corresponding stages entitled: *declarative*, *transitional* and *procedural*. Rasmussen, also influenced by Fitts, proposed a

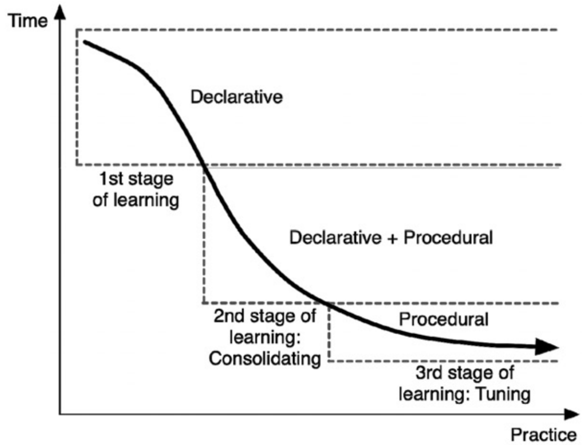

Figure 1. A theory of learning in the three stages, which is based on the theory of Fitts (1964), Anderson (1982), Rasmussen (1986) and VanLehn (1996).

framework pertaining to performance that differentiates task execution as knowledge-based, rule-based and skill-based. Finally, VanLehn (1996) also described three phases of cognitive skill acquisition: *early*, *intermediate* and *late*. These theories provide a consensus understanding of learning – all propose a three-stage process of learning, as shown in Figure 1, that includes: (1) acquiring declarative and procedural knowledge, (2) consolidating the acquired knowledge and (3) tuning the knowledge towards overlearning.

To represent this process, ACT-R encodes facts to support the acquisition of task knowledge; interprets the acquired information to produce behaviour; and, with sufficient practice, converts the acquired knowledge into procedural form using a mechanism called knowledge compilation (Neves and Anderson 1981, Anderson 1982, Anderson and Lebiere 1998, Jones *et al.* 2000), which correspond to chunking in Soar (Newell 1990) and proceduralisation elsewhere (Anderson 1982). After knowledge compilation, further tuning of the knowledge occurs in the third stage, producing speedup of the knowledge application process, which is referred to as the procedural stage.

Learning behaviour generally follows a regularity known as a power law of practice (Seibel 1963, Card *et al.* 1978, Newell and Rosenbloom 1981, Delaney *et al.* 1998, Ritter and Schooler 2001). The law describes the relationship between practice and skill proficiency. As practice trials increase, task completion times decrease but at a diminishing rate, resulting in a power law: $Time = Trials^{-\alpha}$, where α represents the rate at which performance time changes. This learning curve describes an important aspect of behaviour, providing a mathematical account of the general learning rate (Rosenbloom and Newell 1987), and supports building models of learning complex tasks that can be decomposed into smaller components (Lee and Anderson 2001).

2.2. *An example of learning*

To illustrate this theory, we present a simple Ergonomics example of learning typing skills (Anderson 1993). While common, typing has the same structure as many similar skills where declarative knowledge leads to procedural knowledge. This example can help us to better understand the distinction of memory between declarative/procedural knowledge and to illustrate the three stages of learning.

When learning to type, the learner generally first memorises the layout of the keyboard declaratively and learns to use the keyboard procedurally through practice. There are other strategies, of course, but this is commonly taught. Practicing typing enables the learner to memorise the keyboard layout and to type faster with practice. Over time (several months or more), the learner generally loses their declarative knowledge of the keyboard's layout but retains their procedural typing skills. Thus, once fully learned, few participants determine any key position declaratively (i.e. can retrieve directly where 'r' is), but rather rely exclusively on their proceduralised knowledge of the task (i.e. imagining typing a letter and seeing where their finger goes).

This example illustrates how individuals can and do maintain both declarative and procedural knowledge in memory, and how the kinds of memory utilised can be both dependent and independent at different stages. In addition, it suggests that procedural knowledge can be more robust than declarative knowledge. In the first phase, the learner depends almost exclusively on declarative memory elements to perform the task – this initial stage is both cognitively intensive and slow. In the second phase, the learner begins to rely more heavily upon procedural memory elements, but for some 'problematic keys' still rely on their declarative knowledge of the keyboard (*q* is above *a*, for instance). Finally, as the learner becomes more expert, they shift entirely or almost entirely to utilising their procedural memory. In addition, the transition from a primarily declarative to a procedural representation of the keyboard is associated with a reduced need for knowledge maintenance – lack of practice may result in slower typing speeds but not the entire loss of the skill.

This process is consistent with many experimental results. Early experimental work by Posner (1973) showed that procedural memory is more robust. In Posner's experiment, skilled typists were asked to label a diagram of a standard keyboard. He reported that the skilled typists had difficulty in recalling a visual location of a letter from the standard keyboard (declarative memory), whereas the typists could type the letters in a few seconds without errors. This example supports the claim that declarative knowledge of visual location can be degraded while procedural knowledge can remain robust against decay. It also suggests that long-term retention would be improved by turning declarative knowledge into procedural knowledge.

2.3. *A new look: an integrated understanding of learning and forgetting*

Elaborating on our earlier discussion of the stages of learning, we discuss each stage in more detail. We then examine how forgetting might vary reliably across the different stages of learning. We use Figure 2 as an overview of this discussion; the figure depicts a learning curve and a corresponding hypothetical forgetting curve across the three stages of learning. The main continuous line indicates continuous practice. Dashed lines indicate periods of inactivity (lack of practice), with solid lines showing later training. At each stage, the learning and forgetting rates are different. The colour-coding for each stage, shown as darker to lighter grey, indicates the level of performance – i.e. the third stage is a lighter grey, indicating that the performance is faster.

2.3.1. *The first stage: declarative*

In the first stage of learning, skill acquisition occurs and simple training focused on skill acquisition may be adequate. In this stage, learning and forgetting are accounted for by the activation mechanism for declarative knowledge in ACT-R. For this first stage, knowledge

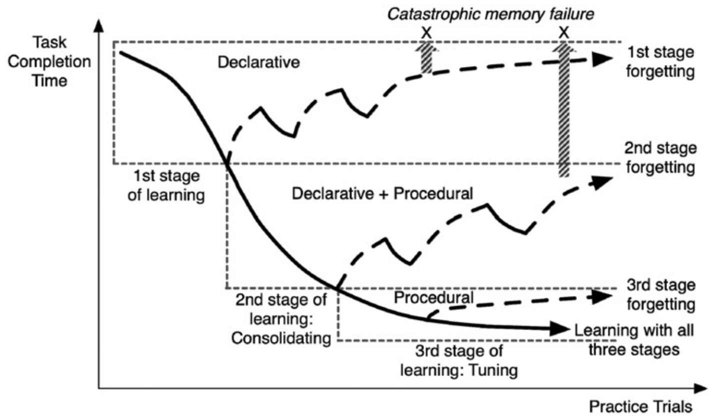

Figure 2. A graph describing a theory of skill retention, showing the three stages of learning and the effects of forgetting.

in declarative memory degrades with lack of use, perhaps catastrophically as indicated by X's in Figure 2, leading to the inability to perform the task. A catastrophic memory failure is a state where declarative memory items needed to perform the task cannot be retrieved from memory due to lack of practice (decay). That is, the activation value drops below a threshold, leading to retrieval failure in ACT-R.

With time the strength of declarative memory items declines. Lack of use leads to decreased memory strength, which leads to increased response times, decreased retention and decreased accuracy. In addition, the ACT-R theory suggests that increases in working memory load in this stage of learning lead to decrements of retrieval performance because the memory load introduces interference in the memory activation mechanism (Anderson *et al.* 1996). Thus, overall performance at this level of knowledge decreases with increases in working memory load. Referring back to our example of learning typing skills, this stage corresponds to the individuals knowing how to hit the keys based on retrieving their spatial location in reference to other keys.

2.3.2. *The second stage: mixed*

In the second stage of learning, task knowledge is represented using a mix of declarative and procedural memory. With lack of use, the declarative knowledge is forgotten, leading to mistakes and missed steps. Procedural memory, on the other hand, is basically immune to decay. This is also predicted by Soar and EPIC. The slope of the forgetting curve in this stage could vary by subtask because the subtasks would vary in their knowledge mix, and different mixes would decay at different rates. In the first and second stages, catastrophic memory failure can occur because the declarative knowledge is not fully activated. This result suggests that, in this mixed stage, training is necessary to keep the declarative knowledge active. Training of declarative memories is also necessary to support proceduralisation because declarative memories have to be active enough for new procedural rules to be generated – procedures can only be created when the declarative memories can be retrieved quickly enough that the architecture can solve the problem in the environment and that the architecture can create procedures.

In this stage, training should occur to avoid catastrophic memory failures in task performance resulting from the learner's inability to retrieve declarative knowledge for steps still requiring declarative knowledge elements, either because the task is incompletely proceduralised or because declarative inputs vary for the task, such as the radio frequencies or flight numbers used by pilots as they enter or exit controlled airspace. The higher decay rates associated with declarative memories also suggest that achieving more complete expertise entails training not only common but also uncommon tasks to the procedural level (e.g. emergency procedures for pilots). Furthermore, while the time between retraining sessions can increase with practice, catastrophic memory failures can still occur if completing the task requires declarative memory elements (i.e. changing inputs like frequencies, coordinates or names) for actions or to move to the next subtask.

In terms of our example of learning typing skills, this stage would correspond with the learner possessing a strong declarative representation of the key locations while also possessing some procedural rules regarding the task as retrievals increasingly get combined with key presses (in other words, the greater association of conditions with actions).

2.3.3. *The third stage: procedural*

In the third stage of learning, task knowledge is available in both declarative and procedural forms, but procedural knowledge predominantly drives performance. Practice will compile declarative knowledge about tasks into procedural knowledge. We describe this type of task knowledge as a *proceduralised skill*. With lack of use, declarative knowledge may degrade. Nevertheless, the learner can still perform the task – if (1) all the knowledge is proceduralised or is available in the environment and thus not forgotten with time and (2) performing the task does not require new declarative inputs.

Conversely, less practiced or infrequently used skills, such as responding to unusual errors, may still exhibit the mixed curve described in the previous section. This effect is shown in Figure 3, where different subskills (Task skills A, B and C) are learned and forgotten at different rates. Consequently, these skills require concerted and structured practice to proceduralise. Unlike the common ASCII keys in the typing task, there is no assurance that routine task execution will compile and proceduralise the declarative

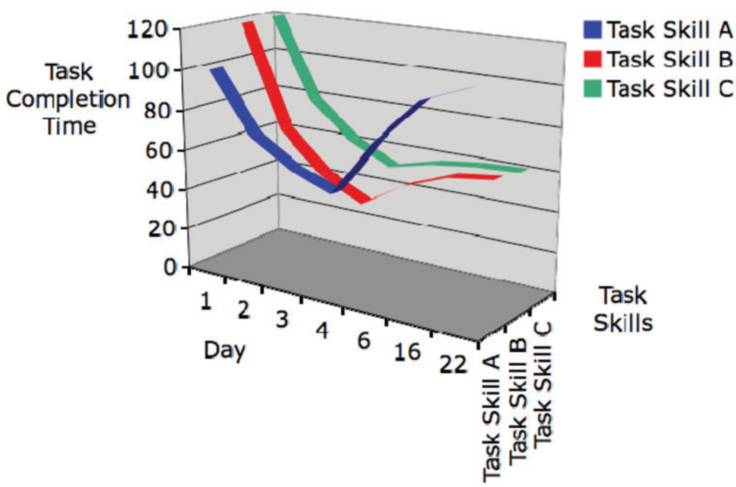

Figure 3. Subskills may be practiced, learned and forgotten at different rates.

memory elements associated with the task, meaning that training is most likely necessary to achieve proceduralisation (noted in Figure 2 as crossing the dashed grids representing the stages' thresholds).

With respect to the example of learning typing skills, in the fully proceduralised knowledge stage the knowledge about how to type would all be in production rules, and the declarative knowledge would not be used to type and would decay with time (Ericsson *et al*. 1993, and related work on expertise documents this effect as well, where experts have difficulty explaining how and why they behave). Infrequently typed characters (e.g. '#' or ']'), however, might still require declarative retrievals if these characters are insufficiently practiced.

3. The three stages of learning and forgetting in ACT-R

In this section, we describe how ACT-R supports this integrated understanding of learning and forgetting. More specifically, we explore to what degree the learning mechanisms in ACT-R are able to reproduce the patterns of skill learning and forgetting that we have discussed. We found that this analysis helped us to not only better understand the architecture's capabilities but also suggest steps for improving the modelling and prediction of learning and retention within a cognitive architecture.

3.1. *Learning mechanisms for declarative and procedural memory*

In ACT-R, the base-level activation of a declarative memory item (a chunk) is dependent on how often (frequency) and how recently (recency) the item is used. Whenever, an item in memory is used, the base-level activation increases and then decreases in accordance with a power function of the time since use. When the sum of the activations is above the retrieval threshold, the memory can be retrieved. When it is below, it cannot. Higher activations lead to faster and more accurate retrievals. Thus, the base-level learning mechanism supports the Power law of learning and the Power law of forgetting. Equation (1) shows that the base-level activation B, for a chunk i, where β is a constant, n the number of presentations for a chunk i, t_j the time since the *jth* presentation and d the decay parameter.

$$B_i = \beta + \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) \tag{1}$$

Production compilation, previously named knowledge compilation, is the current mechanism of production rule learning in ACT-R (Taatgen and Lee 2003). ACT-R uses this mechanism to learn new production rules by collapsing two productions into a single production. Production compilation combines both proceduralisation and composition mechanisms into a single mechanism. ACT-R generates a compiled rule by eliminating the declarative knowledge retrieval request in the first rule and the retrieval condition in the second rule.

ACT-R represents the cognitive intensive nature of declarative knowledge retrieval by allowing only one item in memory to be retrieved at a time. Production compilation allows a speed-up process by generating task-specific procedural knowledge where there previously was a string of declarative retrievals. Like the activation of declarative knowledge, productions have their own utility values. Based on these utility values, one production can be preferred and be selected over another. Also, the utilities can be learned from experience.

3.2. *Forgetting mechanisms for declarative and procedural memory*

If learning stops while the learner is in each of these stages, the degree of forgetting and the rates of forgetting will differ because the structures that the skill is based on are different (e.g. declarative, declarative + procedural or procedural), which is shown in Figure 2. Declarative memory elements in ACT-R decay with the passage of time. Procedural memory elements (i.e. production rules) are not affected by time.

While the ACT-R theory does not directly support the decay of procedural elements, skill can decay in ACT-R. Production rules in ACT-R can reference or rely on declarative memory elements. Thus, the application of procedural knowledge can still require declarative elements to be well learned enough to be retrieved. This suggests that procedural knowledge, in a sense, can be and in some senses has to be primed by declarative knowledge, that is, linked by declarative knowledge elements – generally, the term of prime is used in the context of declarative knowledge but we use the term here in a little bit extended context in which procedural knowledge is linked by declarative knowledge. The declarative memory elements necessary for applying procedural knowledge, however, can become inaccessible through disuse, leading to a phenomenon of forgetting of procedural knowledge – performance decrements of primed knowledge due to the disuse of that knowledge. Thus, within the theory, we can identify instances where procedural knowledge to perform a task may be retained but, without declarative knowledge available in the head or perhaps in the environment of the agent to trigger sections of procedural knowledge, skills can be catastrophically forgotten in that declarative knowledge to trigger or apply them is not available.

It has been noted that most cognitive architectures predict that procedural knowledge does not decay with time (Chong 2004). For example, Soar, ACT-R and EPIC assume that once procedural knowledge is learned, it does not decay. In Soar (Newell 1990, p. 164) and in ACT-R (Anderson 1993, p. 18), this is an explicit decision. Other architectures might not decay procedural knowledge because they are intended not to predict decay but just learning. Interestingly, we notice that little research has been conducted to study mechanisms of decay in procedural memory.

It may be worth exploring the production rule utility equation in ACT-R (i.e. change in the utility value of productions to model procedural memory decay), adding a parameter to the current architecture, or extending the existing architectural mechanisms. In addition, it will be necessary to build a model to test this claim and add this limitation if necessary, as well as to gather data in this area.

4. Training for improved skill retention

In the previous sections, we have introduced the rationale and components of our integrated approach that we will refer to henceforth as our *skill retention theory*. In this section, we explore the theory's potential implications for training.

4.1. *The spacing of training*

When it comes to the spacing of training (i.e. massed or distributed practice), *the skill retention theory* offers a more nuanced understanding about how to create improved learning and retention schedules that are generally found in the literature. During the progression of the learning (and forgetting) curve, the spacing of training theoretically affects performance of learning and retention because of the different knowledge structures used at each stage, which is theoretically straightforward in ACT-R. Thus, the optimal spacing of training can be determined by the learner's progression in the three stages of learning. If the learner's knowledge is within the first stage (the declarative learning stage), the learner's performance might be optimised by distributed practice. On the other hand, if the learner's knowledge is about to move into the third stage (procedural knowledge), the learner's performance would be optimised by massed practice, which makes the declarative knowledge strong enough to proceduralise. Also, for strategy shifts, massed practice may be required to strengthen declarative knowledge necessary to create and learn new strategies.

Upon review, we found that literature on the spacing of training offers contradicting findings on the benefits of massed versus distributed practice. We believe that our theory provides some explanation for these discrepancies, which we now explore through a series of examples.

Bahrick *et al.* (1993) investigated acquisition and retention of English-foreign language word pairs in a 9-year longitudinal study. In this study, Bahrick *et al.* presented a crossover interaction between the acquisition and retention performance in a vocabulary-learning task. Participants received 13 to 26 relearning sessions with three spacing intervals of 14-days, 28-days and 56-days. Then, retention was tested 1, 2, 3 or 5 years after training. On the last training session, the closest spacing intervals (a 14-day interval) produced the highest recall performance. On the other hand, after a year since the last training, the widest spacing of training (i.e. a 56-day) produced the highest recall performance, and the 14-day spacing produced the lowest recall performance.

This pattern in the order of recall performance in terms of the spacing intervals persisted throughout the range up to the 5-year retention measure, indicating that the more widely spaced training provided greater retention on this vocabulary-learning task. This finding shows that using the distributed spacing of training can degrade immediate acquisition and performance but can enhance long-term retention. Schmidt and Bjork (1992) mention that manipulations (e.g. using a spacing between sessions and using a random sequence of presentations) in retrieval practice can enhance retention in some tasks – in motor tasks (Shea and Morgan 1979) and in verbal recall tasks (Landauer and Bjork 1978).

Similarly, Rohrer and Taylor (2006) found distributed practice to be more effective in their study examining the retention of mathematical knowledge. In this investigation, 216 college students solved mathematics problems. In the first experiment, students were divided into four groups with regard to the spacing of training (i.e. massed or distributed) and the retention interval (i.e. 1-week or 4-weeks). The benefit of the distributed practice was significant for the 4-week retention interval, suggesting long-term retention of mathematics knowledge is better achieved when practice is distributed rather than massed.

In a subsequent experiment, Rohrer and Taylor examined the effect of overlearning on retention by varying the number of practice problems within a single session (massed practice). The finding indicates that there are no significant benefits of overlearning on mathematics knowledge retention for 1- and 4-week retention intervals. Based on these results, Rohrer and Taylor argue that long-term retention can be better augmented by distributed practice and is unaffected by overlearning, stating that most mathematics textbooks ironically emphasise massed practice. If this is a procedural task, then learning goes to the third stage, and overlearning cannot take them further.

Alternatively, Vearrier *et al.* (2005) found that massed practice was effective for patients undergoing intensive physical therapy. Vearrier *et al.* examined patients training to regain postural control after a stroke using constraint-induced movement therapy. This intervention is a task-oriented approach that utilises variability of practice with contextual interference. The intervention consisted of an intensive 6 h/day of one-on-one training for 10 consecutive weekdays. Testing the patients' performance, 3 months after the intervention, Vearrier *et al.* found that the patients' follow-up performance was also significantly improved. While this work does not test distributed practice, it suggests that there may be situations where massed practice is desirable for creating proceduralised skills that can be found in the third stage.

How then might we explain the discrepancy between Rohrer and Taylor's (2006) findings and those of Vearrier *et al.* (2005)? We suggest examining the skill types of each study. The types of task skills are different in those two studies – the former is a cognitive skill likely to have a large number of declarative components (i.e. mathematics knowledge) and the latter is a perceptual-motor skill with little declarative knowledge necessary to perform its final form (i.e. postural control task of walking).

These findings seem to suggest that: (1) different knowledge and skill types possess different profiles across the three stages of learning and forgetting, and that (2) skills differ as to which strategy best supports the long-term retention of that knowledge. Furthermore, we believe that the observed differences between skill types result from the cognitive mechanisms used to acquire that knowledge, which we can explain using our *skill retention theory* and potentially expressed in a cognitive architecture such as ACT-R.

4.2. *The amount of training: overlearning*

While Rohrer and Taylor (2006) found that overlearning did not noticeably improve performance, Driskell *et al.* (1992) identify instances where overlearning seems to have a positive effect. To clarify, we define overlearning as the immediate continuation of learning after the learner achieves mastery over a task. In general, it appears that overlearning is a common method to achieve long-term retention of knowledge. In this section, we discuss overlearning and its relationship to declarative memory.

Drawing from the meta-analytic investigation about the benefits of overlearning by Driskell *et al.* (1992), we discuss instances where overlearning appears to be an effective way to support long-term retention for both physical and cognitive tasks. In general, it appears that overlearning has a higher effect on cognitive tasks than physical tasks where there may be more declarative knowledge and multiple strategies and situations to know. Farr (1987) also notes that overlearning can reduce the rate of decay because the amount of overlearning can increase the amount of knowledge acquisition. Furthermore, Bahrick's (1984) long-term language study (i.e. over a 50-year period) suggests that the level of original training can play an important role in the retention performance over the long-term period.

Nevertheless, as noted earlier, there are some contradicting views when it comes to the benefits of overlearning. Particularly, Rohrer and Pashler (2007) argue that the benefits of overlearning can diminish over time and that overlearning can only boost performance for a short time. In an earlier study, it was found that the benefits of overlearning diminished after an extended retention interval of 4 weeks by comparing the effects of overlearning after 1- and 4-week retention intervals in a task of learning word pairs (Rohrer *et al.* 2005). Furthermore, the retention performance with overlearning is similar to the performance without overlearning after a 4-week retention interval.

Rohrer and Pashler (2007) claim that overlearning can produce better performance (e.g. faster task completion time or higher test scores), but that its benefits are primarily associated with short-term declarative knowledge. Thus, they advise that overlearning is helpful in instances where an individual (e.g. pilots, soldiers or nurses) must follow emergency protocols (Rohrer and Pashler 2007). Our theory agrees with this suggestion because the learners can move their knowledge in the third stage with overlearning. They also note, in the meanwhile, that such an approach can be supplemented by distributed training due to the diminishing benefits of overlearning over time, particularly for skills with a large declarative knowledge component.

Krueger's (1929) verbal recall study suggests that a certain degree of overlearning can be economical at a certain interval – 50% overlearning is economical when it comes to retention intervals of 2–28 days. Overlearning, however, seems an expensive strategy for most training situations. One important concern is, therefore, how to determine cost-benefit ratio of overlearning across a wide variety of tasks. Driskell *et al.* (1992) point out that the initial cost for overlearning may be offset by the lowered costs for retraining or refresher training – because participants who overlearned the task required fewer retraining trials to maintain proficiency (Schendel and Hagman 1982).

We have examined these contradicting views concerning overlearning to illustrate that at present there is no consensus understanding as to which stage the learner enters after overlearning – indicating a need for further research. We theorise that the learner can still suffer catastrophic memory failures if overlearning occurs either for declarative knowledge task or a mixed knowledge task because the learner must still rely on declarative knowledge to perform the task. If, however, overlearning helps the learner to move to the third stage, then it seems likely that learning and overlearning can contribute to proceduralisation of skills, making the acquired knowledge basically immune to decay, provided that task performance does not require new declarative inputs. We can then explain studies showing the benefits of overlearning with the acquisition of procedural skills in the third stage, and studies showing the lack of benefits to the learner's heavy dependence upon declarative knowledge for task completion.

The effectiveness of overlearning is, therefore, consistent with our skill retention theory that fully proceduralised knowledge (i.e. knowledge and skills in the third stage) exhibits a low decay rate. The question is then how we achieve moving into the third stage of proceduralised skill and whether overlearning is an efficient way to achieve this transition.

To achieve the third learning stage, first of all, it is necessary to practice the task skill in the first and second learning stages. If learning stops in the first or second stage, the learner can suffer from declarative knowledge retrieval failures. This may help explain when overlearning helps (when moving all of the task skill set into the third stage), and when overlearning does not help (when only the skill in the third stage is being practiced or when the skill is only a declarative retrieval), and when overlearning is inefficient (when only skills in the third stage are being practiced, and there remain subskills in the first and second stages, or the knowledge is only declarative). This provides insights for creating smart refresher training strategies – when trainees should practice and which task subskills to practice for enhanced retention.

When implementing these strategies, external aids may prove useful. Often, when the learners decide what to practice, they may prefer to practice what they already know for a later test. Automatic training has been shown to work better than self-directed learning for vocabulary learning (Atkinson 1972, Pavlik and Anderson 2008). Additionally, intelligent tutoring systems have proven effective for teaching formal tasks such as geometry proofs or learning a programming language (Anderson *et al.* 1985, 1989). Consequently,

Table 1. Suggestion for an optimal training policy in terms of stage and knowledge representation.

Stage	Knowledge representation	Optimal training
First stage	Declarative	Distributed, overlearning will help proceduralise tasks
Second stage	Mix of declarative and procedural	Distributed, but massed to move into Stage 3. Overlearning will help proceduralise tasks
Third stage	Mostly or exclusively procedural	Distributed to retain declarative components. Overlearning does not help

an external aid that identifies what to practice based on what the learner has not yet mastered (i.e. skills not yet known or in the first and second learning stages that can be moved to the third stage) can help to achieve proceduralised skill.

5. Discussions and conclusion

In this article, we have described a theory of how task knowledge is learned and forgotten that consists of three stages, as well as the implications of that theory. We believe our skill retention theory provides a more nuanced understanding of learning and forgetting that trainers can use to evaluate the efficacy of training systems (whether they support skill retention). We also believe that our work can provide useful guidance to training system designers and suggest further studies. Suggestions are noted through the paper, including to proceduralise skills where possible and to reduce the use of declarative memory items in procedural skills or that are needed to trigger procedural skills.

The skill retention theory suggests that skill retention is related to the progression through the three stages of learning. The different learning mechanisms (declarative or procedural) that characterise each stage exist in ACT-R; though, this is less true for forgetting. This study suggests that we may need to apply different strategies of knowledge acquisition (massed or distributed) to a training regimen with regard to both the skill type of interest and the learning stages. We summarise this theory and its implications in Table 1. The table notes the role of declarative and procedural knowledge at each stage of learning.

Unfortunately, we have not been able to find a study comparing different task skill components by their retention across the three stages. Such a study would be large, longitudinal, but greatly help develop the skill retention theory. If we can characterise the nature of decay of task skill components, it is possible to strengthen those component skills that are vulnerable to decay over an extended time. It will be necessary, therefore, to better understand the different attributes of skill components with regard to the three stages of learning and forgetting. This will be useful for better supporting the long-term retention of complex task skills.

Little has been studied on long-term retention and decay of procedural skills. We have seen a computational model based on ACT-R that simulates retention of declarative (vocabulary) memory (Pavlik and Anderson 2005) but not of procedural skill retention and its operational mechanism. Thus, it remains necessary to empirically study retention and decay of procedural skills and disambiguate procedural skill decay from declarative knowledge effects. In addition, decay mechanisms of procedural skills need to be proposed if they indeed exist. This endeavour will provide an advanced understanding of human learning and retention that can be applied to improve education and training systems.

Finally, we present questions that arise out of the skill retention theory. Can we tell which stage the learner is in? After a certain amount of predetermined overlearning, can we ensure that the learner moves into the third stage? The ACT-R theory provides some examples of the progression of learning through each stage, but no studies on procedural skill retention have been reported.

Theoretically, these transitions are clear in Figure 2. Practically, there are and will be difficulties in measuring transitions. We see at least two difficulties. One difficulty is knowing which stage the learner is in for a particular subtask; the second difficulty is that the complex tasks will have multiple subtasks, and managing this information and recognising that the learner may be at different stages in subtasks has been problematic for human and computer tutors in the past.

Acknowledgements

This research was sponsored by a grant from the Division of Human Performance Training, and Education at the Office of Naval Research, under Contract nos. W911QY-07-01-0004 and N000141010401. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Government or the Pennsylvania State University. Jonathan Morgan, Junya Morita and three anonymous reviewers provided useful comments and helped improve this manuscript.

References

Anderson, J.R., 1982. Acquisition of cognitive skill. *Psychological Review*, 89 (4), 369–406.

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

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

Anderson, J.R., Boyle, C.F., and Reiser, B.J., 1985. Intelligent tutoring systems. *Science*, 228 (4658), 456–462.

Anderson, J.R., Conrad, F.G., and Corbett, A.T., 1989. Skill acquisition and the Lisp tutor. *Cognitive Science*, 13, 467–707.

Anderson, J.R. and Lebiere, C., 1998. *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Anderson, J.R., Reder, L.M., and Lebiere, C., 1996. Working memory: activation limitations on retrieval. *Cognitive Psychology*, 30 (3), 221–256.

Anderson, J.R., *et al*., 2004. An integrated theory of mind. *Psychological Review*, 111 (4), 1036–1060.

Atkinson, R.C., 1972. Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, 96 (1), 124–129.

Bahrick, H.P., 1984. Semantic memory content in permastore: fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 113 (1), 1–29.

Bahrick, H.P., *et al*., 1993. Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4 (5), 316–321.

Bolton, M.L. and Bass, W.J., 2009. A method for the formal verification of human-interactive systems. *In*: *Proceedings of the human factors and ergonomics society 53rd annual meeting*, Santa Monica, CA, 764–768.

Booher, H.R. and Minninger, J., 2003. Human systems integration in Army systems acquisition. *In*: H.R. Booher, ed. *Handbook of human systems integration*. Hoboken, NJ: John Wiley, 663–698.

Card, S.K., English, W.K., and Burr, B.J., 1978. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics*, 21 (8), 601–613.

Chong, R.S., 2004. Architectural explorations for modeling procedural skill decay. *In*: M. Lovett, *et al.*, eds. *Proceedings of the sixth international conference on cognitive modeling* [online]. Mahwah, NJ: Lawrence Erlbaum.

Delaney, P.F., *et al.*, 1998. The strategy-specific nature of improvement: the power law applies by strategy within task. *Psychological Science*, 9 (1), 1–7.

Driskell, J.E., Willis, R.P., and Copper, C., 1992. Effect of overlearning on retention. *Journal of Applied Psychology*, 77 (5), 615–622.

Ericsson, K.A., Krampe, R.T., and Tesch-Römer, C., 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100 (3), 363–406.

Farr, M.J., 1987. *The long-term retention of knowledge and skills: a cognitive and instructional perspectives*. Arlington, VA: Springer.

Fitts, P.M., 1964. Perceptual-motor skill learning. *In*: A.W. Melton, ed. *Categories of human learning*. New York: Academic Press, 243–285.

Gray, W.D. and Boehm-Davis, D.A., 2000. Milliseconds matter: an introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied*, 6 (4), 322–335.

Jones, G., Ritter, F.E., and Wood, D.J., 2000. Using a cognitive architecture to examine what develops. *Psychological Science*, 11 (2), 93–100.

Kieras, D.E. and Meyer, D.E., 1997. An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12 (4), 391–438.

Kim, J.W., Koubek, R.J., and Ritter, F.E., 2007. Investigation of procedural skills degradation from different modalities. *In*: R.L. Lewis, T.A. Polk, and J.E. Laird, eds. *Proceedings of the 8th international conference on cognitive modeling*. Oxford, UK: Taylor & Francis/Psychology Press, 255–260.

Krueger, W.C.F., 1929. The effect of overlearning on retention. *Journal of Experimental Psychology*, 12, 71–78.

Laird, J.E., Rosenbloom, P.S., and Newell, A., 1987. SOAR: an architecture for general intelligence. *Artificial Intelligence*, 33, 1–64.

Landauer, T.K. and Bjork, R.A., 1978. Optimum rehearsal patterns and name learning. *In*: M.M. Gruneberg, P.E. Morris, and R.N. Sykes, eds. *Practical aspects of memory*. New York: Academic Press, 625–632.

Lee, F.J. and Anderson, J.R., 2001. Does learning a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology*, 42 (3), 267–316.

McKenna, S. and Glendon, A., 1985. Occupational first aid training: decay in cardiopulmonary resuscitation (CPR) skills. *Journal of Occupational Psychology*, 58, 109–117.

Neves, D.M. and Anderson, J.R., 1981. Knowledge compilation: mechanisms for the automatization of cognitive skills. *In*: J.R. Anderson, ed. *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum, 57–84.

Newell, A., 1973. You can't play 20 questions with nature and win: projective comments on the papers of this symposium. *In*: W.G. Chase, ed. *Visual information processing*. New York: Academic Press, 283–308.

Newell, A., 1990. *Unified theories of cognition*. Cambridge, MA: Harvard University.

Newell, A. and Rosenbloom, P.S., 1981. Mechanisms of skill acquisition and the law of practice. *In*: J.R. Anderson, ed. *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum, 1–55.

Pavlik, P.I. and Anderson, J.R., 2005. Practice and forgetting effects on vocabulary memory: an activation-based model of the spacing effect. *Cognitive Science*, 29, 559–586.

Pavlik, P.I. and Anderson, J.R., 2008. Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, 14 (2), 101–117.

Pew, R.W., 2007. Some history of human performance. *In*: W.D. Gray, ed. *Integrated models of cognitive systems*. New York, NY: Oxford University Press, 29–44.

Pew, R.W. and Mavor, A.S., eds., 2007. *Human-system integration in the system development process: a new look*. Washington, DC: National Academy Press.

Posner, M.I., 1973. *Cognition: an introduction*. Glenview, IL: Scott Foresman.

Proctor, R.W. and Dutta, A., 1995. *Skill acquisition and human performance*. Thousand Oaks, CA: SAGE.

Ramos, M.A.G. and Chen, J.J.G., 1994. On the integration of learning and forgetting curves for the control of knowledge and skill acquisition for non-repetitive task training and retraining. *International Journal of Industrial Engineering*, 1 (3), 233–242.

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

Ritter, F.E. and Schooler, L.J., 2001. The learning curve. *In*: W. Kintch, N. Smelser, and P. Baltes, eds. *International encyclopedia of the social and behavioral sciences*. Vol. 13, Amsterdam: Pergamon, 8602–8605.

Rohrer, D. and Pashler, H., 2007. Increasing retention without increasing study time. *Current Directions in Psychological Science*, 16 (4), 183–186.

Rohrer, D. and Taylor, K., 2006. The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, 20, 1209–1224.

Rohrer, D., *et al.*, 2005. The effect of overlearning on long-term retention. *Applied Cognitive Psychology*, 19, 361–374.

Rosenbloom, P.S. and Newell, A., 1987. Learning by chunking: a production system model of practice. *In*: D. Klahr, P. Langley, and R. Neches, eds. *Production system models of learning and development*. Cambridge, MA: MIT Press, 221–228.

Schendel, J.D. and Hagman, J.D., 1982. On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology*, 67 (5), 605–610.

Schmidt, R.A. and Bjork, R.A., 1992. New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3 (4), 207–217.

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

Shea, J.B. and Morgan, R.L., 1979. Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, 5 (2), 179–187.

St. Amant, R., Horton, T.E., and Ritter, F.E., 2007. Model-based evaluation of expert cell phone menu interaction. *ACM Transactions on Computer-Human Interaction*, 14 (1), 1–24.

Taatgen, N.A. and Lee, F.J., 2003. Production compilation: a simple mechanism to model complex skill acquisition. *Human Factors*, 45 (1), 61–76.

VanLehn, K., 1996. Cognitive skill acquisition. *Annual Review of Psychology*, 47, 513–539.

Vearrier, L.A., *et al.*, 2005. An intensive massed practice approach to retraining balance post-stroke. *Gait and Posture*, 22 (2), 154–163.

About the authors

Jong Wook Kim is a Research Scientist in the Department of Psychology at the University of Central Florida. He has conducted several funded research for the Navy and Army in the areas of Applied Cognitive Science. His research efforts have been focused on skill learning and retention. His current interests include the development of intelligent tutoring systems for enhanced skill retention. He received his PhD in Industrial and Manufacturing Engineering from the Pennsylvania State University.

Frank E. Ritter is a Professor of Information Sciences and Technology, of Psychology, of Computer Science and Engineering, and is affiliated with Industrial Engineering at the Pennsylvania State University. His research is in the development, application, and methodology of cognitive models, particularly as applied to interfaces and behavioural moderators. He helped found the new interdisciplinary College of Information Sciences and Technology at Penn State. Previously he has worked at BBN Labs, where he helped build a tutor for ONR, and taught at the University of Nottingham. He received his PhD in AI and Psychology from Carnegie-Mellon University.

Richard J. Koubek is Dean of College of Engineering at Louisiana State University. He was Professor and Head for the Industrial and Manufacturing Engineering at the Pennsylvania State University. His research focuses on usability, human aspects of manufacturing, and human-computer interaction. He received his PhD in Industrial Engineering from Purdue University.