# An ACT-R Model of the Spacing Effect

**Philip I Pavlik (ppavlik@andrew.cmu.edu)**
Department of Psychology,
Carnegie Mellon University
Pittsburgh, PA 15213 USA

**John R. Anderson (ja+@cmu.edu)**
Department of Psychology,
Carnegie Mellon University
Pittsburgh, PA 15213 USA

## Abstract

The ACT-R computational modeling system encapsulates an activation-based system of declarative memory. While this model has had a variety of successes in modeling memory phenomena, in its current conception it has no mechanism that would produce the spacing effect. A paired-associate memory experiment was conducted and the data from this experiment were fit with an ACT-R memory model using a new decay mechanism. Rather than a set decay rate, this mechanism bases decay for a trial on the current activation at the time of that trial. The spacing effect is a result of this mechanism.

## Introduction

The spacing effect is one of the most ubiquitous phenomena of human memory. Often referred to the as the massed vs. distributed practice effect, the spacing effect refers to the memory benefit that individuals accrue when they increase the duration between practice episodes. It is common to hear the admonishment given to students that they should not "cram" the night before the exam if they really want to learn the material well. They are being told to use the spacing effect to maximize their learning. It is just this sort of maximization that the model we propose attempts to formalize.

As a vehicle for studying memory and investigating the spacing effect, we have chosen a paired-associate memory task in which subjects memorize the English translations of Japanese words. Foreign language vocabulary learning retains a nearness to basic processes of memory while having a certain amount of external validity. The pioneering work of Bahrick (1979) suggests that foreign language vocabulary learning is a task for which the spacing effect should have important implications.

## Experiment

An experiment was conducted to look for evidence of the spacing effect. The intent of this experiment was to provide a strong challenge to any proposed model of spacing.

### Method

#### Participants and Design

In this experiment, participants were tested repeatedly over the course of two sessions on their knowledge of Japanese-English vocabulary pairs. There was either 1 or 7 days between the first and second testing sessions (S1 and S2). In both cases during S1, participants were tested on word pairs that were repeated 1, 2, 4, or 8 times spaced at intervals of 2, 14 or 98 trials. This indicates a 3x4 design; however, due to session length limitations the 8 x 98 condition was not included, resulting in 11 conditions. There were 8 word pairs in each condition, and 16 word pairs used for primacy and warm-up buffers as well as to enable the arrangement of the spacing conditions. An attempt was made to space the trials of conditions across the span of S1. Thus, learning trials could occur at any time since new items were introduced continuously. During S2, participants were tested 4 times with each pair at a spacing of 98 trials between tests.

40 participants were recruited for this study from the Pittsburgh, Pennsylvania community. All participants completed the experiment. 20 subjects were used in each condition. Sessions lasted between 60 and 90 minutes. Only subjects that professed no knowledge of Japanese were recruited.

#### Materials

The stimuli were a collection of 104 Japanese-English word pairs. English words were chosen according to certain criteria. Words had mean familiarity ratings of 548 and mean imagability ratings of 464 in the MRC Psycholinguistic Database (Coltheart, 1981). The data base mean of familiarity and imagability are 488 (s.d. 120) and 438 (s.d. 99) respectively. Japanese translations (from the possible Japanese synonyms) were chosen for dissimilarity to common English words. Only 4 letter English words were used, and 4 to 7 letter Japanese translations were used. Assignment of words to conditions was randomized for each participant.

#### Procedure

Participants arrived at the testing room, signed consent forms, and were told that this was a memory study. Details of the parameters of the testing were given and it was explained that the experimenter was performing this study with the aim of developing measures to predict longer-term retention by measuring retrieval speeds, correctness, and amounts and times of practice.

Participants were instructed not to practice the word pairs during the interval between sessions. Participants were scored for motivational purposes, receiving 6 points for each correct response and losing 12 points for each incorrect response. Failure to provide a response, either by time-out or providing a blank response received 0 score. Payment, $9 to $15 per session for each of 2 sessions, was based on this scoring.

Each pair received one study trial and then a number of subsequent test trials. Study and test trials were intermingled in 12 blocks of 40 trials. Initial study trials occurred at the appropriate prior spacing for the condition of that pair. All trials began with presentation of the word “Study” or “Test” presented for 2 seconds to cue the purpose of the following trial. Study opportunities allowed subjects to view the new pair for 5 seconds. Tests involved presentation of the Japanese word on the left side of the screen. Participants typed the correct English translation on the right. If no response was made, the program timed-out in 7 seconds and registered the lack of response as incorrect.

Following response or failure to respond the program responded “Correct” or “Incorrect” for one second while providing the score adjustment based on the performance. If the response was correct, the following trial commenced. If incorrect, the word “Restudy” appeared for 2 seconds followed by a 5 second restudy opportunity that was identical to the original study presentation. Since model results are based on time averages rather than counts of intervening trials, these extra studies are averaged into the model.

Between all blocks participants proceeded by pressing the space bar when they were ready. Few subjects paused at these opportunities. S2 procedures were identical with the exception that there were no study trials.

## Results and Discussion

Session one data were aggregated across different practice conditions. Figure 1 displays the data for S1. As can be seen, performance improves across the first 4 trials of S1, F(3,114)=412, p<.05, and there was significantly lower performance with wider spacing, F(2,76)=240, P<.05. This lower performance on S1 with wider spacing is likely due to the overall longer retention intervals for these trials. This analysis also confirmed that there were no differences in learning between the two groups that would be experiencing the 1 and 7 day delays, F(1,38)=1.13, p=.296.

In order to establish that forgetting occurred between sessions, last repetitions on S1 and first repetitions on S2 were compared for differences in correctness. Not surprisingly, subjects forgot over the intervals, F(1,38) = 770, p < .05, and the participants forgot more over 7 days as compared to 1, F(1,38) = 5.14, p < .05.

The mean correctness for S2 was .64 for the 1 day retention interval and .52 for the 7 day retention interval. Since the patterns of data were similar between the 2 retention intervals on S2 (a correlation of .959), we aggregated the two conditions for purposes of display. These data are displayed in Figure 2, and there are a number of two-way and three-way interactions that can be noted from the graphs. Repeated -measures ANOVA of S2 data (S1 Repetitions x S1 Spacing x S2 Trial x Retention interval, excluding the 8 repetition conditions due to the incomplete design) revealed strong main effects of spacing, F(2,76) = 58.2, p<.05. Most interestingly, there was a significant repetition x spacing interaction F(4,152) = 4.38, p < .05 reflecting the fact that benefit to spacing increased with increased repetition at a particular spacing. As can be seen from the graphs, there is a rather dramatic increase in the importance of spacing as repetitions increased.

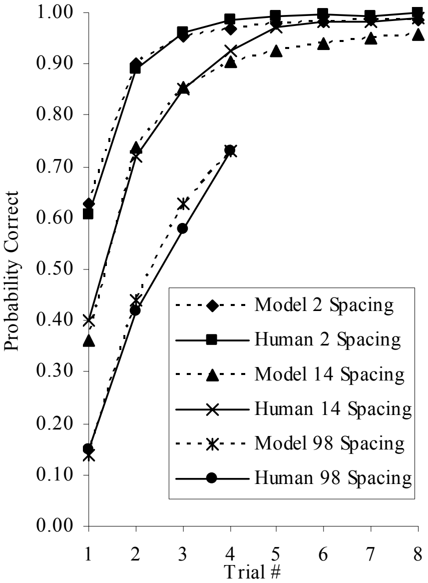


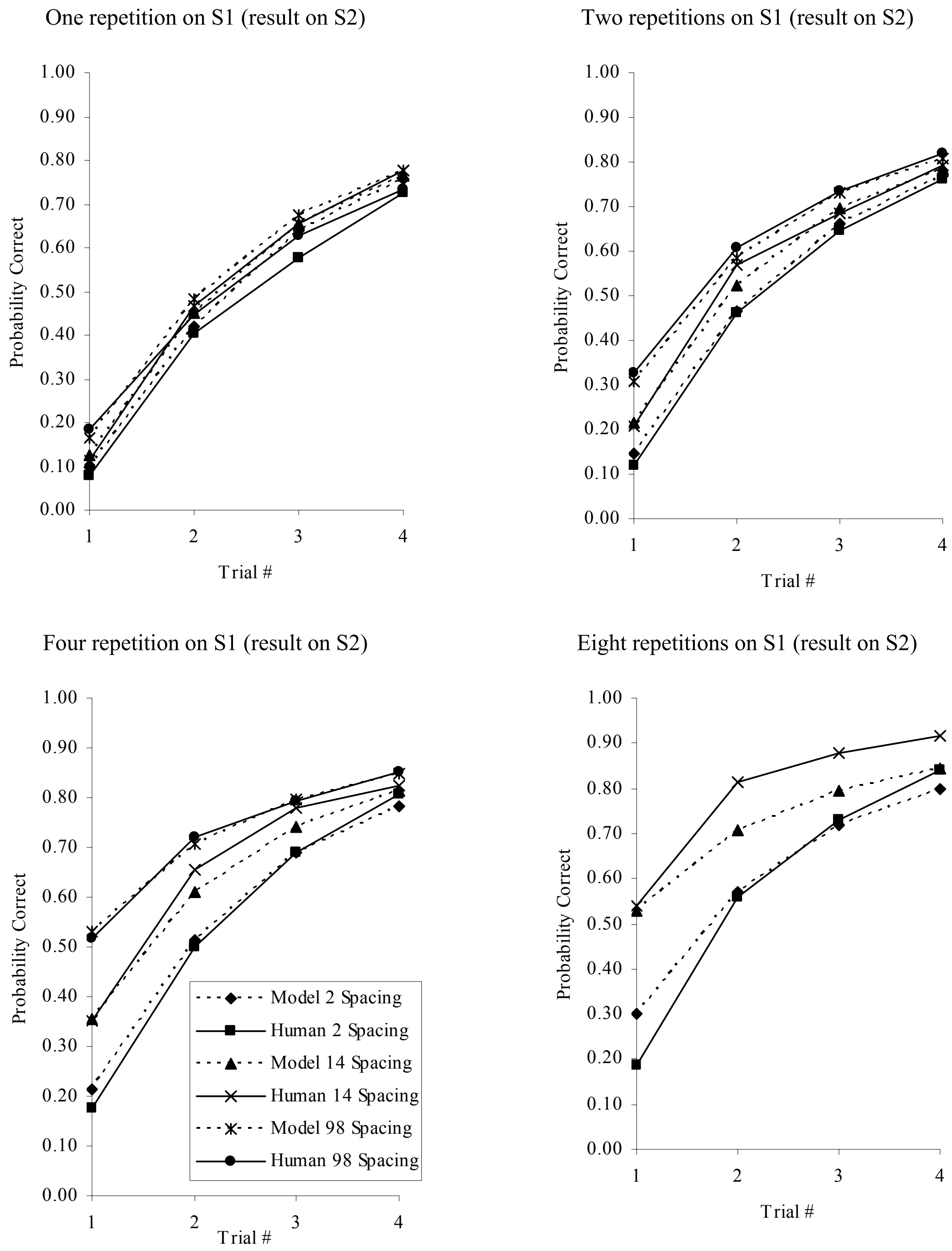
Figure 1: S1 aggregate data for humans and model for spacing conditions.

## Model

ACT-R’s retrieval memory system is based on a unitary trace that is composed of the sum of a number of individual strengthenings. The equation for this summation (Equation 1) proposes that these strengthenings accumulate and decay according to a power function.

$$m_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) \quad (1)$$

In this function $m_i$ is the activation of item i, $t_j$ is how long ago the practice of that item occurred, and n is the number of opportunities (trials) to practice this item.


One repetition on S1 (result on S2)

Two repetitions on S1 (result on S2)

Four repetition on S1 (result on S2)

Eight repetitions on S1 (result on S2)

Figure 2: S2 aggregate data for humans and model for practice conditions by spacing intervals.

The logarithm of the sum is taken to yield observed retention functions and provides a correspondence with log odds of items occurring in the environment as shown by Anderson and Schooler (1991).

The choice of the power function for decay is not arbitrary, but we believe there may be some question as to the functional form of decay. We have found that a power function for activation strength provides activation values that fit well to the data for latencies and correctness. Others investigating practice and retention functions have questioned whether a power function is satisfactory and suggested that exponential functions may be more appropriate (Heathcote, Brown, & Mewhort, 2000; Myung, Kim, & Pitt, 2000). Still others however, have demonstrated that the power function can be appropriate (Newell and Rosenbloom, 1981; Wixted & Ebbesen, 1997). We currently favor the power-law decay function because it corresponds with analyses of how need probability for memory fluctuates in the environment (Anderson and Schooler., 1991). We have also found it to be computationally tractable and parsimonious relative to many options that would entail more parameters.

In ACT-R, activation plays a role in the chunk choice equation (Equation 2). This equation describes the probability of one (of many) alternatives being greatest. In the case of the simplified version below, the correct choice competes solely with the threshold value, rather than with other similar items in memory as in more complex versions of the equation.

$$recall\ probability_i = \frac{e^{\frac{m_i}{s}}}{e^{\frac{m_i}{s}} + e^{\frac{\tau}{s}}} \tag{2}$$

In this equation, $\tau$ is the threshold parameter and s is a measure of noise. As can be seen from an inspection of the formula, as $m_i$ tends higher, the probability of recall approaches 1, whereas, as $\tau$ tends higher, the probability decreases. In fact, when $\tau=m_i$, the probability of recall is .5. The s parameter controls the noise in activation and it describes the sharpness of the difference in recall resulting from activation values above or below threshold. If s is near 0, the transition from near 0% recall to near 100% will be quite abrupt, whereas when s is larger, the transition as activation increases will be a smooth sigmoidal curve.

A further mechanism of the model in question is its ability to handle forgetting over the interval between experimental sessions. Anderson, Fincham and Douglass (1999) found that a simple decay rate did not well fit the data. Instead, they found it necessary to postulate a construct they refer to as psychological time. This refers to the fact that forgetting seems to proceed more slowly in the interval between sessions. A parameter (h) reflects how fast between session intervening events “chip away” at the memory relative to experimental events. It should be noted that the psychological time factor in this paper takes a slightly different form as compared to Anderson, Fincham, and Douglass (1999). In the current conception, we take the h factor to be a direct scaling parameter of the time between experimental sessions. This is implemented by multiplying the amount of time between sessions by the h parameter. Thus if the h parameter is .03, the 23 hours between two sessions is multiplied by .03 for the purposes of calculating activations on session 2.

The new part of our model is the mechanism that accounts for spacing. Anderson and Schooler (1991) first accounted for the spacing effect using an equation that scaled the lag in time from a previous trial in order to calculate a decay rate for that particular trial. Thus, if the lag from the previous trial was great, decay would be low, while if the lag was short, decay would be high. This formula did indeed go a long way to account for spacing effects and fits many data sets well. However, it makes the decay rate only a function of the delay from the last trial. It seemed more reasonable to make decay a function of the current level of activation of the trace, which would reflect the overall mass of practice. We found this idea essential to account for the very little learning we observed in other experiments under conditions of very massed practice.

As a modification of the Anderson & Schooler (1991) proposal, we have developed an equation in which decay for each $t_j$ is a function of the activation at the time it occurs instead of the lag. (See Equation 3.) The implication of this is that higher activation at the time of a trial will result in that trial decaying more quickly. Alternatively, if activation were low, decay would proceed more slowly. Since $e^{-m}$ is the basis of the retrieval time function in ACT-R, Equation 3 can also be interpreted as explaining decay as an inverse function of time effort expended in making a recall. In Equation 3, c is a scaling parameter, and a is the intercept of the decay function. Equation 4 represents how the new decay term integrates into the original activation function where $d_j$ is the decay of item j calculated from the activation strength at its occurrence.

$$d_j = ce^{m_j} + a \tag{3}$$

$$m_i = \ln\left(\sum_{j=1}^{n} t_j^{-d_j}\right) \tag{4}$$

It should be pointed out that our model assumes that each trial counts as one encoding (one time value for the purpose of activation calculations) regardless of whether it was correct or incorrect. In the case of a correct response, we consider it to be one encoding as a result of the retrieval of the memory, in the case of an incorrect response we consider it as one encoding because of the subsequent study. We also consider the initial study trial for each pair to count as one encoding. More complex assumptions could be considered; however, this assumption resulted in a good fit to the data and we favored it for the sake of parsimony. Another assumption we have made is to consider the unit of memorization to be a chunk encoding the association of the Japanese word with the English word.

There were 162 aggregate correctness data point averages to be fit for the experiment, split into 81 points for each between-subjects group, of which 37 points were for the S1 conditions and 44 points were for the S2 conditions. The model was fit to these 162 points through a $\chi^2$ minimization determined from aggregate condition variances ($\chi^2$=260, d.f. 157). Table 1 displays the model parameters.

Table 1: Model parameter values.

| | |
|---|---|
| decay intercept (a) | 0.167 |
| scale decay (c) | 0.232 |
| noise (s) | 0.252 |
| threshold ($\tau$) | -0.669 |
| intervening events (h) | .031 |

Measures of goodness-of-fit show that the model mirrored the absolute and relative patterns in the data closely. (See Figures 1 and 2.) The root mean squared deviation (RMSD adjusted for 5 parameters fit) was .047 indicating that the mean deviation of model points from data was 4.7% in absolute correctness across the 162 points. The $r^2$ value of .971 indicates the model also well captured the pattern of the data across all 162 points.

Of concern in the fits is the serious deviation for the 8 repetition condition with 14 spacing on S2. Here we see that the model is unable to capture the large benefit of spacing the subjects experienced. We are unable to account for this discrepancy, except to suggest that perhaps it is related to the fact that this condition of the experiment involved subjects reaching a sustained and high level of activation for a longer period than any other condition. We believe it is plausible to entertain the notion that some process categorically different than simple declarative memory retrieval may be partially involved in this superior performance by participants.

## A Second Test

In order to gain some sense of the generalizability of the model we decided to test it against a past result in the literature on spacing. Ideally, such a test would involve conditions that were dissimilar to those of the experiment from which the model was designed. Glenberg (1976) experiment 3 fit this criterion. This experiment involved a shorter time course, less semantically meaningful stimuli, and a different task (recognition vs. paired-associate recall). In this experiment subjects proceeded through decks of cards containing consonant trigrams. For each card they had to rate whether it had been seen previously or was new. Some cards were presented three times with an initial spacing lag between first and second presentations of 0, 1, 8, 20, or 40, and a retention interval of either 8, 32, or 64 trials between second and third trials. The dependent measure of interest was the proportion judged old on the third trial. These data are displayed in Figure 3, with the corresponding fit of our model.

The fit of the model was quite good with an $r^2$ of .980 and an RMSD of 1.0%, adjusted for the three parameters fit. In this case, we choose to consider threshold as the stable construct. It remained at the default from the experiment of .669. The parameters fitted were a=.104, c=.374, and s=.4.

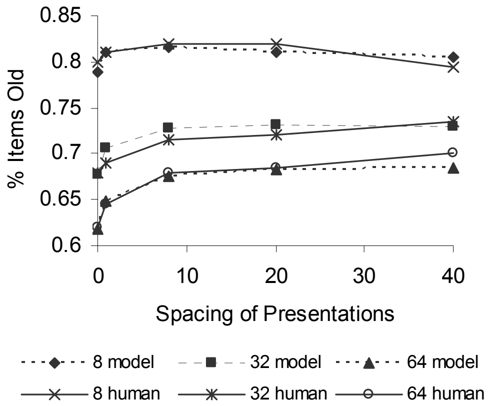


Figure 3: Human and Model results Glenberg (1976). Lines represent different retention intervals.

## Discussion

The most novel aspect of this ACT-R memory model, the calculation of decay as a function of activation, can be supported from several theoretical perspectives. Since ACT-R has traditionally considered decay to be some sort of neural degradation of existing memories, it makes sense to frame the theory behind this mechanism in terms of a plausible neural explanation.

The mechanism in our model purports that the encoding resulting from a study decays more quickly when activation is high. One neural theory of memory is that long-term memory encoding occurs through synaptic changes induced by LTP (long-term potentiation) (Scharf, Woo, Lattal, Young, Nguyen & Abel, 2002). Our mechanism suggests that when LTP is already high, another presentation will provide a temporary increase in activation, but do little to increase the rate of LTP encoding into long-term memory. Thus, there will be relatively rapid decay of this presentation. In fact, (Beggs, 2001) presents a statistical model of LTP information transfer to synapse connection strengths that proposes a limiting mechanism similar to ours and Scharf et al. (2002) show that temporally spaced learning or synaptic stimulation does indeed induce enhanced LTP as compared to massed learning or stimulation. This explanation appeals to the notion that any biological system is limited, and it makes sense to suggest that as activation grows that limit is neared.

Our mechanism is further quite plausible in light of theories of spacing and memory that propose that the benefit of additional practice is mediated by the difficulty of that additional practice. The proposition that difficulty of access at learning may increase recall at test has been advocated by Whitten and Bjork (1977). Whitten, Bjork and Schmidt (1977; 1992) note that often manipulations that cause detrimental effects to acquisition result in better long-term retention. This paradoxical result is quite explicable in terms of our mechanism if we suppose that the current activation of an item determines decay. Any manipulation that depresses that current activation will result in higher long-term activation because of the effect on decay rate. This can be seen as a theoretical advantage of our theory over theories that imply spacing should depend on differential encoding. Theories of differential encoding (Bjork & Allen, 1970) imply that the full effects of spacing should be apparent on a retention trial immediately following a pair of trials at different spacing. Our experimental data suggests this is not the case and that these benefits manifest most strongly at long intervals (Schmidt & Bjork, 1992).

Our conception also corresponds well with recent work by Altman and Gray (2002) involving a functional relationship between decay and interference. This work proposes the decay of current items in memory serves the purpose of reducing interference when an individual needs to switch current items frequently. In this work they present evidence that performance decline is slower when updates are presented less frequently. This evidence corresponds well with the mechanism we are proposing, and highlights the functional nature of forgetting as an adaptive response to the environment.

The fact that the spacing of trials provided a significant and meaningful benefit to later recall is not surprising given the fact that spacing has been conclusively shown to benefit foreign language instruction (Bloom & Shuell, 1981; Bahrick, 1979; Bahrick, Bahrick, Bahrick, & Bahrick, 1993). However, the success of the model in capturing the complexity of the data we generated indicates that the mechanism we propose is a strong candidate explanation for the spacing effect. The most important contribution of this work was the explication of that mechanism

## Acknowledgments

Preparation of this paper was supported by grant BCS 997-5-220 from the National Science Foundation.

## References

Altmann, E. M. & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. Psychological Science, 13, 27-33.

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

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

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. Journal of Experimental Psychology: General, 108, 296-308.

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, 316-321.

Beggs, J. M. (2001). A statistical theory of long-term potentiation and depression. Neural Computation, 13, 87-111.

Bjork, R. A. & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding. Journal of Verbal Learning & Verbal Behavior, Vol. 9, 567-572.

Bloom, K. C. & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. Journal of Educational Research, 74, 245-248.

Coltheart, M. (1981). The MRC psycholinguistic database. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33, 497-505.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. Journal of Verbal Learning & Verbal Behavior, 15, 1-16.

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, 185-207.

Myung, I. J., Kim, C., & Pitt, M. A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. Memory & Cognition, 28, 832-840.

Newell, A., & Rosenbloom, P. S., 1981, Mechanisms of skill acquisition and the law of practice. In Cognitive skills and their acquisition, edited by J. R. Anderson, Erlbaum (Hillsdale).

Scharf, M. T., Woo, N. H., Lattal, K. M., Young, J. Z., Nguyen, P. V., & Abel, T. (2002). Protein synthesis is required for the enhancement of long-term potentiation and long-term memory by spaced training. Journal of Neurophysiology, 87, 2770-2777.

Schmidt, R. A. & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, 3, 207-217.

Whitten, W. B. & Bjork, R. A. (1977). Learning from tests: Effects of spacing. Journal of Verbal Learning & Verbal Behavior, 16, 465-478.

Wixted, J. T. & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. Memory & Cognition, 25, 731-739.