UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
A Multiple-Trace Memory Model Exhibiting Realistic Retrieval Dynamics

Permalink

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Green, Collin
Kittur, Aniket

Publication Date
2004

Peer reviewed

A Multiple-Trace Memory Model Exhibiting Realistic Retrieval Dynamics

Collin Green (cbgreen@ucla.edu)
Aniket Kittur (nkittur@ucla.edu)
Department of Psychology, 1285 Franz Hall
University of California, Los Angeles
Los Angeles, CA 90095

Abstract

A process model of human memory dynamics is proposed as an implementation of Kittur, Green, & Bjork's (2004) mathematical model. Both models are based on an ideal information processing approach, in which an item's accessibility is based on the predicted future need of that item. The proposed model is an adaptation of the multiple-trace architecture of Hintzman's MINERVA2 model (Hintzman 1984; 1986; 1988). We present simulations of complex spacing and practice dynamics encompassing the mechanics of Bjork and Bjork's (1992) New Theory of Disuse, which accounts for diverse phenomena such as massed vs. spaced practice and spontaneous recovery. In addition, we show how the model explains and simulates time-dependent serial position effects (such as the shift from recency to primacy with delay and time-invariant recency effects). The model's potential as a tool for exploring the relationship between the content of items in memory and more general memory dynamics is also discussed.

Memory as a System for Predicting Need

Kittur, Green, & Bjork (2004) described a mathematical model of memory dynamics inspired by Bayesian statistics. The model is driven by the assumption that memory approximates an ideal information processor, keeping memory items accessible to the degree that they are likely to be needed in the future (see Anderson, 1989 for further rationale on this approach). The predicted future need for an item is computed by the model based on the pattern of past retrievals for that item and the time since it was last retrieved. This is best illustrated by analogy.

Imagine that Book A has been checked out of a library once a month for the past year. Book B, on the other hand, has been checked out every week for the last month but never prior to that. If the librarian was forced to choose which of the two books should be kept readily available, the best choice would change over time. Initially, the librarian would probably keep Book B more readily retrievable as it has been needed frequently in the recent past (possibly, an instructor has assigned reading from this book for a class project); however, after a month has passed with neither book being required, the librarian would likely decide that Book A should be more accessible, given its history of being required at regular, if infrequent, intervals.

The Kittur, Green, & Bjork (2004) model functions in a similar way. It calculates the probability that an item will be needed given three key pieces of information: the average interval between past retrievals; the number of times the item was retrieved in the past; and the time since it was last retrieved. The use of these elements allows for a distinction between item accessibility and item storage, a key insight of the New Theory of Disuse (NTD) (Bjork & Bjork, 1992). The model was inspired by and provides a potential algorithmic basis for the NTD and the complex memory dynamics it explains.

The New Theory of Disuse

The NTD accounts for a variety of effects in the human memory literature. The NTD includes the following assumptions about memory (see Bjork & Bjork, 1992):

1) Memory items are associated with two distinct "strengths": a *storage strength* (SS) and a *retrieval strength* (RS). SS indicates how well-learned an item is (that is, the accumulated history of an item is reflected in its SS). RS, on the other hand, indicates how readily accessible an item is for retrieval. RS alone determines the probability that an item will be successfully recalled from memory. SS does not directly influence memory performance, but has important implications for memory dynamics over time[1].

2) SS does not decrease. SS grows during study or retrieval events as a decelerating function of the current SS. That is, all else being equal, items with low SS benefit from study or retrieval events more than items with high SS. The total storage strength across all items in memory is therefore unbounded. Changes in SS are dependent on both RS and SS. An item gains SS as a decelerating function of its current SS, and as a decelerating function of its current RS.

3) RS increases and decreases. As with SS, an item gains RS as a result of study or retrieval events. When the item is not being studied or retrieved, such as when other items are being attended, RS decreases. As a result, gains in RS for one item necessarily result in a loss of RS for the other (unstudied) items in memory, though these are not necessarily changes of the same magnitude. Changes in RS are dependent on both RS and SS: Gain in RS due to a retrieval or study event is a decelerating function of current RS, and an increasing function of current SS. Conversely, RS loss in is faster the larger the current RS is, and slower with larger SS.

4) Generally, retrieval events are more potent than study events. Increments in both SS and RS are larger when an item is retrieved versus when it is studied.

---

[1] The two-strength theory espoused by NTD and implemented in MNEM is an important difference between it and other related need-based models, such as Anderson's ACT-R (1989). We are currently exploring testable differences between the models.

The postulation of two separate strengths whose magnitudes influence each other is at the core of NTD's account of retrieval and memory dynamics.

The MNEM Model

Many models of human memory employ a strategy that assumes each item stored in memory is represented by a single memory trace. For example, the studied item "horse" would be instantiated as a single mental symbol, and further exposure to "horse" would serve to strengthen or heighten the activation (or gain—i.e. sensitivity to excitation) of that symbol. However, such models struggle with the problem that no two exposures to an item are identical: the spatial, temporal, or subjective context of encoding is variable. Additionally, changes in attention or effort may occur during different exposures to an item and attributes of a stimulus that are important at one point may be more or less important at some point in the future. Multiple trace models of memory are better suited to deal with variable encoding, in that they do not assume that all encodings of an item are linked to a single representation. Such models also do not assume a mechanism for reconciling variable encodings with unitary representation.

MNEM (*M*emory *N*eed *E*xpectation *M*odel), like MINERVA2 and other multiple trace models, works on the assumption that every instance of encoding lays down a new memory trace in the long-term store. If a single stimulus is encoded on multiple occasions (studied and re-studied), then MNEM creates and stores separate traces for each encoding event. Because of random information loss during encoding events (see below), recording new traces for every instance produces variability in the Long Term Memory (LTM) representations of a repeated item. This variability occurs in addition to any variability introduced by context, environment or attention, which may also be introduced to the model.

Representation

The representations upon which MNEM operates are simple, and are adapted from Hintzman's (1984) MINERVA2 model. Each trace in MNEM is an ordered vector of size *n*, with each element taking on the values of -1, 0, or +1. The elements can be thought of as corresponding to specific feature dimensions (e.g. "redness", "roundness", "chair-ness", etc.), with values indicating the absence of a specific feature (-1), the presence of the feature (+1), or a lack of information about the feature (0). The format is open to other interpretations, of course.

Consideration of the history of a memory item depends on the ability to examine past encodings of that item. It is unlikely, however, that any two memory traces are actually identical. That is, identifying instances of trace *T* is simple when literal copies of *T* are stored in several LTM locations, but it is more likely that LTM traces containing the same information are encoded with different contexts, or with different features emphasized. Instead of a single strengthened trace *T*, or many literal copies of *T*, we may store several traces similar to *T*: *T'*, *T''*, etc. As such, it is necessary to resolve some ambiguity about *which* traces in LTM should be counted as instances of a single item.

MNEM uses a specific similarity metric to evaluate the similarity of two memory traces. Borrowing again from the MINERVA2 model, the similarity of an LTM trace *T* to some probe trace *P* is calculated as follows:

$$S(P,T) = (\frac{1}{N_R})\sum_{j=1}^{n} P(j)T(j), \qquad (1)$$

where *n* is the number of elements in the trace and N_R indicates the number of relevant features in the pair of traces. Relevant features are defined as features for which at least one of the two traces contains a non-zero value; in other words, if neither trace contains any information about the presence or absence of a feature, then the feature is not counted as relevant. This similarity function approximates a dot product calculated on the feature sets of the two traces, *T* and *P*.

This representational format is admittedly simplistic, though one advantage of this simplicity is that it requires few assumptions. In fact, the MNEM model requires only two key properties of its representations: they must be amenable to *some* systematic similarity metric, and they must be combinable in a systematic way.[2]

Any representational format that meets these requirements is compatible with the MNEM model. This flexibility makes it amenable to incorporation into diverse cognitive architectures, where other components of the system might necessarily place more serious constraints on the representational format. (As an example, as ordered one-dimensional vectors may be too limiting for representing relational structures, an alternative and appropriate format could be used provided it satisfies the above requirements). That human memory traces satisfy these constraints is a common (if sometimes implicit) assumption among cognitive scientists. The ability to judge the degree to which two stimuli are similar is fundamental to human cognition. Schema abstraction, generalization, and conceptual blending are psychological phenomena that may involve the combination of two or more stimuli to form a composite or abstraction.

Architecture

Like MINERVA2, MNEM has two components: a working memory (WM) and a LTM. WM consists of a buffer that holds a single trace. All inputs to and outputs from LTM are buffered by WM. Traces that are in WM may be encoded into LTM, and information retrieved from LTM is brought into WM.[3]

---

[2] The second requirement is not important for simulating retrieval dynamics, but will be critical in future work when the model is used to generate content from a set of memory traces.

[3] The authors have not attempted to model WM except in the sense that it is a buffer between the world and LTM. In MNEM, multiple traces are not maintained simultaneously, and no attention is required for WM trace maintenance. WM traces may be overwrit-

LTM is simply a collection of memory traces that have been encoded from WM. The current model imposes no (theoretical) limit on the capacity of (the number of traces in) LTM. Each LTM trace is associated with an index. The indices are assigned in the order with traces are encoded into LTM, so that traces encoded earlier have lower indices. The authors consider this equivalent to incorporating spatio-temporal tags on memory traces. Extensions of MNEM may attempt to use a subtler form of spatio-temporal tagging.[4]

Operations

Encoding The encoding operation of the model is relatively simple, and amounts to little more than copying a WM trace into LTM. As discussed, MNEM assumes that variability exists in encoding process (i.e. information is randomly lost during encoding).

The accuracy of encoding depends on a learning rate parameter (L) which indicates the independent probability that any trace feature will be properly encoded (where $0 < L \leq 1$). For example, when $L = 0.7$, seven out of ten features in a trace are accurately copied into the LTM trace (on average). The features that are not properly encoded result in gaps in LTM information (zeros are written into the LTM trace where 1 or -1 existed in the WM trace). During the encoding process, information is only lost, not distorted: a 1 in the WM trace is never erroneously encoded as a -1 in the LTM trace, nor vice versa. This encoding procedure is taken directly from MINERVA2.

Every encoding event yields a new LTM trace, regardless of whether the content of the new trace is redundant with existing LTM traces. The similarity of traces is not considered during the encoding process.

Retrieval Calculating RS for an item is also relatively straightforward. The main complication arises from determining which LTM traces should be considered in the RS calculation when variability exists among different encodings of an item. To address this problem, MNEM "marks" the traces in LTM whose similarity to the probe item exceeds a set criterion. (This criterion similarity, C_s, is a parameter of the model). For example, if C_s is set to 0.75, then only traces for which $S(P,T) \geq 0.75$ will be marked for inclusion in the RS calculation. Once LTM traces are marked, the mean retention interval between them is calculated:

$$\overline{RI(P)} = \frac{1}{(N_m - 1)} \sum_{i=2}^{N_m} [index(M_i) - index(M_{i-1})], \quad (2)$$

where P is the item for which RS is being calculated, M_i is the ith marked LTM trace, and N_m is the total number of marked LTM traces.[5] The *index()* operator simply indicates that the model is using the LTM index for a trace and not the trace itself.

This mean interval is multiplied by the number, or "base rate", of similar instances in LTM. The base rate ($BR(P)$) is equal to the number of marked traces in LTM:

$$BR(P) = N_m. \quad (3)$$

The product of the mean retrieval interval and the base rate[6] is divided by the size of the current retrieval interval, which is the number of time steps that have elapsed since the last marked item was encoded:

$$CI(P) = index(P) - index(M_{max}), \quad (4)$$

where $index(M_{max})$ indicates the index of the timestep during which a marked trace was most recently encoded. Also, $index(P)$, the time index for the encoding of the current item, is simply set to the index of the current timestep (which is equal to the number of traces in LTM plus one: $N_{ltm} + 1$).

In summary, RS can be characterized thus:

$$RS(P) = \frac{\overline{RI(P)} * BR(P)}{CI(P)}. \quad (5)$$

That is, the accessibility of an item P, is equal to the product of the average retention interval between instances like P in LTM and the number of such instances, divided by the interval that has elapsed since the last instance of P occurred.[7]

In order to compare forgetting curves, it is necessary to normalize $RS(P)$. This is accomplished by finding the ratio of logarithm of $RS(P)$ to the maximum value that $RS(P)$ obtains for an item (immediate recall).[8] (Because the log may be negative, we add one to both numerator and denominator for convenience). In all simulations, this ratio is reported as RS. That is:

$$RS_{reported}(P) = \frac{\log(RS(P) + 1)}{\log(RS_{max}(P) + 1)}. \quad (6)$$

---

ten, but this is the only way that information is "lost" from MNEM's WM.

[4] The authors are currently exploring the incorporation of a context vector into encoded representations, or giving individual traces an activation value which would be initialized to some maximum at encoding, and would decay over time. In the latter strategy, the activation value would represent a trace's "age" for purposes of calculating RS. The RS calculation would consider the difference between the activations of two traces. This approach remains untested, but seems promising in that the decay function would likely be non-linear, decelerating as it approaches zero. This being the case, two traces equally displaced in absolute time would become less discriminable with age.

[5] When only a single trace in LTM is marked, the average retention interval is defaulted to a value of 1.

[6] This product is the closest analog to SS in MNEM:

$SS(P) = \overline{RI(P)} * BR(P)$. Note that unlike RS, SS is strictly increasing with additional study, and is not subject to decay. SS influences changes in RS, most importantly by retarding the loss of RS over time (see Figure 3).

[7] This definition of RS is at the core of the Kittur, Green, and Bjork (2004) model, which exhibits the same memory dynamics in a single-trace architecture.

[8] See Pavlik & Anderson (2003) for rationale on scaling forgetting using the maximum (current) activation of a trace.

The result of normalization is that immediate recall yields a reported RS of 1, and any delay in recall produces an RS between 0 and 1. This allows for comparison of forgetting curves in terms of probability of recall.

Trace Composition The formation of composite traces from a set of LTM traces is also important in this model. We have specified how one may calculate the RS of a specific item in LTM, but retrieving useful information from the LTM store is another matter entirely. MINERVA2 includes a mechanism that uses similarity to weight traces in LTM, and forms a composite "echo" by averaging these weighted traces. MNEM employs a similar strategy, but instead of *all* traces in LTM, only those that exceed the similarity criterion are weighted and averaged. While this is an important aspect of the model, and may allow simulation of important memory phenomena (e.g. encoding specificity, context effects, etc.) the details of this operation are not directly relevant to the retrieval dynamics discussed here, and we will leave them for another time.

Simulation Results

The NTD was conceived to "post-dict" a number of memory effects. In the previous discussion of that theory, behavioral correlates of RS and SS were noted. MNEM implements the same relationships between RS and SS and its performance is similar to that of humans on a variety of memory tasks.
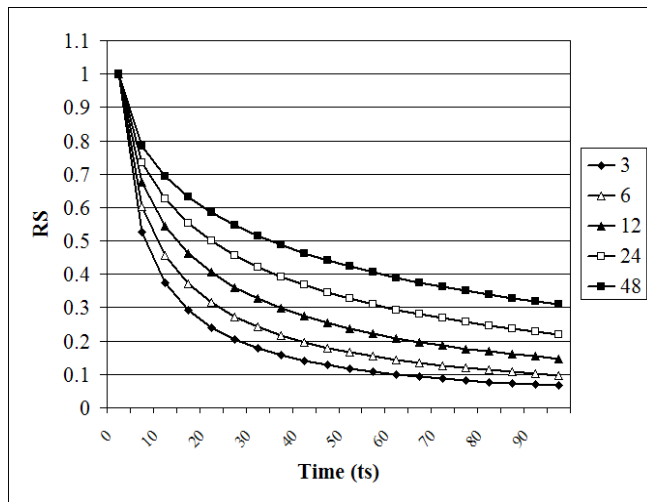

Figure 1: Forgetting curves for items studied three times each, with inter-item intervals of 3, 6, 12, 24, or 48 timesteps.

Forgetting Over Time

MNEM displays forgetting curves that closely resemble those of human subjects. Behavioral data suggest that the probability of recalling a once-studied item declines as a function of the retention interval. More specifically, access to an item declines as a function of intervening experience (Thorndike, 1914; McGeoch, 1932; Bjork & Bjork, 1992).

NTD postulates that probability of recall is linked to RS only, but that changes in RS are mediated by SS. The particular rate of forgetting for an item is influenced by the frequency of exposure to an item (Melton, 1967; Krueger, 1929), as well as the interval between exposures (Peterson, Hillner, & Saltzman, 1962; Whitten & Bjork, 1977). MNEM captures the general shape of forgetting curves, and simulates frequency and spacing effects observed in human data.

In simulation, a single item A is studied according to various schedules. At various delays, the RS of A is calculated, which indicates the probability that it would be recalled at that interval since last study. To simulate the passage of time without study or retrieval events, a randomly generated memory trace is encoded into LTM on each timestep.[9] Note that in simulation, the calculation of RS does not affect the state of LTM.

The simulated practice schedules vary in the number of exposures of A, as well as in the spacing of exposures. Forgetting curves generated by MNEM for items studied with equal frequency, but different inter-item intervals are shown in Figure 1. Figure 2 shows forgetting curves for items studied at equal intervals, but with different frequencies.
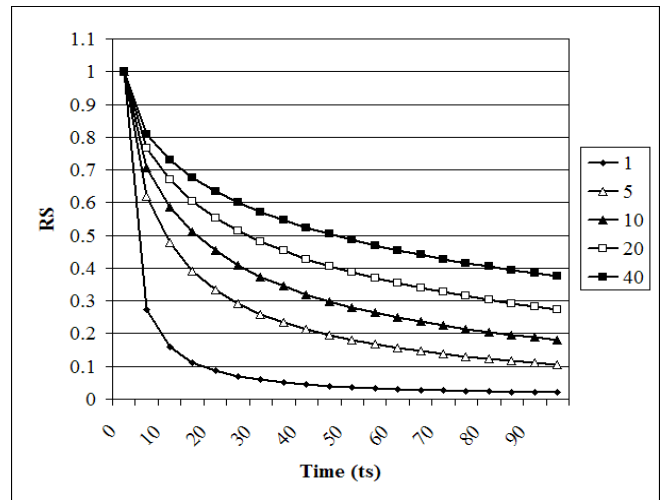

Figure 2: Forgetting curves for items studied with equal spacing (3 timesteps between exposures) and frequencies of 1, 5, 10, 20, or 40 exposures.

Spacing and frequency effects are important aspects of human memory in that they give rise to more complicated phenomena. For example, in some circumstances an old habit may be replaced with a new behavior, only to re-emerge at a later time, a phenomenon known as spontaneous recovery (Estes, 1955; Koppenaal, 1963).

---

[9] It is worthwhile to note that the noise introduced to the LTM system is relatively unconstrained. In fact, the same method that generates the "studied" trace for these simulations is used to generate the "noise" traces that are interpolated between the study event and the sampling of RS.

NTD and the MNEM model account yield spontaneous recovery as a natural consequence of different forgetting rates. Simulation data in Figure 3 show spontaneous recovery. Item *A* represents an old response that has been learned over a long period of time. Item *B* is a new response intended to replace *A*. As *B* is acquired, *A*'s RS decays substantially. However, we observe that *A* gains an advantage after a certain delay. If *B* is not practiced, the larger SS of item *A* yields a shallower forgetting curve. The decay of RS is slower for trace *A* than for trace *B* and the curves cross over. The older habit will remain more accessible thereafter.

Primacy & Recency

Primacy and recency are well-known memory phenomena. When a list is studied, items that appeared early in the study list are more recallable than items near the middle of the list. Primacy effects have been attributed to covert rehearsal between study presentations (Glenberg et al., 1980). Effectively, subjects create extra study opportunities in the gaps between item exposures.
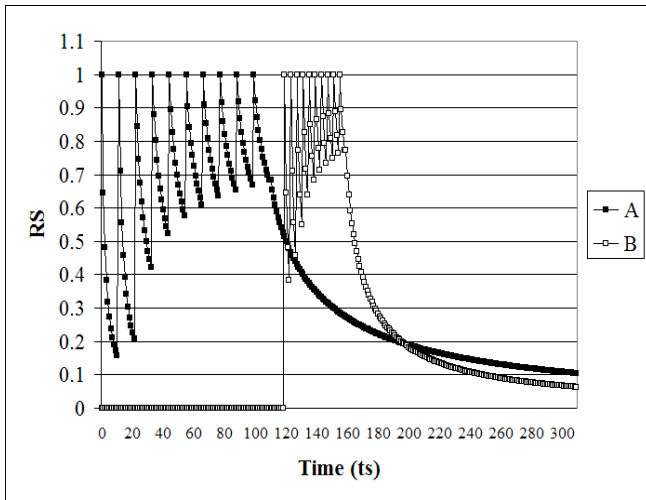

Figure 3: Spontaneous recovery of response *A* occurs after learning response *B*. This is due to a larger increase in response *A*'s SS at reminding, owing to a lower RS at that time.

Similarly, items presented near the end of the list are also recalled better than mid-list items. Recency results from the relatively short retention interval between study and test. Prior work has demonstrated that there is a shift from recency to primacy over increasing retention intervals (Craik, 1970; Knoedler, Hellwig, & Neath, 1999). The MNEM model shows similar behavior.

In simulation, a list of 20 items is studied, with five timesteps between study events. Between trials, the simulated subject is assumed to perform covert rehearsal on some of the items presented so far, in order, for the duration of the interval. This strategy lasts for a limited number of presentations (three, in this simulation), at which point the simulated subject is assumed to become overwhelmed by

the number of items and therefore abandons the covert rehearsal strategy. Beyond this point, inter-trial intervals are filled with random traces, as in previous simulations.[10]

At the end of the study phase, the RS for each of the 20 items is calculated at five different retrieval intervals. The serial position curves that result are shown in Figure 4. Three features are notable: the prominence of recency effects in immediate recall; the presence of primacy in all serial position curves; the shift of recency to primacy as the dominant pattern in the data as the retention interval grows.
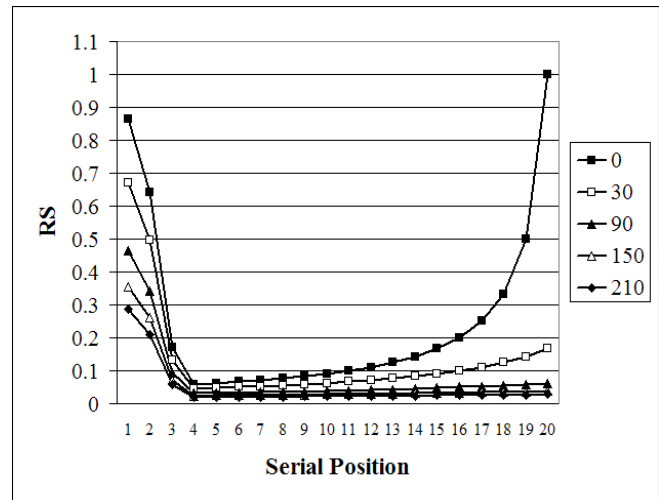

Figure 4: Serial position curves at delays of 0, 30, 90, 150, and 210 timesteps. Note the rapid decay of recency effects relative to the slower decay of primacy.

The recency effects observed in simulation share a subtle property with human behavioral data: time-invariance. Some data from humans suggest that the magnitudes of recency effects follow a ratio rule (Glenberg et al, 1980; 1983; Bjork & Whitten, 1974). This phenomenon was described mathematically by Bjork & Whitten (1974). Specifically, recency effects scale with the log of the ratio of mean presentation interval divided by the current retention interval:

$$recency \propto \log(\frac{\overline{RI(P)}}{CI(P)}) \qquad (7)$$

This behaviorally-derived ratio rule is inherent in the MNEM model (see Kittur, Green & Bjork, 2004). Figure 5 shows serial position curves for various ratios of mean retentional interval to current retention interval.

Conclusions and Future Directions

The model described here shows memory dynamics that are consistent with human behavioral data. Forgetting curves, spacing and frequency effects, and serial position curves are generated in simulation by following the assump-

---

[10] Glenberg, et al., (1980) observed that primacy effects were eliminated when participants were prevented from performing cumulative rehearsal on early list items.

tions of NTD, and allowing items to accumulate independent SS and RS.

The relative simplicity of this model (and its more general mathematical formulation in Kittur, Green, & Bjork, 2004), makes it a useful tool for exploring subtle issues in memory and generating concrete experimental predictions. There is potential to extend our understanding of retrieval dynamics to a greater diversity of memory phenomena by manipulating the content of the memory traces used in simulation. For example, MNEM provides a natural platform for exploring the influence of inter-item associations on memory dynamics. In addition, MNEM may prove to be informative on issues surrounding schema abstraction, categorization, and other arenas where knowledge content is an issue. Context, encoding specificity, and variability effects may also be amenable to analysis with this model in the future.
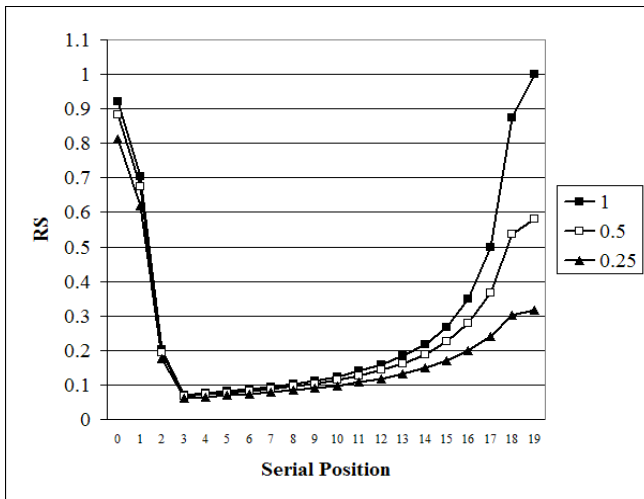

Figure 5: The magnitude of recency effects in MNEM scale with the ratio of mean retention interval to current retention interval. Serial position curves are shown for spacing to retention interval ratios of 1, 0.5, and 0.25.

Acknowledgments

References

Anderson, J. R. (1989). A rational analysis of human memory. In H. L. Roediger, III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 195-210). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Bjork, R.A. & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Schiffrin (Eds.), *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes* (Vol. 2, pp.35-67). Hillsdale, NJ: Erlbaum.

Craik, F.I.M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior, 9*, 143-148.

Estes, W.K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62*, 145-154.

Glenberg, A.M., Bradley, M.M., Stevenson, J.A., Kraus, T.A., Gretz, A.L., Fish, J.H., & Turpin, B.N. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 355-369.

Hintzman, D.L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*, 96-101.

Hintzman, D.L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93(4)*, 411-428.

Hintzman, D.L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95(4)*, 528-551.

Kittur, A., Green, C., & Bjork, R.A. (2004, July). A need-based model of human memory retrieval. Poster presented at the 112th Annual Meeting of the American Psychological Association: Honolulu, HI.

Knoedler, A.J., Hellwig, K.A., & Neath, I. (1999). The shift from recency to primacy with increasing delay. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 474-487.

Koppenaal, R.J. (1963). Time changes in strength of A-B, A-C lists: Spontaneous recovery? *Journal of Verbal Learning and Verbal Behavior, 2*, 310-319.

Krueger, W.C.F. (1929). The effects of overlearning on retention. *Journal of Experimental Psychology, 12*, 71-78.

McGeoch, J.A. (1932). Forgetting and the law of disuse. *Psychological Review, 39*, 352-370.

Melton, A.W. (1967). Repetition and retrieval from memory. *Science, 158*, 532.

Pavlik Jr., P. I., & Anderson, J. R. (2003). An ACT-R model of the spacing effect. In F. Detje, D. Dorner & H. Schaub, *Proceedings of the Fifth International Conference of Cognitive Modeling* , 177-182.

Peterson, L.R., Hillner, K., & Saltzman, D. (1962). Time between pairings and short-term retention. *Journal of Experimental Psychology, 64*, 550-551.

Thorndike, E.L. (1914). *The psychology of learning*. New York: Teachers College.

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