Optimizing Memory Retention with Cognitive Models

Robert Lindsey (robert.lindsey@colorado.edu)
Michael Mozer (mozer@colorado.edu)
Department of Computer Science and Institute of Cognitive Science
University of Colorado at Boulder

Nicholas J. Cepeda (ncepeda@yorku.ca)
Department of Psychology, York University

Harold Pashler (hpashler@ucsd.edu)
Department of Psychology, University of California at San Diego

Abstract

When individuals learn facts (e.g., foreign language vocabulary) over multiple sessions, the durability of learning is strongly influenced by the temporal distribution of study (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Computational models have been developed to explain this phenomenon known as the *distributed practice effect*. These models predict the accuracy of recall following a particular study schedule and retention interval. To the degree that the models embody mechanisms of human memory, they can also be used to determine the spacing of study that maximizes retention. We examine two memory models (Pavlik & Anderson, 2005; Mozer, Pashler, Lindsey, & Vul, submitted) that provide differing explanations of the distributed practice effect. Although both models fit experimental data, we show that they make robust and opposing predictions concerning the optimal spacing of study sessions. The Pavlik and Anderson model robustly predicts that contracting spacing is best over a range of model parameters and retention intervals; that is, with three study sessions, the model suggests that the lag between sessions one and two should be larger than the lag between sessions two and three. In contrast, the Mozer et al. model predicts equal or expanding spacing is best for most material and retention intervals. The limited experimental data pertinent to this disagreement appear to be consistent with the latter prediction. The strong contrast between the models calls for further empirical work to evaluate their opposing predictions.

Keywords: distributed practice effect; optimization; study schedules

Introduction

In educational settings, individuals are often required to memorize facts such as foreign language vocabulary words. A question of great practical interest is how to retain knowledge once acquired. Psychologists have identified factors influencing the durability of learning, most notably the temporal distribution of practice: when individuals study material across multiple sessions, long-term retention generally improves when the sessions are spaced in time. This effect, known as the *distributed practice* or *spacing* effect, is typically studied via an experimental paradigm in which participants are asked to study items over two or more sessions, and the time between sessions—the *interstudy interval* or *ISI*—is varied. Retention is often evaluated via a cued recall test at a fixed lag following the final study sessionthe *retention interval* or *RI* (Figure 1).

Typical experimental results are shown in the data points and dotted lines of Figures 2a (Glenberg, 1976) and 2b (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). In both experiments, participants studied material at two points in time, with a variable ISI, and then were tested following a fixed RI. The graphs show recall accuracy at test as a function of ISI for several different RIs. The curves, which we will refer to as *spacing functions*, typically show a rapid rise in memory retention as ISI increases, reach a peak, and then gradually drop off. From the spacing function, one can determine the *optimal ISI*, the spacing of study that yields maximal retention. The exact form of the spacing function depends on the specific material to be learned and the RI. The distributed practice effect is obtained over a wide range of time scales: ISIs and RIs in the Glenberg study are on the order of seconds to minutes, and in the Cepeda et al. study are on the order of weeks to months. On the educationally relevant time scale of months, optimally spaced study can double retention over massed study. Thus, determining the optimal spacing of study can have a tremendous practical impact on human learning.

Pavlik and Anderson (2005; 2008) used the ACT-R declarative memory equations to explain distributed practice effects. ACT-R supposes a separate trace is laid down for each study and that the trace decays according to a power function of time. The key feature of the model that yields the distributed practice effect is that the decay rate of a new trace depends on an item's current memory strength at the point in time when the item is studied. This ACT-R model has been fit successfully to numerous experimental datasets. The solid lines of Figure 2a show the ACT-R fit to the Glenberg data.

Mozer, Pashler, Lindsey, and Vul (submitted) have recently proposed a model providing an alternative explanation of the distributed practice effect. In this model, when an item is studied, a memory trace is formed that includes the current *psychological context*, which is assumed to vary randomly over time. Probability of later recall depends in part on the similarity between the context representations at study and test. The key feature of this model that distinguishes it from related past models (e.g., Raaijmakers, 2003) is that the context is assumed to wander on multiple time scales. This

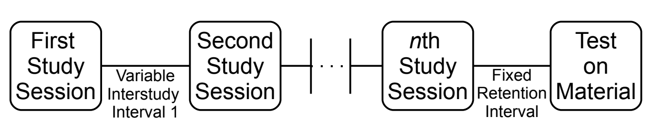

Figure 1: Structure of a study schedule.

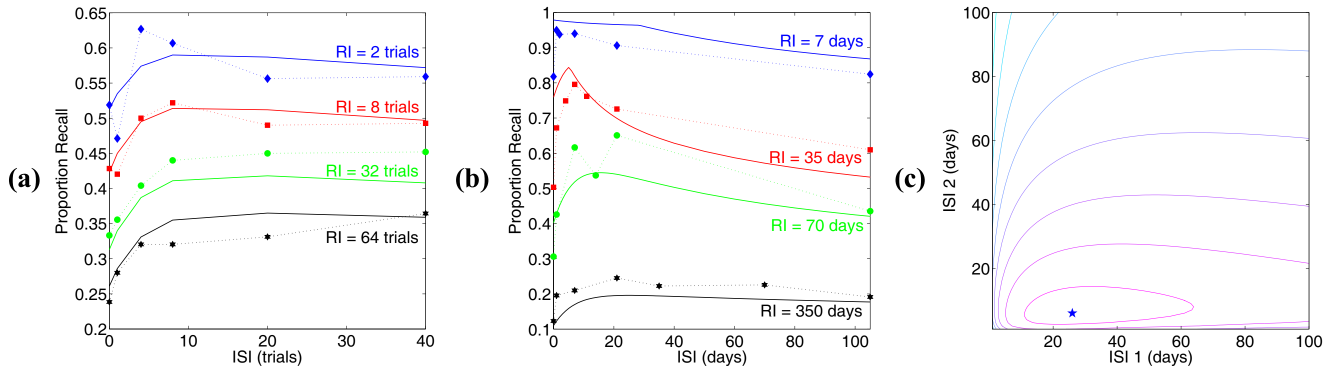

Figure 2: Results from **(a)** Glenberg (1976) and **(b)** Cepeda et al. (2008) illustrative of the distributed practice effect. The dotted lines correspond to experimental data. The solid lines in **(a)** and **(b)** are the ACT-R and MCM fits to the respective data. **(c)** A contour plot of recall probability as a function of two ISIs from ACT-R with parameterization in Pavlik and Anderson (2008).

model, referred to as the *multiscale context model* (MCM), has also been successfully fit to numerous empirical datasets, including the Glenberg study. In Figure 2b, we show the MCM prediction (solid lines) of the Cepeda et al. data.

Both ACT-R and MCM can be parameterized to fit data post hoc. However, both models have been used in a predictive capacity. Pavlik and Anderson (2008) have used ACT-R to determine the order and nature of study of a set of items, and showed that ACT-R schedules improved retention over alternative schedules. Mozer et al. (submitted) parameterize MCM with the basic forgetting function for a set of items (the function relating recall probability to RI following a single study session) and then predict the spacing function for the case of multiple study sessions. Figure 2b is an example of such a (parameter free) prediction of MCM.

Most experimental work involves two study sessions, the minimum number required to examine the distributed-practice effect. Consequently, models have mostly focused on this simple case. However, naturalistic learning situations typically offer more than two opportunities to study material. The models can also predict retention following three or more sessions. In this paper, we explore predictions of ACT-R and MCM in order to guide the design of future experiments that might discriminate between the models.

Study Schedule Optimization

A cognitive model of the distributed practice effect allows us to predict recall accuracy at test for a particular study schedule and RI. For example, Figure 2c shows ACT-R's prediction of recall probability for a study schedule with two variable ISIs and an RI of 20 days, for a particular parameterization of the model based on Pavlik and Anderson (2008). It is the two-dimensional generalization of the kind of spacing functions illustrated in Figures 2a and 2b. Recall probability, shown by the contour lines, is a function of both ISIs. The star in Figure 2c indicates the schedule that maximizes recall accuracy.

Models are particularly important for study-schedule optimization. It is impractical to determine optimal study schedules empirically because the optimal schedule is likely to depend on the particular materials being learned and also because the combinatorics of scheduling $n+1$ study sessions (i.e., determining n ISIs) make it all but impossible to explore experimentally for $n > 1$. With models of the distributed practice effect, we can substitute computer simulation for exhaustive human experimentation.

In real-world learning scenarios, we generally do not know exactly when studied material will be needed; rather, we have a general notion of a span of time over which the material should be retained. Though not the focus of this paper, models of the distributed practice effect can be used to determine study schedules that maximize retention not only for a particular prespecified RI, but also for the situation in which the RI is treated as a random variable with known distribution. The method used in this paper to determine optimal study schedules can easily be extended to accomodate uncertain RIs.

Pavlik and Anderson ACT-R Model

In this section, we delve into more details of the Pavlik and Anderson (2005; 2008) model, which is based on ACT-R declarative memory assumptions. In ACT-R, a separate trace is laid down each time an item is studied, and the trace decays according to a power law, t^{-d}, where t is the age of the memory and d is the power law decay for that trace. Following n study episodes, the activation for an item, m_n, combines the trace strengths of individual study episodes:

$$m_n = \beta_s + \beta_i + \beta_{si} + \ln\left(\sum_{k=1}^{n} b_k t_k^{-d_k}\right),$$

where t_k and d_k refer to the age (in seconds) and decay associated with trace k, and the additive parameters β_s, β_i, and β_{si} correspond to participant, item, and participant-item factors that influence memory strength, respectively. The variable b_k reflects the salience of the kth study session (Pavlik, 2007); larger values of b_k correspond to cases where, for example, the participant self-tested and therefore exerted more effort.

The key claim of the ACT-R model with respect to the distributed-practice effect is that the decay term on study trial k depends on the item's overall activation at the point when study occurs, according to the expression:

$$d_k(m_{k-1}) = ce^{m_{k-1}} + \alpha,$$

where c and α are constants. If spacing between study trials is brief, the activation m_{k-1} is large and consequently the new study trial will have a rapid decay, d_k. Increasing spacing can therefore slow memory decay of trace k, but it also incurs a cost in that traces $1...k-1$ will have substantial decay.

The model's recall probability is related to activation by:

$$p(m) = 1/(1+e^{\frac{\tau-m}{s}}),$$

where τ and s are additional parameters. The pieces of the ACT-R model relevant to this paper include 3 additional parameters, for a total of 10 parameters, including: h, a translation of real-world time to internal model time, u, a descriptor of the maximum benefit of study, and v, a descriptor of the rate of approach to the maximum.

Pavlik and Anderson (2008) use ACT-R activation predictions in a heuristic algorithm for scheduling the trial order *within* a study session, as well as the trial type (i.e., whether an item is merely studied, or whether it is first tested and then studied). They assume a fixed intersession spacing. Thus, their algorithm reduces to determining how to best allocate a finite amount of time within a session.

Although they show a clear effect of the algorithm used for within-session scheduling, we focus on the complementary issue of scheduling the lag between sessions. The ISI manipulation is more in keeping with the traditional conceptualization of the distributed-practice effect. Fortunately, the ACT-R model can be used for both within- and between-session scheduling. To model between-session scheduling, we assume—as is true in controlled experimental studies—that each item to be learned is allotted the same amount of study (or test followed by study) time within a session.

Pavlik and Anderson (2008) describe their within-session scheduling algorithm as optimizing performance, yet we question whether their algorithm is appropriately cast in terms of optimization. They argue that maximizing probability of recall should not be the goal of a scheduling algorithm, but that activation gain at test should be maximized so as to encourage additional benefits (e.g., improved long-term retention). We believe that had Pavlik and Anderson (2008) sought simply to maximize probability of recall at test and had more rigorously defined their optimization problem, they would have seen results of the ACT-R within-session scheduler even better than what they achieved. In light of these facts, we contend that our work is the first effort to truly optimize memory retention via cognitive models.

Multiscale Context Model

One class of theories proposed to explain the distributed-practice effect focuses on the notion of encoding variability. According to these theories, when an item is studied, a memory trace is formed that incorporates the current psychological context. Psychological context includes conditions of study, internal state of the learner, and recent experiences of the learner. Retrieval of a stored item depends partly on the similarity of the contexts at the study and test. If psychological context is assumed to fluctuate randomly, two study sessions close together in time will have similar contexts. Consequently, at the time of a recall test, either both study contexts will match the test context or neither will. A longer ISI can thus prove advantageous because the test context will have higher likelihood of matching one study context or the other.

Raaijmakers (2003) developed an encoding variability theory by incorporating time-varying contextual drift into the Search of Associative Memory (SAM) model and used this model to explain data from the distributed-practice literature. The context consists of a pool of binary-valued neurons which flip state at a common fixed rate. This behavior results in exponentially decreasing similarity between contexts at study and test time as a function of the study-test lag.

In further explorations, we (Mozer et al., submitted) found a serious limitation of SAM: Distributed-practice effects occur on many time scales (Cepeda et al., 2006). SAM can explain effects for study sessions separated by minutes or hours, but not for sessions separated by weeks or months. The reason is essentially that the exponential decay in context similarity bounds the time scale at which the model operates.

To address this issue, we proposed a model with multiple pools of context neurons. The pools vary in their relative size and the rate at which their neurons flip state. With an appropriate selection of the pool parameters, we obtain a model that has a power-law forgetting function and is therefore well suited for handling multiple time scales. The notion of multiscale representations comes from another model of distributed-practice effects developed by Staddon, Chelaru, and Higa (2002) to explain rat habituation. We call our model, which integrates features of SAM and Staddon et al.'s model, the Multiscale Context Model (MCM).

MCM has only five free parameters. Four of these parameters configure the pools of context neurons, and these parameters can be fully constrained for a set of materials to be learned by the the basic forgetting function—the function characterizing recall probability versus lag between a single study opportunity and a subsequent test. Given the forgetting function, the model makes strong predictions concerning recall performance at test time given a study schedule.

MCM predicts the outcome of four experiments by Cepeda et al. (in press, 2008). These experiments all involved two study sessions with variable ISIs and RIs. Given the basic forgetting functions for the material under study, MCM accurately predicted the ISI yielding maximal recall performance at test for each RI. Although MCM is at an early stage of development, the results we have obtained are sufficiently promising and robust that we find it valuable to explore the model's predictions and to compare them to the well-established ACT-R model.

Comparing Model Predictions

Having introduced the ACT-R model and MCM, we now turn to the focus of this paper: obtaining predictions from the two models to determine whether the models are distinguishable. We focus on the most important, practical prediction that the models can make: how to schedule study to optimize memory retention. We already know that the models make similar predictions in empirical studies with two study sessions (one ISI); we therefore turn to predictions from the models with more than two sessions (two or more ISIs). Even if the models make nonidentical predictions, they may make predictions that are quantitatively so similar the models will in practice be difficult to distinguish. We therefore focus our explorations on whether the models make qualitatively different predictions. Constraining our explorations to study schedules with three study sessions (i.e., two ISIs), we test whether the models predict that optimal study schedules have *expanding*, *contracting*, or *equal* spacing, that is, schedules in which ISI 1 is less than, greater than, or equal to ISI 2, respectively. For the sake of categorizing study schedules, we judge two ISIs to be equal if they are within 30% of one another. The key conclusions from our experiments do not depend on the precise setting of this criterion.

In all simulations, we used the Nelder-Mead Simplex Method (as implemented in Matlab's `fminsearch`) for finding the values of ISI 1 and ISI 2 that yield the maximum recall accuracy following a specified RI. Because this method finds local optima, we used random restarts to increase the likelihood of obtaining global optima. We observed some degenerate local optima, but for the most part, it appeared that both models had spacing functions like those in Figures 2a and 2b with a single optimum.

Our first exploration of the models' spacing predictions uses parameterizations of the models fit to the Glenberg (1976) data (Figure 2a for ACT-R, not shown for MCM). Because the models have already been constrained by the experimental data, which involved two study opportunities, they make strong predictions concerning memory strength following three spaced study opportunities. We used the models to predict the (two) optimal ISIs for RIs ranging from ten minutes to one year. We found that both models predict contracting spacing is optimal regardless of RI. The spacing functions obtained from the models look similar to that in Figure 2c. Because the models cannot be qualitatively discriminated based on the parameters fit to the Glenberg experiment, we turn to exploring a wider range of model parameterizations.

Randomized Parameterizations

In this section, we explore the predictions of the models across a wide range of RIs and model parameterizations, in order to determine whether we can abstract regularities in the models' predictions that could serve to discriminate between the models. In particular, we are interested in whether the optimality of contracting spacing predicted by both models for the Glenberg paradigm and material is due to peculiarities of that study, or whether optimality of contracting spacing is a robust parameter-independent prediction of both models.

Methodology. We performed over 200,000 simulations for each model. In our simulations, we systematically varied the RIs from roughly 10 seconds to 300 days. We also chose random parameter settings that yielded sensible behavior from the models. We later expand on the notion of "sensible."

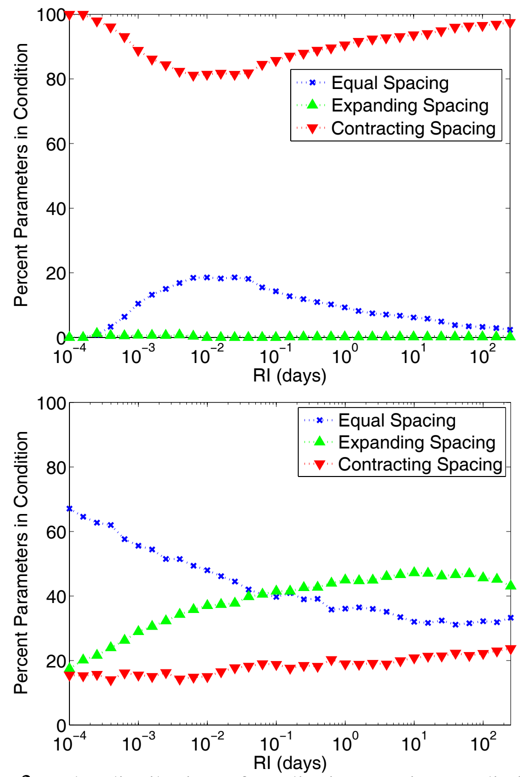

Figure 3: The distribution of qualitative spacing predictions of ACT-R (upper panel) and MCM (lower panel) as a function of RI, for random model variants. Each point corresponds to the percentage of valid model fits that produced a particular qualitative spacing prediction.

For the ACT-R model, we draw the parameters β_i, β_s, β_{si} from Gaussian distributions with standard deviations specified in Pavlik and Anderson (2008). The parameters h, c, and α are drawn from a uniform distribution in [0, 1]. The study weight parameter b is fixed at 1, which assumes test-practice trials (Pavlik & Anderson, 2008). Remaining parameters of the model are fixed at values chosen by Pavlik and Anderson (2008). For MCM, we vary the four parameters that determine the shape of the forgetting function.

To ensure that the randomly generated parameterizations of both models are sensible—i.e., yield behavior that one might expect to observe of individuals studying specific materials—we observe the forgetting function for an item studied once and then tested following an RI, and place two criteria on the forgetting function: (1) With an RI of one day, recall probability must be less than 0.80. (2) With an RI of thirty days, recall probability must be greater than 0.05. We thus eliminate parameterizations that yield unrealistically small amounts of forgetting and too little long-term memory.

Results. Results of our random-parameter simulations are presented in Figures 3 and 4. The upper graphs of each figure are for the ACT-R model and the lower graphs are for MCM. Figure 3 shows, as a function of the RI, the proportion of simulations that yield contracting (red curve), expanding (green curve), and equal (blue curve) optimal spacing. The ACT-R model (Figure 3, upper) strongly predicts that contracting spacing is optimal, regardless of the RI and model parameters. In contrast, MCM (Figure 3, lower) suggests that the qualitative nature of the optimal study schedule is more strongly dependent on RI and model parameters. As the RI increases, the proportion of expanding spacing predictions slowly increases and the proportion of equal spacing predictions decreases; contracting spacing predictions remain relatively constant. Over a variety of materials to be learned (i.e., parameterizations of the model), MCM predicts that expanding spacing becomes increasingly advantageous as the RI increases.

Each scatter plot in Figure 4 contains one point per random simulation, plotted in a log-log space that shows the values of the optimal ISI 1 on the x-axis and the optimal ISI 2 on the y-axis. In other words, each point is like the star (point of optimal retention) of Figure 2c, plotted for a unique parameterization and RI. The two solid diagonal lines represent the decision boundary between the different qualitative spacing predictions. Points between the decision boundaries are within 30% of each other (in linear space) and fall under the label of equal spacing. Points above the upper diagonal line are classified as expanding spacing, and points below the lower diagonal line are classified as contracting spacing. The color of the individual points specifies the corresponding RI.

The spacing functions produced by the ACT-R model are fairly similar, which is manifested not only in the consistency of the qualitative predictions (Figure 3, upper), but also the optimal ISIs (Figure 4, upper). The relationship between optimal ISI 1 and optimal ISI 2 appears much stronger for the ACT-R model than for MCM, and less dependent on the specific model parameterization. Not only do we observe a parameter-independent relationship between the optimal ISIs, but we also observe a parameter-independent relationship between the RI and each of the ISIs. The apparent linearity in the upper panel of Figure 4 translates to a linear relationship in log-log space between RI and each of the optimal ISIs. The least-squares regression yields:

$$\log_{10}(ISI_1) = 1.0164 \log_{10}(RI) + 0.5091$$
$$\log_{10}(ISI_2) = 1.0237 \log_{10}(RI) + 0.9738$$

with coefficient of determination (ρ^2) values of 0.89 and 0.90, respectively. We emphasize that these relationships are predictions of a model, not empirical results. The only empirical evidence concerning the relationship between RI and the optimal ISI is found in Cepeda et al. (2006), who performed a meta-analysis of all cogent studies of the distributed-practice effect, and observed a roughly log-log linear relationship between RI and optimal ISI for experiments consisting of two study sessions (one ISI). Were this lawful relationship to exist, it could serve as an extremely useful heuristic for educators who face questions such as: If I want my students to study this material so that they remember it for six months until we return to the same topic, how should I space the two classes I have available to cover the material?

In further contrast with ACT-R, MCM's optimal ISI predictions are strongly parameter dependent (Figure 4, lower). Is

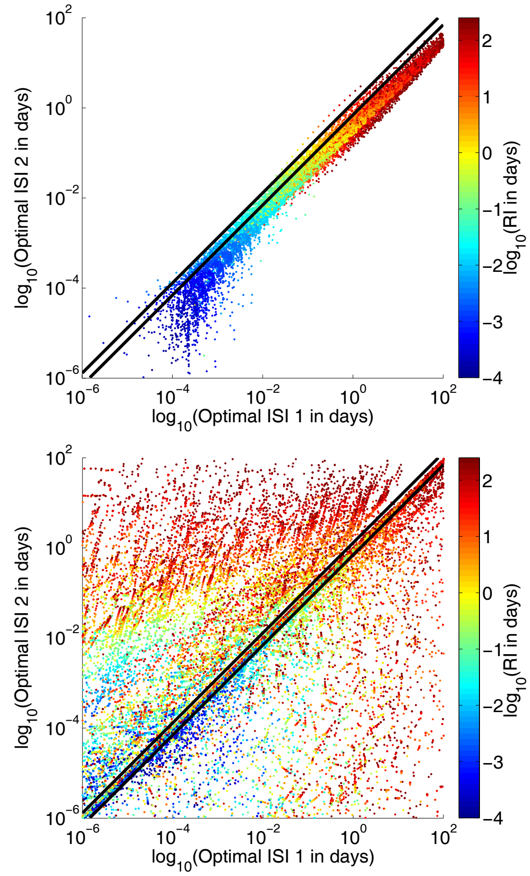

Figure 4: Optimal spacing predictions in log-space of ACT-R (upper figure) and MCM (lower figure) for random parameter settings over a range of RIs. Each point corresponds to a parameter setting's optimal spacing prediction for a specific RI, indicated by the point's color. The black lines indicate the boundaries between expanding, equal, and contracting spacing predictions.

this result problematic for MCM? We are indeed surprised by the model's variability, but there are no experimental data at present to indicate whether such variability is observed in optimal study schedules for different types of material (as represented by the model parameters).

Although ACT-R shows greater regularity in its predictions than MCM, as evidenced by the contrast between the upper and lower panels of Figure 4, note that both models make optimal spacing predictions that can vary by several orders of magnitude for a fixed RI. No experimentalist would be surprised by the prediction of both models that optimal spacing of study for a given RI is material-dependent, but this point has not been acknowledged in the experimental literature, and indeed, the study by Cepeda et al. (2008) would seem to suggest otherwise: two different types of material yielded spacing functions that appear, with the limited set of ISIs tested, to peak at the same ISI.

Another commonality between the models is that both clearly predict the trend that optimal ISIs increase with the RI. This is evidenced in Figure 4 by the fact that the long RIs (red points) are closer to the upper right corner than the short RIs (blue points). Although the experimental literature has little to offer in the way of behavioral results using more than two study sessions, experimental explorations of the distributed-practice effect with just two study sessions do suggest a monotonic relationship between RI and the optimal ISI (Cepeda et al., 2006).

Discussion

In this paper, we have explored two computational models of the distributed practice effect, ACT-R and MCM. We have focused on the educationally relevant issue of how to space three or more study sessions so as to maximize retention at some future time. The models show some points of agreement and some points of fundamental disagreement.

Both models have fit the experimental results of Glenberg (1976). With the parameterization determined by this fit, both models make the same basic prediction of contracting spacing being optimal when three study sessions are involved. Both models also agree in suggesting a monotonic relationship between the RI and the ISIs. Finally, to differing extents, both models suggest that optimal spacing depends not only on the desired RI, but also on the specific materials under study.

When we run simulations over the models' respective parameter spaces, we find that the two models make remarkably different predictions. ACT-R strongly predicts contracting spacing is best regardless of the RI and materials. In contrast, MCM strongly predicts that equal or expanding spacing is best, although it shows a greater dependence on both RI and the materials than does ACT-R. This stark difference between the models gives us a means by which the models can be evaluated. One cannot ask for any better set-up to pit one model against the other in an experimental test.

In reviewing the experimental literature, we have found only four published papers that involve three or more study sessions and directly compare contracting versus equal or contracting versus expanding study schedules (Foos & Smith, 1974; Hser & Wickens, 1989; Landauer & Bjork, 1978; Tsai, 1927). *All four studies show that contracting spacing leads to poorer recall at test than the better of expanding or equal spacing.* These findings are consistent with MCM and inconsistent with ACT-R. However, the findings hardly allow us to rule out ACT-R, because it would not be surprising if a post-hoc parameterization of ACT-R could be found to fit each of the experimental studies.

Nonetheless, the sharp contrast in the predictive tendencies of the two models (Figure 3) offers us an opportunity to devise a definitive experiment that discriminates between the models in the following manner. We conduct an experimental study with a single ISI and parameterize both models via fits to the resulting data. We then examine the constrained models' predictions regarding three or more study sessions. If ACT-R predicts decreasing spacing and MCM predicts equal or increasing spacing, we can then conduct a follow-on study in which we pit the predictions of two fully specified models against one another. We (Kang, Lindsey, & Pashler, in preparation) have just begun this process using Japanese-English vocabulary pairs that Pavlik and Anderson (2008) have modeled extensively with ACT-R. Without extensive simulation studies of the sort reported in this paper, one would not have enough information on how the models differ to offer an approach to discriminate the models via experimental data.

Acknowledgments

This work was supported NSF Science of Learning Center grant SBE-0542013 (Garrison Cottrell, PI), and by NSF BCS-0339103, BCS-720375, and SBE-0518699. We thank Dr. Pavlik for considerately helping us with the ACT-R model.

References

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (in press). Optimizing distributed practice: Theoretical analysis and practical implications. *J. Exp. Psych.* (In press)

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psych. Bull.*, *132*, 354–380.

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

Foos, P. W., & Smith, K. H. (1974). Effects of spacing and spacing patterns in free recall. *J. Exp. Psych.*, *103*, 112–6.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associated and recognition memory paradigms. *J. Verb. Learn. & Verb. Behav.*, *15*, 1–16.

Hser, Y. I., & Wickens, T. D. (1989). The effects of the spacing of test trials and study trials in paired-association learning. *Ed. Psych.*, *9*, 99–120.

Kang, S. H. K., Lindsey, R. V., & Pashler, H. (in preparation). *Investigations into study schedules.* (In preparation)

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In *Practical aspects of memory* (pp. 625–632). London: Academic Press Inc.

Mozer, M. C., Pashler, H., Lindsey, R. V., & Vul, E. (submitted). *Predicting the optimal spacing of study: A multiscale context model of memory.* (Submitted)

Pavlik, P. I. (2007). Understanding and applying the dynamics of test practice and study practice. *Instruc. Sci.*, *35*, 407–441.

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

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *J. Exp. Psych.: Applied*, *14*, 101–117.

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: application of the sam model. *Cog. Sci.*, *27*, 431–452.

Staddon, J. E. R., Chelaru, I. M., & Higa, J. J. (2002). A tuned-trace theory of interval-timing dynamics. *J. Exp. Anal. of Behav.*, *77*, 105–124.

Tsai, L. S. (1927). The relation of retention to the distribution of relearning. *J. Exp. Psych.*, *10*, 30–39.