Psychological Review
1989, Vol. 96, No. 4, 703–719

0033-295X/89/$00.75

Human Memory: An Adaptive Perspective

John R. Anderson and Robert Milson
Carnegie Mellon University

It was argued that the basic principles of operation of human memory can be understood as an optimization to the information-retrieval task that human memory faces. Basically, memory is using the statistics derived from past experience to predict what memories are currently relevant. It was shown that the effects of frequency, recency, and spacing of practice can be predicted from the statistical properties of information use. The effects of memory prompts, cues, and primes can be predicted on the assumption that memory is estimating which knowledge will be needed from past statistics about interitem associations. This analysis was extended to account for fan effects. Memory strategies were analyzed as external to the process of statistical optimization. Memory strategies are attempts to manipulate the statistics of information presentation to influence the optimal solution derived by memory. The classic buffer-rehearsal model for free recall is analyzed as a strategy to manipulate the statistics of information presentation.

Human memory is typically viewed by lay people as quite a defective system. For instance, over the years we have participated in many talks with artificial intelligence researchers about the prospects of using human models to guide the development of artificial intelligence programs. Invariably, the remark is made, "Well, of course, we would not want our system to have something so unreliable as human memory." Actual memory researchers seldom comment on the adaptiveness of memory (but see Bjork & Bjork, 1988; Sherry & Schacter, 1987). One seldom finds arguments for a theory of memory mechanisms cast in terms of the adaptiveness of these mechanisms. Rather, the typical argument for a memory mechanism is by reference to its ability to fit the data at hand. The implied inference is that the actual combination of mechanisms is pretty arbitrary. Certainly, this characterization is fairly accurate of our own writings on memory despite our advocacy of the ACT* theory (Anderson, 1983) whose initials stand for "Adaptive Control of Thought".

In this article, we argue that human memory is adaptively designed and that we can understand a great deal about memory phenomena by understanding its adaptiveness. The analysis being offered here is not meant to supplant existing mechanistic accounts such as ACT* but to supplement them by showing that they implement a rational memory design. This article begins with a framing of the information-processing problem that memory faces. Then we derive the optimal memory behavior, given this framing. We show that this predicts many of the major results in human memory. These results are very general trends that appear across particular memory paradigms. As a final topic, we consider what would be involved in getting this framework to apply to the classic free-recall paradigm.

This research was supported by Grant 87-51890 from the National Science Foundation.

We would like to thank Robert Bjork, Bill Estes, Gary Gillund, Bill Jones, Clayton Lewis, Timothy McNamara, and Lynne Reder for their comments on various drafts of this article.

Correspondence concerning this article should be addressed to John R. Anderson, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213-3890.

Framing of the Information-Retrieval Problem

The framing we offer for human memory comes from a subfield of computer science called, curiously, *information retrieval* (Salton & McGill, 1983).[1] The generic information-retrieval system has a data base of stored items and must respond with a subset of these, given a query that consists of some keywords. Perhaps the most frequent use of such systems in academia is in library searches in which we provide some content words and the system responds with a list of possible books and their abstracts. Such systems have to deal with access to very large data bases in the presence of very limited and uncertain cues. We think this is essentially the human situation. Such systems, like the human system, have to deal with balancing two types of errors that can be made in the face of this uncertainty. As with human memory, the system may fail to retrieve the desired item, which clearly is a costly error. However, there is also a cost associated with retrieving an inappropriate item—that is, the user's cost in considering it and rejecting it. Thus, the information-retrieval system cannot just deal with the problem of undergeneration by retrieving everything. In the field of information retrieval, the problem of generating the desired items is called *recall* and the problem of not generating irrelevant items is called *precision*.

According to Salton and McGill (1983), a typical information-retrieval system consists of four components:

1. There are files that are the items to be retrieved. In human memory, these units are propositions, productions, images, associations, schemata, or whatever your favorite flavor of cognitive chunk.

2. These files are indexed by a number of terms. These terms

[1] The symmetric possibility has also been explored—namely, designing information retrieval systems to correspond to human memory—see Jones (1987).

can be things like the keywords that appear in a book's abstract. In the human case, these terms are the items unified in the chunk. Thus, if we remember a paired associate, the stimulus and response would be among the terms.

3. There are queries that are presented to the system that consist of sets of terms. For instance, we might be asked "What is the response associated with *dog* in list 2?" In this case *dog* and *list 2* would be the terms.

4. There are a set of target structures or files that provide an answer to the query.

A basic assumption is that there is a cost associated with our system's retrieving an irrelevant item and our having to process it. Let us denote this cost as C. Our analysis will deliberately avoid inquiring as to the mechanisms of retrieval behind this cost. We want to see how far we can get just with the assumption of a retrieval cost. The most obvious form this cost takes in models of memory is retrieval time, but there may be other dimensions to this cost such as metabolic expenditure.

The Optimization Problem

Now we are in the position to define the optimization problem for human memory. Let G be the gain associated with retrieving the target memories. Let $p[\mathrm{A}]$ be an estimated probability that memory structure A is relevant—that is, a target. A rationally designed information-retrieval system would retrieve memory structures ordered by their probabilities $p[\mathrm{A}]$ and would stop retrieving when

$$p[\mathrm{A}]G < C. \tag{1}$$

That is, the system should stop retrieving when the probabilities are so low that the expected gain from retrieving the target is less than the cost of retrieving that item. A basic assumption is going to be that $p[\mathrm{A}]$ is monotonically related to latency and accuracy of recall, which are the two major dependent measures used in the literature. It is related to latency because memory structures are ordered according to $p[\mathrm{A}]$, and it is related to accuracy because of this threshhold on what items will be considered. To be able to predict speed and accuracy of recall, we need to inquire as to what factors memory can use to estimate $p[\mathrm{A}]$, and our prediction will be that these factors determine memory performance.

Before turning to an examination of $p[\mathrm{A}]$—the major focus of this article—it is worth saying something about C and G and how they might vary. G should vary with the importance of the task, implying that people should try longer before quitting in more important tasks, and recall more. Recall is known to improve gradually as subjects try longer (Buschke, 1974). C should vary with the time spent inspecting an item before accepting or rejecting it as relevant to the current task. Varying it should lead to speed–accuracy trade-off functions in which longer recall times are associated with more accurate memories.

This discussion of the optimization problem is framed in serial terms—first, the subject considers one target structure, then another, and so on. However, it should be clear, given our knowledge of the parallel-serial equivalence (Townsend, 1974), that this is nothing more than an expository convenience. Indeed, we think of this all as being implemented in the parallel pattern-matching machinery of ACT* (Anderson, 1983). In ACT*, the system can assign resources to the structures it is processing according to their plausibility, and the system can effectively ignore structures below some threshold of plausibility. Thus, whether parallel or serial, the critical feature is that knowledge structures are ordered in terms of plausibility until they become too implausible to consider. It is not the goal of this article to inquire in detail as to the mechanisms that achieve this, only to inquire whether we can predict memory performance, assuming that memories are so ordered.

Estimation of Likelihoods of Memory Structures

We are now one big step from having a theory that specifies the behavior of memory from purely rational considerations. That one big step is to specify the $p[\mathrm{A}]$ in the preceding discussion. We will refer to $p[\mathrm{A}]$ as the *need probability,* inasmuch as it is the probability that A is needed.

One solution to the estimation of $p[\mathrm{A}]$ that appears in the computer information-retrieval literature (Bookstein & Swanson, 1974, 1975) is to use Bayesian estimation procedures. The two obvious pieces of information for evaluating whether a memory structure will be relevant are its past history and the terms in the query. Thus, each structure A has some history $\mathrm{H_A}$ of being relevant in the past. The current context consists of a set of terms that we will call *cues* and denote by indices, i. We will denote the set of cues as Q, for query. In doing Bayesian estimation, we are trying to calculate the posteriori probabilities, giving us the following equation:

$$\frac{P(\mathrm{A}\,|\,\mathrm{H_A}\ \&\ \mathrm{Q})}{P(\bar{\mathrm{A}}\,|\,\mathrm{H_A}\ \&\ \mathrm{Q})} = \frac{P(\mathrm{A}\,|\,\mathrm{H_A})}{P(\bar{\mathrm{A}}\,|\,\mathrm{H_A})} \times \prod_{i \in \mathrm{Q}} \frac{P(i\,|\,\mathrm{A})}{P(i\,|\,\bar{\mathrm{A}})}. \tag{2}$$

That is, the odds ratio for Item A is the product of the odds ratio for Item A, given history $\mathrm{H_A}$ times the product of the ratio of the conditional probabilities for each cue in the context. This equation makes certain assumptions about conditional independence, namely, that the degree to which A affects the probability of i in the context does not depend on A's past history or the other cues in the context. Formally, the assumption is

$$\frac{P(i\,|\,\mathrm{H_A}\ \&\ \mathrm{A}\ \&\ \{\mathrm{Q}-i\})}{P(i\,|\,\mathrm{H_A}\ \&\ \bar{\mathrm{A}}\ \&\ \{\mathrm{Q}-i\})} = \frac{P(i\,|\,\mathrm{A})}{P(i\,|\,\bar{\mathrm{A}})}. \tag{3}$$

This assumption is typically made in the computer information-retrieval literature for purposes of tractability.[2]

The first term in Equation 2, $P(\mathrm{A}\,|\,\mathrm{H_A})/P(\bar{\mathrm{A}}\,|\,\mathrm{H_A})$, is basically a prior odds ratio for the item, given its frequency and recency of occurrence. This is the *history factor.* The other quantities, the $P(i\,|\,\mathrm{A})/P(i\,|\,\bar{\mathrm{A}})$ are the odds ratios of the conditional probabilities of the cues, given that the structure is relevant versus not relevant. These ratios can be thought of as associative strengths. They constitute the *context factor.* We will discuss each at length in subsequent sections.

If one is willing to make a somewhat different assumption:

[2] Human memory may not be so constrained, and it is interesting to inquire as to which of our predictions might be upset by nonindependence.

$$\frac{P(i|H_A \& A \& \{Q-i\})}{P(i|H_A \& \{Q-i\})} = \frac{P(i|A)}{P(i)}, \quad (4)$$

then we can write an equation that is for some purposes more useful.

$$P(A|H_A \& Q) = P(A|H_A) * \prod_{i \in Q} \frac{P(i|A)}{P(i)}. \quad (5)$$

Equation 5 gives us a direct formula for the need probability rather than the odds ratio. Note, $P(i|A)/P(i) = P(A|i)/P(A)$.

The basic behavioral assumption will be that memory performance will be monotonically related to the conditional need probability, $P(A|H_A \& Q)$. Later in the article we offer a proposal for how increased need probability maps into higher recall probability and faster reaction times.

The History Factor

To address the history factor, $P(A|H_A)$, we need to determine how the past history of a structure's usage predicts its current usage. To determine this in the most valid way, we would have to follow people about in their daily lives, keeping a complete record of when they use various facts. Such an objective study of human information use is close to impossible. What is possible is to look at records from nonhuman information-retrieval systems that can be objectively studied. Such studies have been done, for instance, of borrowing from libraries (Burrell, 1980; Burrell and Cane, 1982) and of access to computer files (Stritter, 1977). Both systems tend to yield rather similar statistics. If we believe that the statistics of human memory information retrieval mirror the statistics of these nonhuman systems, we are in a position to make predictions about how the human should estimate need probabilities given past history.

Should we really believe that information retrieval by humans has the same form as library borrowings and file accesses? The fact that two very different systems display the same statistics suggests that there are "universals" of information retrieval that transcend device (library, file system, or human memory) and that these systems all obey the same form but differ only in parameterization. Also, when we look at Burrell's (1980, 1985) explanation of library borrowings, it seems plausible that it would extend to human memory and other information-retrieval systems. Finally, the success of the model in accounting for human memory suggests that it applies.

In this section we develop a mathematical model of information use for the human system, assuming it is analogous to these objectively observable information-retrieval systems. From this we will derive predictions for human memory by an optimization analysis. In case the basic point gets lost in the mathematics to follow, we want to state it up front: A system that is faced with the same statistics of information usage as a library or a file system and that is optimized in the sense already defined will produce the basic human memory functions.

Burrell (1980, 1985) has developed a mathematical theory of usage for information-retrieval systems such as libraries (a similar model appears in Stritter, 1977, for file usage). His theory involves two layers of assumptions. First, Burrell assumed that the items (books, files, memory structures) in an information-retrieval system vary in terms of their desirability. He assumed that they vary as a gamma distribution with parameter b and index v. Such a distribution produces mean desirability values of v/b and variances v/b^2. Desirabilities can be interpreted as mean rates of use (in Burrell's case, usage is borrowing). Burrell's second assumption is that uses are described by a Poisson process. This means that, given an item with desirability λ, the time to next use is an exponential process with mean $1/\lambda$.

Anderson (1989) has developed an analysis of human memory using this model of Burrell's (1980). Here we want to consider a sophistication of this model based on some ideas in Burrell (1985). Burrell noted that there are problems with the ahistorical character of the exponential distribution of times until the next borrowing. The problem he is concerned with in library systems is that there is an aging phenomenon—books become less used with time. He chose to model the aging process by assuming that borrowings are a nonhomogeneous Poisson process whose rate varies as $r(t)\lambda$, as a function of time. In the simple homogeneous case, $r(t) = 1$. In his model for aging, Burrell assumed $r(t) = e^{-at}$. In our initial derivations, we derive some general equations that are independent of the form of $r(t)$. Then we will develop a specialization to correspond to our assumptions.

Formally (see Berger, 1985, for discussion of conjugate Bayesian families), what the Burrell (1985) model gives us is a prior distribution for the desirability, λ, of an item. We can specify this as the gamma distribution,

$$\pi(\lambda) = \frac{b^v \lambda^{v-1} e^{-b\lambda}}{(v-1)!}. \quad (6)$$

What we observe of a particular item is that it was used n times in the first t time units since its creation. What Burrell's (1985) model tells us is the probability that we would see such a history, given a book with desirability λ, is described by the equation for a nonhomogeneous Poisson process:

$$p(n, t|\lambda) = \frac{e^{-\lambda M(t)}[\lambda M(t)]^n}{n!}, \quad (7)$$

where

$$M(t) = \int_0^t r(s)ds.$$

What we are interested in is the posteriori distribution of λ, given n and t. A Bayesian estimate of λ be derived as

$$\pi(\lambda|n, t) = \frac{\pi(\lambda)p(n, t|\lambda)}{\int_0^\infty \pi(x)p(n, t|x)dx}, \quad (8)$$

which has the solution

$$\pi(\lambda|n, t) = \frac{(M(t)+b)^{v+n}\lambda^{v+n-1}e^{-\lambda(M(t)+b)}}{(v+n-1)!}, \quad (9)$$

which is itself a gamma distribution with parameters $M(t) + b$ and $v + n$. The mean of this distribution is $(v + n)/(M(t) + b)$. Because this is a rate, it could potentially be greater than 1. However, if we set our time scale so that we are looking at a rate for a small enough unit, this quantity effectively becomes a probability of a use in that interval—that is, a quantity that will

vary from 0 to 1. For instance, if we measured time in seconds, this would give us probability of use in a second. Because there is a decay process, we can take our estimate of $P(A|H_A)$ to be this quantity times $r(t)$ for decay. Thus,

$$P(A|H_A) = \frac{v+n}{M(t)+b} r(t). \quad (10)$$

In the case where $r(t) = 1$ (i.e., no aging or loss of desirability), this becomes

$$P(A|H_A) = \frac{v+n}{t+b}. \quad (11)$$

Burrell (1980) tended to get best fits for library borrowing by estimating b to be about 1 year and v to be about one. This yields the average borrowing rate for a book of about once per year. The low value of v implies that most books have very low borrowing rates and a few have high rates. In the applications in this chapter, we have set $v = 2$ to give a somewhat more normal distribution of desirabilities for human memory. (Unlike Burrell, we do not have the same data base to directly estimate v.) The b parameter defines the time scale. We have set it arbitrarily at 100 and have tried to scale the results of experiments to fit the time scale of the model. It remains a future goal to try to get a systematic estimation of b and v for the human case. Many of the analyses are sensitive only to the general function and not to the exact values of b and v.

$P(A|H_A)$ is not the same thing as need probability because it is only conditioned on the history and not on the context factor. Equation 2 is needed to integrate the two. Nonetheless, in many situations the context is the same for a set of items, and only their experimental history varies. For such situations it is reasonable to treat $P(A|H_A)$ as the need probability because it is directly related to the actual need probability. We do this in discussing recency, frequency, and spacing effects.

Frequency and Recency

There are further developments that will complicate Equation 10, but it is of interest to inquire how its current simple form relates to the basic variables of presentation frequency and recency. Consider the retention function in which we wait t seconds after an item is studied and first test it. In this case, $n = 0$ and Equation 10 takes the form $vr(t)/(M(t) + b)$. Thus, depending on the form of $r(t)$, the function in Equation 10 could give a very good mimicry of human retention functions, which are typically characterized as rapidly negatively accelerated. We will specify $r(t)$ after considering the spacing effect.

The function in Equation 10 can also be analyzed to predict the relation between frequency of exposure and memory performance. In this case, n is our independent variable. Let us assume that t is constant—that is, we are manipulating number of exposures in a fixed interval. The form of Equation 10 is $I + sn$, where $s = r(t)/(M(t) + b)$ and $I = v*s$. Such a linear growth model is the strengthening model in ACT*, which has been shown to yield a good approximation to human learning data.[3] Again we will develop a more precise mapping in subsequent sections.

Spacing Effects

According to Burrell's (1985) model, it does not matter what the spacing is of these n presentations. All that matters in Burrell's model is the total number of uses (n) and the total elapsed time (t).[4] This is a consequence of the fact that the rate of a Poisson process depends only on the time (according to the function $r(t)$) and not on the history of past events. The question arises as to whether Burrell's model correctly describes the likelihood function. Is it the case that in information-retrieval systems there is no massing of need? Burrell's model implies that the probability of next presentation should depend only on the elapsed time since the item was introduced and not on the time since the past presentation of the item. In fact, Burrell's model is not descriptively accurate here, as one might expect. For instance, in Carnegie Mellon University's library system, there are definite clusterings of borrowings, and one can reject the hypothesis that the n borrowings of a book in a fixed time interval are independent samples from a monotonically decreasing probability density. There are lots of reasons for such massings, such as a book being relevant to a course taught only one semester. Stritter (1977) noted such deviations from independence of accesses, but chose to ignore them in developing his model of file system access. It is fairly intuitive that the same is true of human memory, although it is hard to objectively verify what the human likelihood function is.

If some use is massed and some is not, then the intervals between successive uses should predict the probability of the item being needed now. Thus, compare one item that has been used fairly uniformly n times over the year and another item whose n uses all occurred in a 3-month period, 6 months ago. Clearly, the first is more likely to be needed now. Thus, we would predict better memory for spaced items, as long as we are not comparing with a massing of study that has just occurred.

The Augmented Burrell Model

The question arises as to what kind of formal model might underlie the observed behavior of libraries and file systems. Burrell's (1985) model provides us with two assumptions that he was able to justify in the library domain: (a) There is a distribution of desirability of items in which the desirability of an item controls the rate at which it is accessed, and (b) there is an aging process for items and their effective rate of use decays with time.

To deal with spacing, we need to complicate this model in two ways, both of which seem plausible. (c) There is also variation in the rate of decay across items. For simplicity, we will assume that the decay rates are exponentially distributed. This means that the probability of a decay rate, d, is $\alpha e^{-\alpha d}$. In a library system this gives us the distinction between classics and flashes-in-the-pan. (d) Items in the set undergo random revivals of interest

[3] Here and elsewhere we will be making the assumption that every time an item is presented for study, we incur another need for the memory trace of the item.

[4] Burrell (personal communication) indicated that he is aware of discrepencies from this assumption, such as those we have described. However, these were not important to his applications.

in which they return to their original level of desirability. The probability of a revival at time t is an exponential process with probability $\beta e^{-\beta t}$. In a library system this gives us the effect of current events (e.g., a course) suddenly making a book relevant again. It is these revivals that will give us the massing that we observe in library systems and file systems.

If Assumptions c and d do describe features of the environment, then the system will have to adapt to them, just as the features in the original Burrell (1980) model. We derived predictions from our augmented Burrell model by estimating mean need probabilities by Bayesian means, assuming that the environment is as described in Assumptions a to d. In deriving predictions from the augmented Burrell model, the following settings were used for the environmental parameters: $v = 2$, $b = 100$, which establishes our time scale, $\alpha = 2.5$, and $\beta = 0.04$. The last two are arbitrary and can be questioned. They are set so that the mean rate of decay ($1/\alpha$) is 10 times the revival rate.

The augmented model poses serious complications. Because of the revival process, we no longer have independence of rate from past history. Thus, it is necessary to estimate need probabilities by Monte Carlo means. It takes on the order of 100,000 runs to get stable estimates for 100 time units, given the aforementioned parameters. Each run involves (a) choosing a random revival pattern and decay rate and (b) calculating the expected need probability under that revival pattern and decay rate. The more time units, the more runs to get stable estimates and the longer each run takes. The actual algorithm that performs these calculations is described in the Appendix.

Figure 1 examines the relationship between delay and the need probability—that is, the retention function.[5] We have plotted on a log–log scale the relationship between need probability and t. We are assuming that the item was introduced t time units ago and has not been used in the intervening interval. The reader can confirm the linear relationship that exists, implying that retention is a power function of delay. Such power functions are typically found in the experimental literature. One might have thought the rate of decay would be a more rapid exponential to reflect the aging process. However, the revival process slows down this decay process. In the long term, the retention function is more dominated by the fact that the Bayesian estimation becomes more and more biased to low desirabilities if the item has not been used. If we think of Equation 10 as providing the retention function, this amounts to saying that $r(t)$ reaches a nonzero asymptote because of the revival process, but $M(t)$ continues to grow and dominates the retention function in the long run.

When it comes to looking at practice effects, one is forced to try to deal with conflicting variables—the number of exposures to the item, the spacing of exposures, the total interval, and the time from last exposure to test. One cannot hold all of these variables constant and have only the number of exposures vary. Typically, time from last exposure is held constant because of the large retention effects. One either holds spacing constant and lets total time vary or holds total time constant and lets spacing vary. Thus, we have graphed both functions in Figures 2 and 3. In Figure 2, an exposure was given every 10 time units, and the test was 10 time units after the last exposure. Thus, total time is $10n$, where n is the number of exposures. In Figure 3, the total exposure was held constant at 100 and the last study

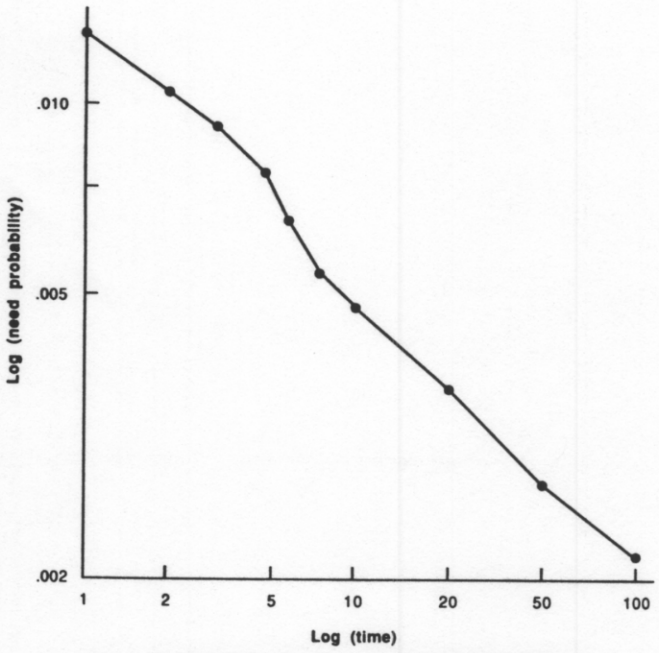

Figure 1. Relationship between need probability and delay since an item has been studied.

was at 80; the remaining studies were placed from 0 to 80 at intervals of $80/(n - 1)$, where n was the number of exposures. (In this case, the minimum number of trials is 2—one at Time 0, one at 80, and then the test at 100.) The functions are plotted on a log–log plot, and again they are linear, implying that need is a power function of use. Such power functions are typically found (Newell & Rosenbloom, 1981). The linear functions in Figures 2 and 3 have slopes close to 1, which is what we would have predicted from Equation 10. Thus, the complications we introduced after deriving Equation 10 have not changed its basic prediction about the practice function.

One might wonder about the match of our theoretical functions for need probability to empirical power functions for speed and accuracy. There has to be some transformation from need probability to these dependent variables. When we examine plausible transformations in later sections, we will see that they basically raise need probabilities to some power (typically, less than 1). Such transformations would preserve the power relationship between the measure and the independent variables of delay and practice.

Until recently, there was only Crossman's (1959) theory that predicted power-law learning. He attributed power-law learning to subjects adjusting their sampling among problem-solving methods. More recently there has been a flurry of theories that are able to predict power-law learning. Newell and Rosenbloom (1981), expanding on ideas in Lewis (1978), attributed power-law learning to chunking macrooperators in exponentially complex problem spaces. MacKay (1982) attributed it to a strengthening process. Anderson (1982) derived it from the power-law

[5] Note that in this section we are assuming that the contextual factor is constant and are just looking at the history factor.

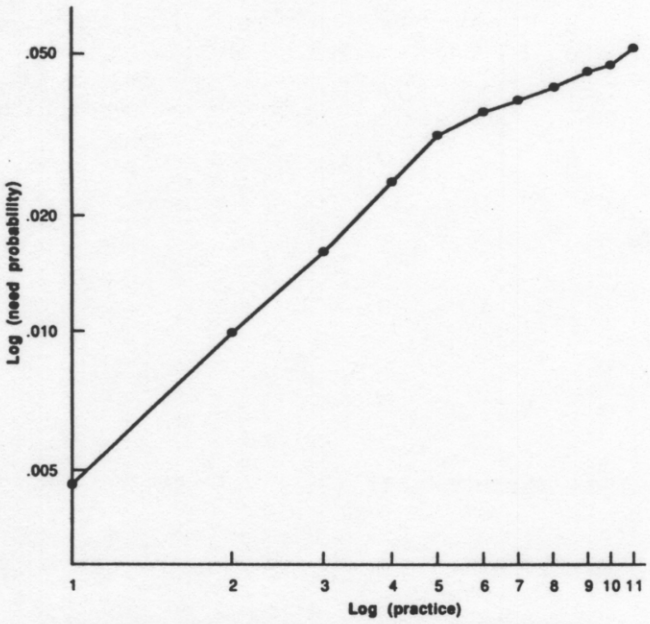

Figure 2. Relationship between practice and need probability. (In Figure 2, spacing between studies is held constant.)

form of strength decay. Shrager, Hogg, and Huberman (1988) produced power-law learning by a procedure that learns new operators and optimizes the decision procedure. Logan (1988) attributed it to a race among retrieval of previous experiences. The rational analysis given here does not necessarily contradict any of these models. They could be interpreted as simply proposing a mechanism to achieve the rational analysis. The wide variety of mechanisms proposed should convince us that there

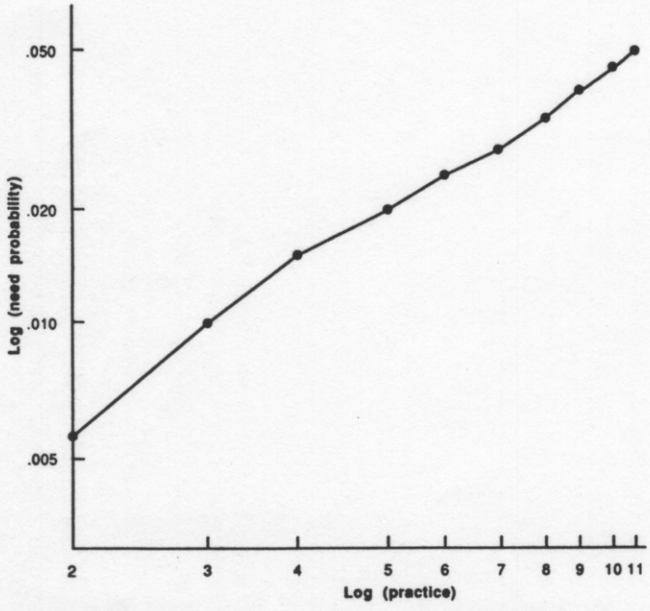

Figure 3. Relationship between practice and need probability. (In Figure 3, time since initial study and time since last study are held constant.)

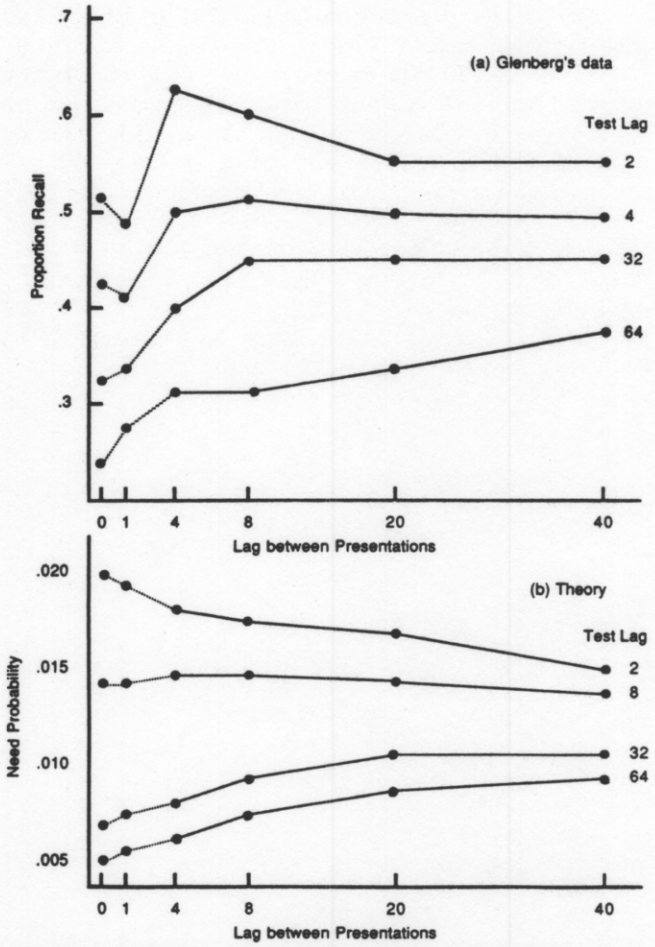

Figure 4. (a) Glenberg's (1976) data and (b) our estimates of need probability. (Glenberg studied the effect of spacing between two studies [the abscissa] as a function of the spacing from second study to test [the different curves].)

is not a unique mechanism to get power-law learning. There is one critical feature that these models lack, however. They focus only on the effect of practice and do not take into account its relationship with spacing and forgetting effects (Anderson does not integrate in spacing). A theory that fails to deal with such fundamental variables of practice cannot really claim to explain practice effects. Indeed, if one tried to integrate directly into these models facts such as the power-law decay with delay or the nearly total ineffectiveness of highly massed training, the models might fail to predict power-law learning.

We now turn to whether our model can predict the spacing effect. This is where Burrell's (1985) model breaks down as being descriptively accurate of information usage. Although usages are not independent, it remains a question of whether our formulation of the process will match the human spacing functions. Some of the richest data on human spacing functions come from Glenberg (1976), who looked at the interaction between the spacing effect and the retention function. That is, he orthogonally manipulated the delay between the two studies and the delay between the second study and test. Figure 4a plots his data, and Figure 4b plots estimated need probability. The

Table 1
Mean Percentage of Recall and Need Probabilities (in Parentheses) for Bahrick

Intercession interval (in days)	Session 2	3	4	5	6	7
	Tested after three training sessions					
0	77 (.0201)	89 (.0298)	33 (.0057)			
1	60 (.0144)	87 (.0192)	64 (.0087)			
30	21 (.0032)	51 (.0069)	72 (.0108)			
	Tested after six training sessions					
0	82 (.0201)	92 (.0298)	96 (.0394)	96 (.0488)	98 (.0576)	68 (.0075)
1	53 (.0144)	86 (.0192)	94 (.0256)	96 (.0329)	98 (.0424)	86 (.0214)
30	21 (.0032)	51 (.0069)	72 (.0108)	79 (.0144)	82 (.0167)	95 (.0183)

Note. The final test was always after 30 days.

unit of time in Glenberg's study was a 3-s presentation of an item. The unit of time on our simulation was 0.33 time units. Thus, one of our time units equals nine of Glenberg's seconds.

Glenberg's (1976) basic result is that at short testing lags, recall decreases with spacing between studies, whereas at long lags it increases. The one exception is that at very short lags (0 and 1 item), there is universally poor performance. This poor performance is typically (e.g., Crowder, 1976) attributed to inattention—something not modeled in the current approach. However, except for this, we do a remarkably good job of reproducing the data of Glenberg. The correlation between his recall probabilities and our need probabilities is .86, which is good considering the peculiarities at 0 and 1 lags and the fact that the two measures are only monotonically related and not linearly related (the rank-order correlation without the 0 and 1 points is .99).

Bahrick (1979) did an interesting experiment in which he studied the interaction between spacing and repetition. He presented subjects with a number of memory trials for 50 English–Spanish vocabulary pairs. The first trial was a study, and the remainder were tests followed by study. There was either 0 delay, a 1-day delay, or a 30-day delay between trials. This was followed by a final test after a 30-day interval. For the subjects who had been studying with 30-day intersession intervals, this was like another test. Bahrick ran two groups of subjects—one that had an initial study session, two test–study sessions, and a final test session, and another that had a study session, five intermediate test–study sessions, and a final test. The results are shown in Table 1 both for performance on the test–study trials after the first and performance on the final test. Bahrick's basic point is that there is an advantage if the retraining interval matches the retention interval.

Bahrick's (1979) experiment was simulated with our system, using 0.01 time-unit intervals to simulate Bahrick's massed condition, 1 time-unit interval to simulate his 1-day delay condition, and 30 time-unit intervals to simulate his 30-day conditions. The results are also shown in Table 1. Certain of Bahrick's six-session data points are replications of the three-session data points. Altogether there are 20 distinct conditions in Table 1. There is a rank-order correlation of .95 between our need probabilities and his percentage recall (averages of two percentages in the case of replicates).

In short, our rational analysis of the historical factor does a remarkably good job of accounting for the effects of frequency and recency of presentation.

The Contextual Factor

The preceding analysis has been concerned with analyzing the history factor, which was the first term in Equation 2. Now we turn to calculating the remaining quantities, $P(i|A)/P(i|\bar{A})$, which are the cue strengths. It is almost certainly the case that $P(i)$ and $P(i|\bar{A})$ are going to be nearly identical, because conditionalization on the nonrelevance of one memory structure out of millions cannot much change the probability of any cue. Thus, our discussion of cue strength focuses on the simpler form of $P(i|A)/P(i)$. As noted earlier, $P(i|A)/P(i) = P(A|i)/P(A)$. The cue strength thus reflects either the degree to which the context element is more or less probable when a trace is needed, or equivalently, the degree to which it is more or less probable that a trace is needed when a context element is present. Intuitively, these cue strengths reflect the degree of association between the terms i and the memory structure A. However, we have not been satisfied with any formal analysis that we have been able to develop of this intuition. One idea is incorporated in the Search of Associative Memory model (SAM; Gillund & Shiffrin, 1984). This is to make these cue strengths reflect the frequency with which i and A have cooccurred in the past. This is a sensible proposal, but it does have two related difficulties: (a) When a memory structure is just created, there is a poor data

base for estimating such strengths, and estimates will fluctuate wildly. (b) This scheme loses information by looking only at direct cooccurrences. If *school* has occurred in the context when other memories about *teacher* are needed, then there is reason to expect that it will occur in the context when a new memory about *teacher* is needed—even if we have not yet experienced that particular cooccurrence.

In Anderson (1989), we proposed a scheme for estimating these conditional ratios based on work from information retrieval (Salton & McGill, 1983). This work avoids the objections listed earlier, although we have to confess that we have never read a deep explanation of why this approach works in information retrieval.

Given that this analysis has been developed elsewhere and that its motivation is weak, we do not repeat it here. Rather, we present a similar analysis, but one that we believe is conceptually clearer: A memory structure A consists of a set of terms a, $b, \ldots, n$. The assumption is that $P(i|\mathrm{A})$, the probability of cue i given that A is a target, can be thought of as the probability that i is present, given that a structure with a, b, $\cdots$ and n is needed. Thus, we may rewrite our cue strength expression

$$\frac{P(i|\mathrm{A})}{P(i)} = \frac{P(i|a, b, \ldots, n)}{P(i)}. \tag{12}$$

The key idea is that the relationship between A and i can be decomposed into a set of relationships between the elements of A and i. Let $n(x)$ denote the probability that a trace with element x is needed to distinguish it from $c(x)$, which is the probability that the context contains element x. Similarly, $n(x|y)$ is the probability that a trace with x is needed, given that y is in the context, whereas $c(y|x)$ is the probability that y is in the context, given that a trace with x is needed. Then, under fairly strong independence assumptions, we may write

$$\frac{P(i|\mathrm{A})}{P(i)} = \prod_{x \in \mathrm{A}} \frac{c(i|x)}{c(i)} = \prod_{x \in \mathrm{A}} \frac{n(x|i)}{n(x)}. \tag{13}$$

One requisite independence assumption is that

$$P(\mathrm{A}) = \prod_{x \in \mathrm{A}} n(x), \tag{14}$$

or that the probability structure A is needed is the product of the probabilities that structures involving the individual components are needed. The other requisite independence assumption is

$$P(x|i \,\&\, \{\mathrm{A} - x\}) = n(x|i), \tag{15}$$

or that the probability that x is needed depends only on i and not on the other elements that are needed. These are strong assumptions, but they are required to make the standard move in the information-retrieval literature of decomposing the connection between a trace (or file) and a cue (or query element) to the connections between the components of the trace (or file) and the cue (or query element).

It is worthwhile to take an overview of what has happened in developing Equation 13. We started out interested in how the set of elements in the context predicted a particular memory trace. The system has little or no experience with this exact pairing and so has no direct basis for making this estimation. Thus, earlier, we decomposed (Equation 2) the set of contextual elements into individual elements, and in Equation 13 we decomposed the trace into its elements. We are now looking at the degree to which a specific element in the context predicts a specific element in the trace. The advantage we get out of these moves is that we have much more experience with the pairings of the individual elements than we have with whole contexts and whole traces.

This still leaves open the issue of how to estimate the $n(x|i)/n(x)$. A simple idea is to base it on frequency in experience. That is, define $n(x|i)$ to be the proportion of times that a trace with x is needed when i is present in the context, and $n(x)$ to be the unconditional proportion of times that a trace with x is needed. This is a reasonable solution in cases of large samples. If x is an established concept, experience would give a good basis for estimating $n(x)$. If, in addition, i is a frequently occurring contextual element, one can accurately estimate $n(x|i)$. However, such estimates would still fluctuate radically in a case of infrequent elements.

Consider the predicament of moving to a new psychology department and meeting Professor a and Professor b. How does one set $n(a|b)/n(a)$, where $n(a|b)$ is the probability that a memory about a will be needed if b is mentioned and $n(a)$ is the base probability that a memory about a will be needed. One should set these probabilities at some default value and adjust with experience. The initial default value should be influenced by one's knowledge. Thus, if a and b are both professors of social psychology, $n(a|b)$ should be set higher than if one is a professor of social psychology and one is a professor of cognitive psychology. Eventually, with enough experience, one would adjust from these initial estimates to estimates that reflect proportions in experience. We are not in possession of a precise model of how to set initial values based on knowledge and how to adjust with experience, so we have to leave our analysis at the informal level.

Frequency Effects

The probabilities $p(x)$ and $n(x)$ would be related to frequency norms, although there is no reason to believe that the relationship will be perfect. Similarly, $n(x|i)$ should be related to free association norms, but again there is no reason to believe that the relationship will be perfect. We can therefore inquire as to whether empirical results involving such norms can be predicted within this framework. Consider word-frequency effects in recognition memory. In this case, the cue is the word and the target is a trace involving the same word. Our ratio becomes[6]

$$\frac{n(\text{word}|\text{word in context})}{n(\text{word})}.$$

The numerator is to be read as the probability that a memory trace involving the word is needed, given that the word is present. Presumably, this conditional probability is relatively high, although not 1. It is probably fairly constant for all words. In

[6] In this and subsequent analyses we are focusing on the effect of one element in the context, assuming the effects of the other elements do not vary.

contrast, the denominator will vary with the frequency of the word. Thus, we predict that low-frequency words will be better recognized, a well-documented result (e.g., Kintsch, 1970). The basic point is that low-frequency words are statistically better predictors of traces involving them than are high-frequency words.

Second, consider the case of paired associates in which we present a pair, word1–word2, for study, and test with word1 for the recall of word2. Our ratio in this case is

$$\frac{n(\text{word1}\,|\,\text{word1 in context})}{n(\text{word1})} \times \frac{n(\text{word2}\,|\,\text{word1 in context})}{n(\text{word2})}.$$

If we assume that words are of equal frequency, the relevant variable becomes $n(\text{word2}\,|\,\text{word1 in context})$, which will vary with the associative strength of the words. This predicts the result (Eich, 1982) that it is easier to learn experimental associations between words with strong prior associations. It is worth noting the rational basis for this effect. It is based on the assumption that new knowledge involving highly interassociated terms is, in fact, more likely to be needed. Thus, for instance, if we are told that Ronald Reagan believes that Howard Baker is dishonest, the prediction is that this is more likely to be a fact that we will want to re-use when we hear Ronald Reagan than if we are told that Ronald Reagan believes that Wayne Gretzky is dishonest—because of the greater interassociation between Reagan and Baker.

An interesting feature of this analysis is that it does not predict any clear effect of response frequency on paired-associate learning—a result that is approximately correct. In a typical experiment looking for word-frequency effects, nouns of different frequency are basically paired randomly. This means that as we increase the frequency of a word, and so $n(\text{word2})$, its prior probability of association, $n(\text{word2}\,|\,\text{word1 in context})$, will, on average, grow proportionately. Thus, the numerators and denominators in these ratios should tend to cancel themselves out. Because of $n(\text{word1})$ in the denominator, the analysis does predict a negative effect of stimulus frequency, a result that is supported in Paivio (1971). Gillund and Shiffrin (1984) failed to find any significant effects in recall, but they did find considerable advantage for low-frequency words in paired-associate recognition.

Finally, we consider a free-recall situation in which the subject is given no cues except the random elements in the environment, which we denote as *context*. Then, the probability of recall is governed by

$$\frac{n(\text{word}\,|\,\text{context})}{n(\text{word})}.$$

In this case, we might assume the frequency of the word, given that the context matches its base frequency, and so we have a ratio of 1 for all words independent of their frequency. However, it is known that high-frequency words are better recalled in free-recall tests (Kintsch, 1970).

One possible explanation of this discrepancy is to relate it to an organizational strategy by which a subject tries to interrelate items in the list. High-frequency words, having more traces involving them, will be more easy to interrelate. In fact, there is considerable evidence that the frequency effect is due to organizational strategy. First, if one designs high-frequency lists without strong interitem associations, the word-frequency effect disappears (Deese, 1960). Second, recall of high-frequency lists shows more subjective organization (Postman, 1970). Third, and most critically, when subjects are given distractor activity during study to prevent them from engaging in an organizational strategy, the word-frequency effect disappears (Gregg, Montgomery, & Caslano, 1980). Thus, it seems that the word-frequency effect for free recall is a strategy effect. In contrast, the word-frequency effect for recognition seems more robust across such manipulations.

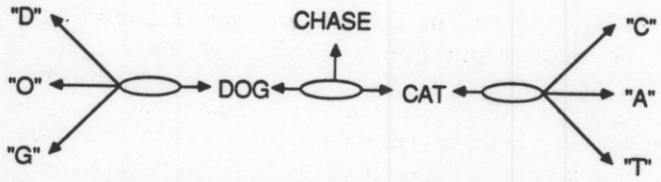

Figure 5. Interassociations relevant to recognizing the spelling of a word like *cat* in the presence of a word like *dog*.

This explanation of the word-frequency effect in free-recall places it outside of our rational model, which by itself clearly fails to predict the phenomena. The fact that it does so clearly should help allay doubts that rational models do not have clear predictions. However, there have to be such things as memory strategies that will effect the behavior of a subject over and above the basic tendencies of human memory. At the end of this article, we discuss how this rational analysis is to be related to the phenomena of memory strategies in free recall.

As a final comment, note that our predictions for word-frequency effects are quite similar to those developed by Gillund and Shiffrin (1984). Indeed, one might look to their model as a reasonable proposal for how such word-frequency effects might be implemented.

Priming Effects

We can relate this analysis fairly directly to the priming literature. Consider a typical experiment in which the subject is asked to judge whether a string of letters like *cat* is a word. This judgment can be more rapidly made when that word is preceded by an associated word like *dog*. This situation is illustrated in Figure 5. *Dog* and *cat* are interassociated through their appearance in common memory structures. To access the word *cat* we have to access information about its spelling. A priming experiment looks at this access as a function of context. In this example the relevant context is the cue term *dog*. The ratio that we are looking at is $P(\text{cat's spelling is needed}\,|\,\text{dog in context})/P(\text{cat's spelling is needed})$. The assumption is that this ratio is greater than 1 in this case because of the interitem association between *dog* and *cat*. Said in other words, our Bayesian estimation procedure increases the probability that the spelling of *cat* will be relevant in the presence of the word *dog*.

In explaining such priming effects, the Bayesian analysis, as currently developed, offers no more predictive power than any of the many competing analyses. However, it adds some explanatory power. It embodies the claim that we can recognize the spelling of *cat* faster in the presence of *dog* because, in actual

fact, there is a higher than average probability that we will have to recognize the spelling of *cat* when *dog* is present.

An interesting feature of this Bayesian analysis is that it predicts the inhibitory priming effects that occur. That is, recognition is worse in the presence of an unrelated stimulus (like *lip*) than it is in the presence of a neutral cue (like *XXX*). These conditional likelihood ratios have to average out to 1 (in a probability definition of average). Thus, because related terms have greater-than-1 ratios, unrelated terms will have less-than-1 ratios. Moreover, it again makes sense. In fact, the odds are lower than chance that we will have to recognize the word *dog* in the presence of an unrelated word like *lip*.

It has been argued that the mechanisms underlying the inhibition effect are different than the mechanisms underlying the facilitation effect (Lorch, Balota, & Stamm, 1986; Neely, 1977). It is argued that early on there is only an automatic facilitation, which gives way to a strategic process that can produce both inhibition and facilitation. The principle evidence for this is the observation that inhibitory effects appear later than do facilitatory effects. This rational analysis does not really take a position on mechanisms and certainly does not deny the possibility that different mechanisms may implement different aspects of rational prescriptions. There is neural evidence that inhibitory processes are slower because their paths involve more synapses (Shepherd, 1979). This may be a case in which the constraints of the brain impact on rational derivations. The differences are not huge—facilitatory effects make themselves known in 100 ms, whereas inhibitory effects take 500 ms. The ideal would be instantaneous priming effects, which is clearly not possible in any physical system.

Finally, this analysis can predict a result that ACT* fails to handle. This is the observation that one cannot seem to get second-order priming. De Groot (1983; see also Balota & Lorch, 1986; Ratcliff & McKoon, 1988) used triplets of words like bull–cow–milk, where there is a strong association between the first and second and between the second and third, but not between the first and third. The first did not prime the third in contrast to what would be predicted by a spreading-activation model in which activation would spread from the first to the second and, hence, to the third. However, on the aforementioned analysis, the first and third terms would have low relatedness. This is, in fact, the rational thing to do: If *milk* is never processed in the presence of *bull,* one should not prime structures involving *milk* when *bull* appears.

The exact status of this result is somewhat in debate. Balota and Lorch (1986) did find second-order priming in a word-naming task but not in a lexical-decision task. Even more recently, McNamara and Altarriba (1988) found evidence for weak, second-order priming more generally. One could argue that, for low-frequency words, second-order priming might reflect a rational estimation procedure. That is, if one has seldom seen A, B, or C, but has seen A and B together and B and C together, it might be reasonable to guess that A and C will occur together. However, the words in these experiments were not low frequency. The relative frequency in experience of B, given A, should have been an adequate basis for estimating the conditional need probability.

If there are second-order priming effects, we think that these effects reflect more the definition of second-order associates. Just because subjects do not give *milk* to *bull* does not mean that *bull* and *milk* are never encountered together. Indeed, Balota and Lorch (1986) reported that subjects rate these as more related than random pairs. Looking over the second-order associates of Balota and Lorch and of McNamara and Altarriba (1988) it seems to us that the probability of the second-order associate is raised by the stimulus—that is, $n(\text{milk}|\text{bull}) > n(\text{milk})$. The basic problem is that free-association norms are only imperfect predictors of the underlying probabilities.

Fan Effects

This analysis also relates fairly directly to fan effects (Anderson, 1983). The fan effect involves manipulating the number of facts in which a particular concept appeared (the fan of the concept). The basic fan result is that a particular fact is more slowly retrieved when the concepts that compose it occur in more other facts. The result can be seen as directly arising from the conditional probability ratio $P(\text{A}|i)/P(\text{A})$. The denominator, the probability of the trace, will be constant for the traces in a particular experiment, whereas the numerator should decrease with the fan of i. As the fan of i increases, it is associated with more traces, and so the probability of any particular trace goes down. This analysis predicts that it is probability of the association and not fan that is the critical variable. It is just that as we increase fan, we decrease probability on the average. Anderson (1976) reported an experiment in which fan was decorrelated from probability by studying different facts about a concept with different frequencies. The clear outcome of that research was that it was probability and not fan that was the controlling variable.

It should also be clear in this analysis why there are fan effects for foils—that is, it takes longer to reject a foil the more sentences were studied about the concepts in the foil. The more facts there are about a concept, the more things will have to be considered before rejecting a fact that involves this concept.

Anderson, 1983, has done a number of fan experiments looking at the effect of the number of cues or terms in the sentence to be recognized. It turns out that it is easy to confound number of cues with the complexity of the memory task. However, when this is avoided, recall increases with number of cues. Each relevant term should increase the odds ratio for the target trace in Equation 2. The current analysis would also predict that retrieval time would be a function of the product of the fans of the individual cues, a prediction that is also generally confirmed (Anderson, 1976). A final prediction is that the fan of existing cues will be attenuated if an additional relevant cue is added. This prediction also has been confirmed (Anderson, 1983).

Relationship of Need Probability to Probability and Latency of Recall

The analysis to date has really been concerned with need probability. That is, we have been developing a theory of how the probability that an item will be needed varies as a function of its history and the cues presented. We have only assumed that it will be related monotonically to the two principal behavioral measures of memory, namely probability of recall and latency in recall. The basic assumption was that a subject would con-

sider items in memory in the order of their need probability until that fell below some threshold. This produces a monotonic relationship between need probability and recall latency. It would seem to define probability of recall as a step function of need probability. Indeed, as much of the research on all-or-none recall has demonstrated, there is a step-functionlike quality to probability of recall so that if an item can be recalled on one occasion, there is a high probability (often near 1) of its being recalled on later occasions, whereas if an item cannot be recalled on one occasion there is low probability (often near 0) that it will not be recalled on later occasions (for reviews see Battig, 1968; Restle & Greeno, 1970). Latency has been shown (Anderson, 1981) to be much more sensitive to presentation variables.

This does not imply, however, that all items with the same experimental history will be recalled or that all will not. There are at least two reasons why we cannot predict a particular item's recall from its history. The first is that we do not know its pre-experimental history. The second is that we are not sure of its experimental history. The subject may not have attended to it during some presentations, and the subject may have rehearsed it covertly during other time periods. There is evidence that we can improve our ability to predict subject's recall if we try to track attention (Loftus, 1972) or if we try to monitor rehearsals (Rundus, 1971).

Analysis of Latency of Recall

We need some theory to relate the probability that an item is needed (which is the quantity that we have analyzed to this point) to the time it will take to retrieve it. In the current analysis, need probability determines the order in which knowledge is examined. Given a particular item with need probability p, we do not necessarily know its exact ordinal position. This will depend on how many higher probability items there are in the situation. However, as an approximation we can assume probabilities are distributed according to Zipf's law. This law has been found to describe such things as distributions of words in prose samples by their frequency of occurrance, distribution of scientists by numbers of articles published, distributions of cities by population, distribution of names by size, and distribution of biological genes by number of species (Ijiri & Simon, 1977). Given such a range of application, it seems only a mild generalization to propose that it describes distribution of memories by their need probability.

A close approximate form for Zipf's law is $f(i) = ri^{-d}$, where i is the ordinal position of the item, f is the measure (count, income, need probability, etc.), d is a constant (often estimated to be 2), and r is a scale factor (Ijiri & Simon, 1977). In the case of need probabilities of memories, we can use the following form

$$p = rt^{-d}, \tag{16}$$

where p is the need probability and t the time it will be retrieved. Inverting this, we get the relationship between time and need probability,

$$t = [r/p]^{1/d},$$

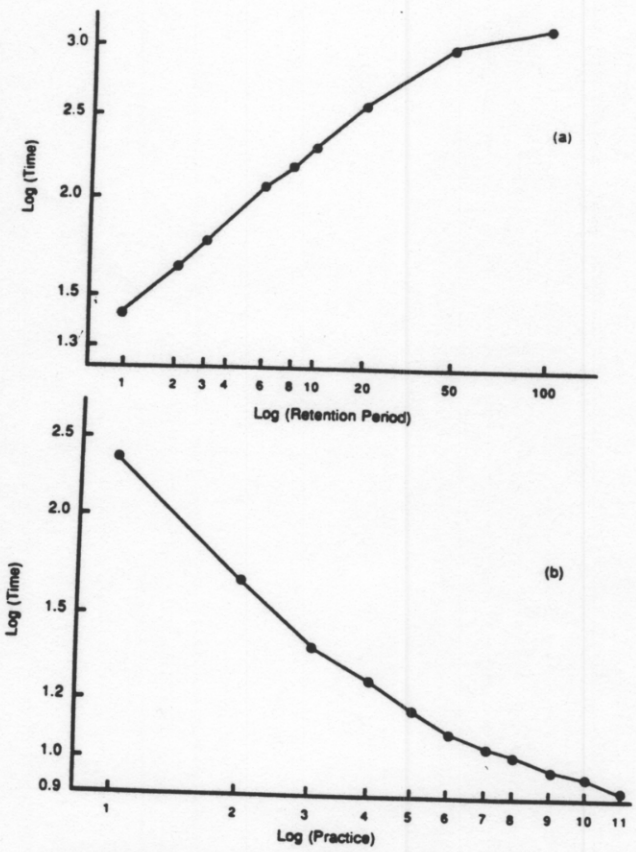

Figure 6. Transformations of Figures 1 and 2 to introduce the dependent measure of response latency.

and adding an intercept gives us a formula for time,

$$T = I + [r/p]^{1/d}. \tag{17}$$

Figures 1 and 2 earlier related need probability to frequency and recency of presentation. Figure 6 shows these functions transformed to give times setting $I = .3$, $r = .02$, and $d = 2$. As can be observed, we get typical practice effects and retention effects when measured by latency. Note that the relationship between the slopes of the functions in Figures 1 and 2 and the slopes of the functions in Figure 6 is basically $1/d$.

Analysis of Probability of Recall

In the current theory, memory fails to recall an experience because the need probability for that memory is below the threshold that the system is willing to consider. As we mentioned earlier, one might expect a perfect step function from probability of 0 to 1 as the threshold is crossed. However, it is quite possible that the threshold will vary depending on the situation. It is also possible that the actual evidence will vary from our estimate, depending on past history, lapses of attention, hidden rehearsals, and so forth, that we cannot observe. It is also possible, all consideration of rationality aside, that there is some noise in the system. We can summarize all of this by saying that there is some variance in our estimate of the distance between the evidence and the threshold and the actual distance.

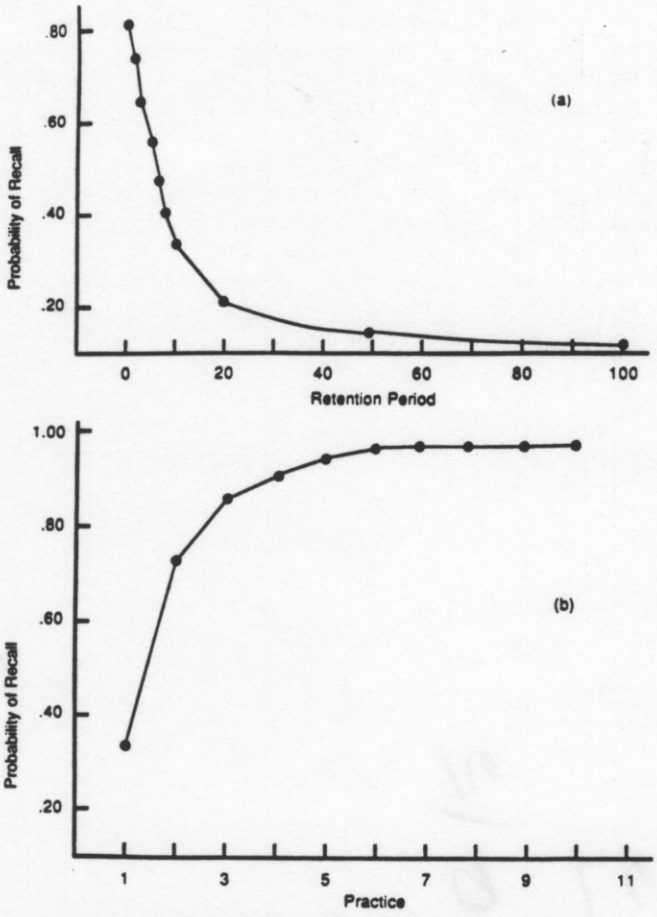

Figure 7. Transformation of Figures 1 and 2 to introduce the dependent measure of probability of recall.

The natural distribution for this variance would be a normal distribution, but for purposes of analytic tractability we will replace it by the similar logistic distribution. Formally, our assumption will be that the distance between the log odds estimated, v, and the threshold θ, is distributed as a logistic function with parameter s. (The parameter s is related to the variance—variance = $\pi^2 s^2/3$.) Then, the probability that a memory with log need, v, will be recalled is

$$P = \frac{1}{1 + e^{(\theta - v)/s}}. \tag{18}$$

This implies that the relationship between the odds of recalling an item and the need odds, $O = e^v$, is

$$\frac{P}{1-P} = AO^{1/s}, \tag{19}$$

where $A = e^{-\theta/s}$.

Figure 7 shows transformations of the retention and practice functions (Figures 1 & 2) setting $\theta = -5$ and $s = .5$. This plots probability of recall, delay, and practice untransformed. However, if we were to take log transforms and plot odds recall rather than probability of recall, we would get linear function with slopes $1/s$ of what they are in Figures 1 and 2.

Effects of Subject Strategy

So far we have ignored any systematic analysis of possible effects of subject strategy on memory performance. However, it is well documented that subjects engage in numerous strategies for processing to-be-learned material, and these strategies can substantially affect their memory performance. Moreover, the evidence is that such strategies are acquired, and younger children may reflect different memory performance simply as a function of their different memory strategies (Flavell, 1977).

The question naturally arises as to how one is to conceive of memory strategy in a rational analysis. One attitude would be to ignore its role and assume that memory strategy is just part of the black box that is being optimized for memory performance. However, the radically different behavior that can occur as a consequence of strategy choice causes fundamental problems for this approach. For instance, in the same experiment, one subject will choose to repeat the items over and over again in a rote fashion, whereas another subject will engage in an elaboration strategy and enjoy a much better memory performance as a consequence. How can one argue that both subjects are engaging optimal behavior with respect to optimizing their memory performance to the same environment?

We think a better way to conceive of this is that subjects are manipulating the information that is presented to human memory by their choice of strategy. Given different strategy-determined experiences, memory is behaving optimally in response to those experiences. Thus, for instance, given multiple redundant traces created by an elaboration strategy, memory has more traces to call upon and more interassociated traces. Thus, the advantage of elaborations is to be understood in terms of the same redundancy analysis that has been given for the ACT* theory (Anderson, 1983).

Our view is that subjects can essentially manipulate the input to their memories and that human memory, blind to the intentions of the subject and to the fact of a deliberate manipulation, behaves as rationally as it can, given the statistics of the input it receives. This leaves open the question of the rationality of subjects' strategy choice—that is, whether their manipulations are optimal by some criteria. We do not address this question here.

The consequence of this approach is (a) that understanding the details of human memory performance in many circumstances will require that we specify subjects' memory strategies and (b) that we cannot simply derive their behavior from an analysis of the information the experimenter is presenting to them. This is because these strategies are intervening between the experimenter and their memories and transforming the information presentation. Indeed, in some situations, the subjects' memory performance will be more a function of strategy choice than of any direct properties of the experimental manipulation. In the next section, we give one token of this in simulating the traditional free-recall experiment.

Simulation of Free Recall

A basic observation about subjects in a free-recall experiment is that they often covertly rehearse items in addition to the item currently being presented. Modeling this particular phenomena

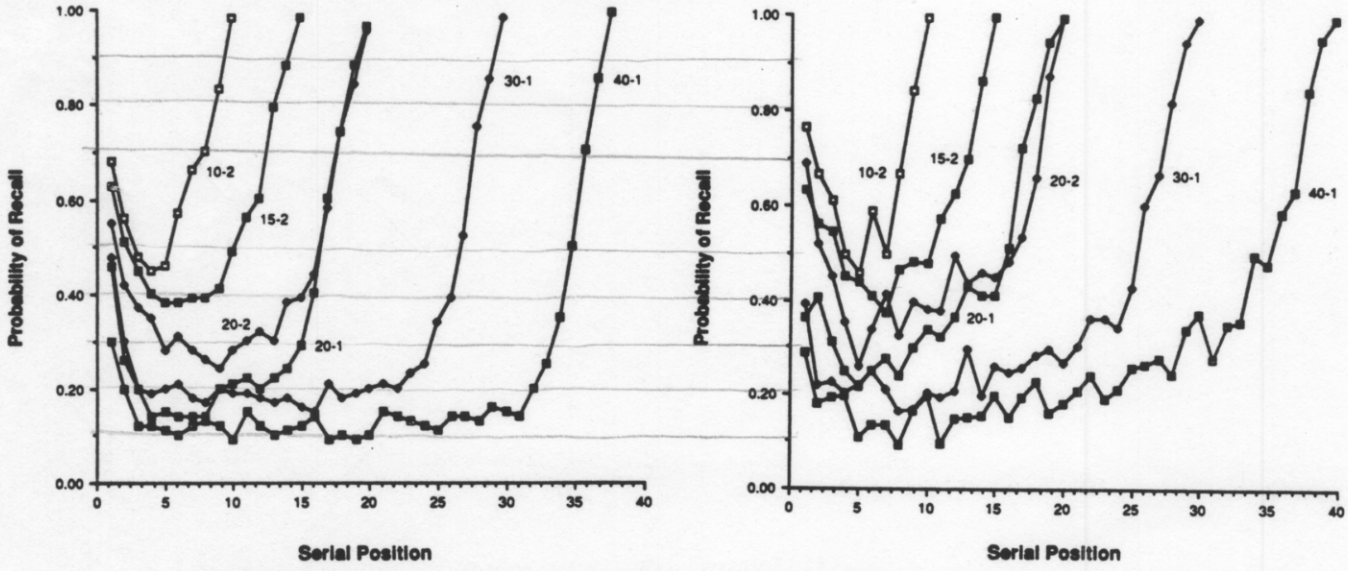

Figure 8. Serial-position effects for lists of various lengths and rates of presentation (seconds per word). (Part a is Murdock's 1962 data, and Part b is our simulation.)

has been part of a good many models of memory, including the original Atkinson and Shiffrin (1968) model, and was incorporated into the Free Recall in an Associative Net model (FRAN; Anderson, 1972). It has also found its way into the more recent SAM (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981) model of free recall. We decided to incorporate a simple version of this rehearsal process, not unlike the SAM model. We assumed that subjects could maintain a buffer of four items for rehearsal. Each time an item was presented, it entered the buffer and, if the buffer was full, an existing item from the buffer was thrown out. The choice of which item to throw out was made randomly, with each item equally likely. When an item was first presented to the subject, it was encoded by memory. Every second it was in the buffer it had a .2 chance of being encoded anew. The benefit of residing in the buffer was these opportunities for further encodings.

We assumed that the memory system responded to each encoding of the item as a new use, whether that encoding came from experimenter presentation or buffer residence. Upon completion of the study, we assumed that the subject would recall items with need probabilities above a certain threshold. This is an obvious simplification in that subjects adopt various strategies for organizing their recall as well as their study.

To map need probability onto probability of recall requires a description of the role that noise plays in converting from estimated need probabilities to probabilities of recall. A preceding section, Analysis of Probability of Recall, provides the theoretical discussion of a noise factor, and the actual transformation from need probabilities to probabilities of recall is given by Equation 18. A fuller description of the free-recall model is given in the Appendix.

This model undoubtedly underestimates the complexity of the memory strategies that are actually occurring in a free-recall experiment. The reader is invited to read the Appendix to Anderson (1972) to see how rich free-recall strategies can really be. However, this gives us a first-order approximation from which we can try to predict some of the basic statistics of recall.

Serial-Position Effects

A classic datum from free recall concerns the serial-position function—how probability of recall varies as a function of serial position in the study list. Murdock's (1962) data is shown in Figure 8a, and our simulation of it is shown in Figure 8b. The correspondence is quite good. The basic features of these data are the strong recency effect in which the recall drops off from the end, and the lesser primacy effect in which recall drops off from the beginning and from the flat region between. The recency effect is produced by the decay in need probability, and the primacy effect is produced because the first items in the buffer have an advantage. It is not obvious that we would get the right orders of magnitude among these three regions of the serial-position curve.

The primacy effect deserves a little comment. It is produced directly by the assumption of a buffer model and by the fact that it takes a few items for the buffer to fill up, and so the first items are not pushed out right away. Thus, in contrast to the recency effect, which is a result of the rational model, the primacy effect is a consequence of strategy choice. One might wonder whether this attribution of the primacy effect is correct. Perhaps the first things in a new context tend to repeat more often or are more important (i.e., perhaps there is a rational explanation of the primacy effect). Certainly, prose (especially newspaper stories) tends to be structured with the important things first, but one can view this as writing adapting to the primacy effect rather than as the cause of the primacy effect. Outside of human communication, it is unclear whether there is any validity to the idea that first things are more important. The analogy in the library system would be something like the first borrowings of the day identifying the more often borrowed books. In a file system it

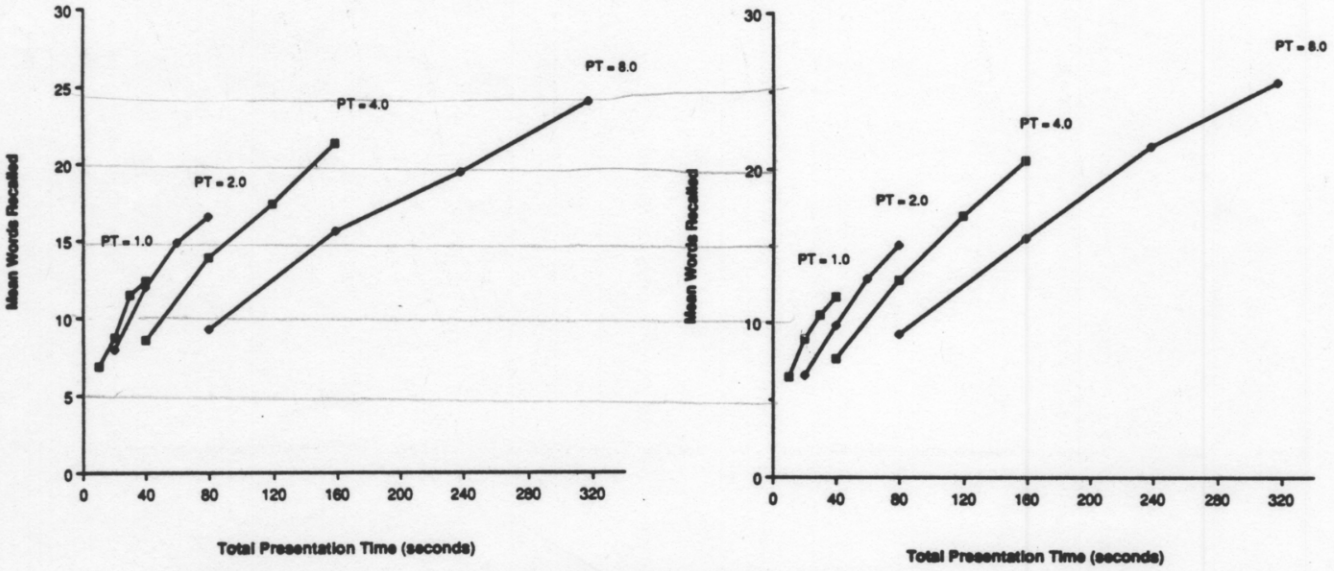

Figure 9. Mean number of words recalled as a function of total presentation time (abscissa) and presentation time (PT) per item. (Part a is Roberts's 1972 data, and Part b is our simulation.)

would be the first files accessed in the day being more important. It is unknown whether these effects exist in such systems, and even if they do they might exist for reasons that do not seem to apply for the human system. For instance, borrowers might be lining up outside the door of the library to get the book that that night's newscast made critical.

On the other hand, recent evidence (Baddeley, 1986; Glenberg et al., 1980) has suggested that the primacy effect in free recall may largely be due to the strategy of rehearsal. Manipulations that discourage rehearsal largely eliminate the primacy effect, whereas they leave the recency effect unaffected. Thus, the current analysis may be correct in its attribution of the recency effect to rational factors and the primacy effect to rehearsal strategy factors.

Inspection of the simulation in Figure 8b reveals that the curves do not really flatten in contrast to the empirical curves in Figure 8a. This is because there is a retention effect throughout the serial positions, and earlier positions suffer because of their greater lag from test. We suspect that we could make these curves indistinguishable from the data by suitably playing with the parameters. Nonetheless, we think that Figure 8b is more informative because it reveals the prediction of a continually decaying serial-position curve until the primacy portion. We would predict that if the empirical serial-position curve were long enough (i.e., stretched out over enough time), the decay would be apparent also.

Effects of Presentation Time and List Length

Roberts (1972) reported an experiment that systematically manipulated list length and study time per item. Figure 9a reports Roberts's data for list lengths of 10, 20, 30, and 40 items and for presentation rates of 1, 2, 4, or 8 s per item. Figure 9b reports the predictions derived from our theory. Both figures plot recall as a function of total study time, and the simulation can be observed to do a good job. An interesting fact to note is that both Roberts's data and simulation data confirm the phenomenon that for the same total presentation time, recall increases as list length becomes larger (and presentation time decreases). This is not at all an obvious prediction of our theory. On the one hand, the longer the list of presented items, the more things there will be to choose from, but on the other hand, the lower will be their mean need probability (presentation time decreases for each item). Apparently, the former factor overwhelms the latter.

Conclusions

This article began with a effort to derive some of the most robust results in human memory from a rational analysis. Many of the results dropped out of an analysis that assumed that subjects were responding to the objective statistics of information presentation. However, our analysis of free recall showed potential effects of subject strategy. This means that any particular memory phenomenon is going to be a joint function of two factors—general properties of memory, which we have argued are rationally determined, and specific strategies adopted by the subject to process the information in that situation. In general, the effect of one memory strategy versus another is to shift the relative need probabilities for different memories, making some more available at the expense of others. A strategy can make experimental traces more available than extra-experimental traces, it can make elaborative traces more available at the expense of verbatim traces, it can make items in the beginning of the serial position more available at the expense of later items, and so forth. From this point of view, there is no better strategy in general but just a better strategy for certain purposes. This is a perspective like the concept of transfer-appropriate processing (Bransford, Franks, Morris, & Stein, 1979). As we said earlier, this leaves open the possibility that

there may be some framework in which the strategy choices might be rationally determined. However, a precise formulation of that framework eludes us.

References

Anderson, J. R. (1972). FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. V, pp. 315–378). New York: Academic Press.

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1981). Interference: The relationship between response latency and accuracy. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 311–325.

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

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1989). A rational analysis of human memory. In H. L. Roediger & F. I. M. Craik (Eds.), Varieties of memory and consciousness (pp. 195–210). Hillsdale, NJ: Erlbaum.

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation* (pp. 90–197). New York: Academic Press.

Baddeley, A. (1986). *Working memory.* Oxford, England: Oxford University Press.

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

Balota, D., & Lorch, R. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 336–345.

Battig, W. F. (1968). Paired-associate learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 146–171). Englewood Cliffs, NJ: Prentice-Hall.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analyses.* New York: Springer-Verlag.

Bjork, E. L., & Bjork, R. A. (1988). On the adaptive aspects of retrieval failure in autobiographical memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory II.* London: Wiley.

Bookstein, A., & Swanson, D. R. (1974). Probabilistic models for automatic indexing. *Journal of the ASIS, 25,* 312–318.

Bookstein, A., & Swanson, D. R. (1975). A decision theoretic foundation for indexing. *Journal of the ASIS, 26,* 45–50.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. Cermak & F. Craik (Eds.), *Levels of processing in human memory* (pp. 331–354). Hillsdale, NJ: Erlbaum.

Burrell, Q. L. (1980). A simple stochastic model for library loans. *Journal of Documentation, 36,* 115–132.

Burrell, Q. L. (1985). A note on aging on a library circulation model. *Journal of Documentation, 41,* 100–115.

Burrell, Q. L., & Cane, V. R. (1982). The analysis of library data. *Journal of the Royal Statistical Society, Series A*(145), 439–471.

Buschke, H. (1974). Spontaneous remembering after recall failure. *Science, 184,* 579–581.

Crossman, E. R. F. (1959). A theory of the acquisition of speed-skill. *Ergonomics, 2,* 153–166.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Deese, J. (1960). Frequency of usage and number of words on free recall: The role of association. *Psychological Reports, 7,* 337–344.

de Groot, A. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22,* 417–436.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Flavell, J. H. (1977). *Cognitive development.* Englewood Cliffs, NJ: Prentice-Hall.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

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

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

Gregg, V. H., Montgomery, D. C., & Caslano, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior, 19,* 240–245.

Ijiri, Y., & Simon, H. A. (1977). *Skew distributions and the sizes of business firms.* Amsterdam: North Holland.

Jones, W. P. (1987). 'As we may think'?: Psychological considerations in the design of a personal filing system. In R. Guindon (Ed.), *Cognitive science and its application for human/computer interaction.* Hillsdale, NJ: Erlbaum.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 333–374). New York: Academic Press.

Lewis, C. H. (1978). *Production system models of practice effects.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology, 3,* 525–551.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Lorch, R. F., Balota, D. A., & Stamm, E. G. (1986). Locus of inhibition effects in the priming of lexical decisions: Pre- or post-lexical access? *Memory & Cognition, 14,* 95–103.

MacKay, D. G. (1982). The problem of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review, 89,* 483–506.

McNamara, T. P., & Altarriba, T. P. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language, 27,* 545–559.

Murdock, Bennet, B., Jr. (1962). The serial-position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482–488.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, & Winston.

Postman, L. (1970). Effects of word frequency on acquisition and retention under conditions of free-recall learning. *Quarterly Journal of Experimental Psychology, 22,* 185–195.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

Restle, F., & Greeno, J. G. (1970). *Introduction to mathematical psychology.* Reading, MA: Addison-Wesley.

Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypothesis. *Journal of Experimental Psychology, 92,* 365–372.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77.

Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval.* New York: McGraw-Hill.

Shepherd, G. M. (1979). *The synaptic organization of the brain.* New York: Oxford University Press.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review, 94,* 439–454.

Shrager, J. C., Hogg, T., & Huberman, B. A. (1988). A dynamical theory of the power-law learning in problem-solving. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 468–474). Hillsdale, NJ: Erlbaum.

Stritter, E. P. (1977). *File migration.* Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133–186). Hillsdale, NJ: Erlbaum.

Appendix

In this Appendix we will specify the algorithm used to estimate the historical component of the need probability, $P(A|H_A)$, and the algorithm used to calculate our free-recall predictions.

Estimating $P(A|H_A)$

Let T denote the time elapsed since an item's original occurrence—that is, the amount of time the item has been under observation. Let $R^* = (R_1, R_2, \ldots, R_n)$ be the retrieval history of the item, where R_i is the time between the item's original occurrence and the ith retrieval of that item. We will have occasion to refer to the original occurrence of the item as a retrieval. We will denote this by R_0, and it is always equal to 0. By convention, $R_i < R_{i+1}$ for all i.

Let $P(R^*, T)$ be the probability that a retrieval history R^* will occur between time 0 and T. This is infinitesimal for all R^*, except for the null retrieval history when $R^* = (\ )$. Let RT^* denote the same retrieval history as R^*, except there is an additional retrieval at T. To calculate $P(A|H_A)$, we need to calculate $P(RT^*, T)/P(R^*, T)$. Essentially, this ratio represents the conditional probability that a retrieval will occur at time T, given that retrievals have already occurred at times $R_1, R_2, \ldots, R_n$. The ratio actually gives us a retrieval rate, but as was discussed in the main part of the article, this is equivalent to a probability if we are defining the rate with respect to a small-enough time unit.

We will describe an algorithm that will compute quantities like $P(R^*, T)$ and $P(RT^*, T)$. One can think of this algorithm as separately applying to calculate $P(R^*, T)$ and $P(RT^*, T)$ in the ratio $P(R^*, T)/P(RT^*, T)$. However, our program calculates them simultaneously as a computational efficiency. This program takes a set of retrieval times that defines an R^* and a recall time T. It calculates an estimate of this ratio that is an estimate of $P(A|H_A)$ at time T. This program is implemented in *C* and can be obtained by writing to the authors.

Computation of $P(R^*, T)$

Before describing the algorithm, we need to derive a critical equation that the algorithm uses. This equation is going to give us $P(R^*, T)$ conditional on a set of parameters, and then our algorithm is going to integrate this over the possible values of the parameters.

Our augmented Burrell model stipulates that retrieval for a specific item is a nonhomogeneous Poisson process with parameters $\lambda, d, r_1, r_2, \ldots, r_m$, where λ is the initial desirability of the item, d is the item's decay rate, and the r_i are times when the item has undergone revivals. Because items get introduced at full desirability, we can for convenience assume a revival at time 0 and define $r_0 = 0$. By convention, $r_i < r_{i+1}$ for all i.

The intensity of the retrieval process is a function:

$$I(t|\lambda, d, r_1, r_2, \ldots, v_m) = \lambda e^{-d(t-r')}, \tag{A1}$$

where $r' = \max\{r_i : r_i < t\}$—that is, r' is the most recent revival before t. The expected number of uses before time T, given the desirability, decay, and revivals is

$$M(T|\lambda, d, r_1, r_2, \ldots, r_m) = \int_0^T I(x|\lambda, d, r_1, r_2, \ldots, r_m)dx, \tag{A2}$$

which has the solution:

$$M(T|\lambda, d, r_1, r_2, \ldots, r_m) = \frac{\lambda}{d}\sum_{j=0}^{j=m}(1 - e^{(-d(r_{j+1}-r_j)}), \tag{A3}$$

where $r_0 = 0$, and $r_{m+1} = T$.

Given Equations A2 and A3, it is possible to calculate the probability of a retrieval history R^* occurring between time 0 and T:

$$P(R^*, T|\lambda, d, r_1, r_2, \ldots, r_m) = [\prod_{i=1}^{n} I(R_i|\lambda, d, r_1, r_2, \ldots, r_m)]e^{-M(T|\lambda,d,r_1,r_2,\ldots,r_m)}, \tag{A4}$$

where $R^* = (R_1, R_2, \ldots, R_n)$. Each term in the product reflects the probability that a retrieval will occur at time R_i. $P(R^*, T)$ is calculated by integrating the preceding expression over the distributions of the retrieval intensity parameters $\lambda, d, r_1, r_2, \ldots, r_m$. We can integrate over λ using Equation 6, which gives the distribution of λ:

$$P(R^*, T|d, r_1, r_2, \ldots, r_m) = \int_0^\infty P(R^*, T|\lambda, d, r_1, r_2, \ldots, r_m)\pi(\lambda)d\lambda, \tag{A5}$$

which has the solution:

$$P(R^*, T|d, r_1, r_2, \ldots, r_m) = \frac{b^v(m+v)!}{n!D^{m+v+1}}\prod_{i=1}^{n} e^{-d(R_i - R_i')}, \tag{A6}$$

where $R_i' = \max(r_j | r_j < R_i)$, and where

$$D = b + \frac{1}{d}\sum_{j=0}^{m} 1 - e^{-d(r_{j+1}-r_j)}.$$

Our expression $P(R^*, T|d, r_1, r_2, \ldots, r_m)$ is still dependent on the decay rate and revival history. What we would like to do is to integrate over decay rate and revival history, but such an explicit integration is impossible given the complexity of the formula. Furthermore, standard numeric integration techniques fail to apply. So we use Monte Carlo integration. This involves randomly choosing $d, r_1, r_2, \ldots, r_m$ from the appropriate distributions, computing $P(R^*, T|r_1, r_2, \ldots, r_m)$, summing the result into a counter S, and repeating the procedure. At the end of N iterations, the value of the integral is approximated by S/N. The steps that occur in each iteration follow:

1. Randomly generate d. This is done by selecting a random variable

X from a uniform distribution between 0 and 1. Because decays are exponentially distributed, we can transform X into a random variable d distributed exponentially by the formula $d = -ln(X)/\alpha$.

2. Randomly generate a set of revivals $r^* = (r_1, r_2, \ldots, r_m)$. We can use the fact that interrevival times are exponentially distributed. The revival set is generated by the following algorithm:

Tot←0
R^*←()
L1: Generate X as a uniform random varible between 0 and 1
Tot←Tot $- ln(X)/\beta$
if (Tot > T) stop
R^*←R^* U (Tot)
goto L1.

3. Calculate $P(R^*, T|d, r_1, r_2, \ldots, r_m)$ according to Equation A6.

Free-Recall Algorithm

The free-recall algorithm works with the following parameters:

α, β, b, v—necessary for need probability computation
N—number of items being presented
t—the interpresentation time
t'—the rehearsal opportunity time step
p_r—the probability of rehearsal when in buffer
z—buffer size
θ—the recall threshhold
s—the variance of the recall noise function

We run a number of simulated free-recall subjects and then average the results. Each run of a simulated subject involves (a) generating a rehearsal pattern, (b) calculating a need probability for each item, and (c) calculating a probability of recall for each item. We describe these three steps in order.

Generation of Rehearsal Pattern

At time $t \times k$ ($k = 0$ to $N - 1$), item $k + 1$ is introduced into the buffer. If the buffer has an empty slot, the item is placed there. Otherwise, a slot is picked randomly and the new item is placed there, bumping the item that was in that slot previously. Hence, the contents of the buffer change every t time units. Every t' time units, the contents of the buffer are checked and every item in it has a probability p_r of being rehearsed. This generates a set R^* of rehearsal times.

Calculation of Need Probabilities

The sets R^* of rehearsal times generated in the previous step can be treated as sets of retrieval times measured from the time of initial presentation of the item. The other quantity required to calculate need probability is the time of recall, which is also measured from the initial presentation of the item. So, given an item presented at Time 18, rehearsed at Times 20 and 24, and tested at Time 60, we would calculate $P(RT^*, T)/P(R^*, T)$, where $R^* = (2, 6)$, $RT^* = (2, 6, 42)$, and $T = 42$. We calculated this ratio by the program described in the previous section of the Appendix.

Calculation of Recall Probability

Given a need probability p, define the need odds $O = p/(1 - p)$ and log need odds $v = ln(O)$. According to the noisy threshhold model discussed in the Analysis of Probability of Recall section, an item with need odds v has the following probability of recall:

$$P(\text{recall}) = \frac{1}{1 + e^{(\theta - v)/s}}. \tag{A7}$$

These probabilities of recall are averaged over iterations of the algorithm to give serial-position curves as in Figure 8.

Received September 22, 1988
Revision received March 10, 1989
Accepted April 4, 1989 ■