COGNITIVE PSYCHOLOGY **32,** 219–250 (1997)
ARTICLE NO. CG970652

The Role of Process in the Rational Analysis of Memory

LAEL J. SCHOOLER

Indiana University

and

JOHN R. ANDERSON

Carnegie Mellon University

The rational analysis of memory (Anderson, 1990) proposes that memory's sensitivity to statistical structure in the environment enables it to optimally estimate the odds that a memory trace will be needed. We have analyzed sources of informational demand in the environment: speech to children and word usage in the front page headlines of the New York Times. In a previous paper (Anderson & Schooler, 1991) we have shown that factors that govern memory performance, including recency, also predict the odds that an item (e.g., a word) will be encountered. In the present paper we develop the theory to make precise predictions about how the odds of encountering an item now varies as a joint function of (1) the statistical associations between the item and elements of the current context and (2) how long it has been since the item was last encountered. The prediction was confirmed environmentally by analyses of the New York Times and speech to children. The corresponding behavioral prediction was tested, using a cued recall task in which the cues were either strongly associated or unassociated to the targets. In contrast to the environmental results, recall performance is more sensitive to the length of the retention interval in the presence of unassociated cues than in the presence of associated cues. Further modeling shows that incorporating estimates of the influence of non-retrieval processes (e.g., reading a word, deciding to respond, etc.) on overall performance reduces the discrepancy between the theoretical predictions and the observed data. © 1997 Academic Press

This research was supported by NSRA Fellowship 1F32HD/MHC7787-01A1, and an NSF/NATO Fellowship to Schooler, and ONR-N00014-90-J-1489 and NSF 92-08524 to Anderson. We thank William Estes, Marsha Lovett, Bennet Murdock, Carmi Schooler, Jonathan Schooler, Richard Shiffrin, John Wixted, and an anonymous reviewer for their comments on this paper. The first author thanks the members of his dissertation committee (which the second author chaired), James McClelland and Lynne Reder.

Address correspondence and reprint requests to Lael Schooler, Department of Psychology, Pennsylvania State University, Moore Building, University Park, Pa 16802.

0010-0285/97 $25.00
Copyright © 1997 by Academic Press
All rights of reproduction in any form reserved.

INTRODUCTION

Estes's (1955) goal for his Stimulus Fluctuation Theory was ''to shift the burden of explanation from hypothesized processes in the organism to statistical properties of environmental events.'' We have revisited Estes's basic hypothesis that memory is sensitive to the patterns with which environmental stimuli occur and reoccur. Our work in this area builds on Estes's initial insights by combining information retrieval theory (Anderson, 1990), with empirical studies of the statistical structure of the environment (Anderson & Schooler, 1991; Schooler, 1993).

The rational analysis of memory (Anderson, 1990) characterizes the function of memory as a database retrieval problem. In this analysis, the elements in the current context are treated as constituting a query to long term memory. The claim is that the memory system responds optimally to this query by retrieving those memories that are most likely to be relevant in the current situation. Few would disagree with the claim that the cognitive system retrieves relevant memories. It is the assertion that the retrieval processes are optimal that runs against intuition. Anderson (1990) did not suggest that these processes are perfect; no system could be given the dual constraints of limited resources and a non-deterministic environment. Rather the retrieval processes do as well as can be done, given these constraints.

Like Estes's (1955) Stimulus Fluctuation Theory, the goal of the pure rational analysis approach is to provide accounts for behavior with as few assumptions as possible about internal representations and processes. This paper can be viewed as an effort to see how far we can take that approach in the case of human memory. To foreshadow, we will find that while the rational analysis of memory yields behavioral predictions that are approximately correct, we need to relax our process assumptions to account for the details of our behavioral results.

REVIEW OF RATIONAL ANALYSIS

A central hypothesis of the rational approach is that human cognition optimally solves the problems that it faces. This means that if we can find the problem a cognitive system is trying to solve and find the optimal solution to this problem, then the rational analysis makes the strong prediction that the behavior of the system will correspond in form to this solution. When applied to memory, this implies a program of research which involves a strategy of: (a) specifying a minimal cognitive architecture; (b) identifying the functional role of memory retrieval in this architecture; (c) deriving prescriptions for how an optimal system would fill this role, given reasonable computational constraints; and (d) comparing the implications of this optimal solution to human behavior. We start with an outline of a rudimentary architecture that will clarify terms that are necessary for the subsequent analysis.

The basic architectural assumptions made by the rational analysis are quite minimal. Perhaps the most significant is that information is stored in discrete memory structures, or traces, in long term memory. There are any number of memory theories which share this assumption, including images in SAM (Raaijmakers & Shiffrin, 1981), the chunk structure of ACT-R (Anderson, 1993), the propositional structures of the Construction Integration Model (Kintsch & Van-Dijk, 1978), and the traces of MINERVA (Hintzman, 1988). For other models, such as TODAM (Murdock, 1982) and CHARM (Metcalfe, 1982), the representations are distributed and overlapping.

As in SAM (Raaijmakers & Shiffrin, 1981), the basic idea is that recall depends on search through the structures stored in long-term memory. We assume that: (a) there are limits on the capacity of the memory system to retrieve structures relevant to the current context; (b) the system tries to find the one structure that is needed most in the current context. We will call it the target structure; (c) the memory system estimates the probability that each structure will be needed to achieve processing goals of the system. This probability will be termed need probability for convenience, and can also be thought of as the probability that a particular memory structure is the sought after target structure; (d) the system performs a best first search considering structures in order of their need probabilities until the need probability of the next structure to be considered is so low that it is no longer worth considering; (e) the system correctly recognizes when the target structure has been retrieved; and (f) the system correctly rejects those structures that are not needed. The retrieval search is limited in capacity, processes structures in serial order, and is self-terminating.

In the next two sections we show how the simple search architecture described above can make predictions about recall, when combined with the central hypothesis that memory retrieval depends on need probability. In particular, we derive predictions for how likely the system is to find the needed target structure (probability of correct recall) and how long it will take to retrieve the structure (latency of recall).

Mapping Need Probability to Probability of Recall

If we let p be need probability, C be the cost of considering a memory, and G be the gain associated with successfully finding the target, the system should stop considering structures when

$$pG < C. \tag{1}$$

Much of the discussion in this paper will be in terms of odds rather than probability. If we let $n = p/(1 - p)$ be the odds that a structure will be needed (the need odds), the corresponding equation to 1 is

$$n < \frac{C}{G - C}\,. \tag{2}$$

In the simple architecture used here, Eq. (2) can be thought of as a recall criterion, because when the system considers a structure it always recognizes the target (assumption e) and rejects those structures which are not needed (assumption f). Equation 2 implies that a step function should relate a target structure's need odds to the odds that it will be retrieved. Target structures with need odds below $C/(G - C)$ will never be retrieved and those above will always be retrieved. It is unlikely, however, that a biological system would demonstrate such sharp behavior. There is bound to be noise injected into the system. For example, C and G might vary with the situation, which would shift the retrieval criterion. So the observed recall criterion would be distributed about $C/(G - C)$. If there were many factors responsible for noise in the system, then it would be natural to model this noise with a normal distribution. Anderson (1990) chose to work with the similar, but more analytically tractable logistic distribution. The domains of the logistic and normal distributions lie between negative and positive infinity. The recall criterion discussed earlier is measured in terms of odds that vary from 0 to positive infinity, which does not coincide with domain of the logistic and normal distributions. Log odds, however, does have the desired domain, varying from negative to positive infinity. Therefore, it is more convenient to use log need odds. Figure 1A illustrates a logistic distribution with standard deviation of 1.81 centered around -5. Note that $-5 = \ln(.0067)$ implying that the recall threshold, $C/(G - C)$, corresponds to a need odds of $n = .0067$.

The area under the curve to the left of any point is the probability that x exceeds the observed retrieval criterion. Suppose that the need odds n of a memory structure S is .05. The area to the left of ln(.05) (i.e., -3) is the probability that .05 exceeds the observed recall criterion and therefore that the structure would be retrieved. This area, shaded in Fig. 1A, equals .88. More generally, Fig. 1b plots the probability that any value of $ln(n)$ exceeds the retrieval criterion, which in turn is the probability that a structure with need odds n will be retrieved.

Anderson (1990) (reviewed here in the Appendix) showed that the assumptions described above imply that the odds of a structure, with need odds n, being above threshold is

$$\text{Odds of Recall} = Jn^{f}, \tag{3}$$

where $J = (C/(G - C))^{-f}$ and f is related to the variance of the criterion distribution ($f = \pi/(\sqrt{3}\sigma)$). In words, Equation 3 says that in an optimal system recall odds should be a power function of need odds. A power function means that some term (in this case n) is raised to an exponent (in this case f).

Mapping Need Probability to Latency of Recall

The architecture we proposed performs a serial search which terminates when one of the following two conditions is met: (1) the needed structure is

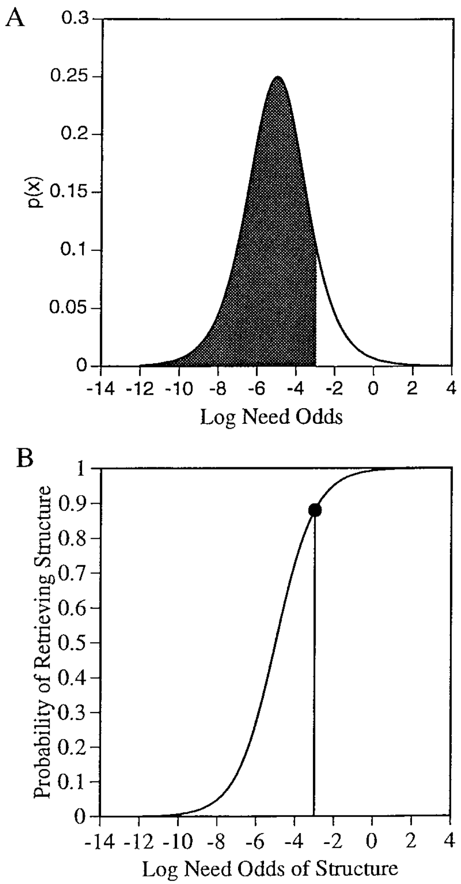

FIG. 1. (A) A plot of a logistic distribution (Eq. (A.1) from the Appendix), which describes noise about the retrieval criterion (Eq. (2)). The shaded region is the probability that a structure with need odds .05 ($ln(.05) = -3$) would exceed the recall criterion. (B) A plot of Eq. (A.2) which shows the probability that a structure would be retrieved as a function of its log need odds.

found or (2) the need odds of the next structure falls below $C/(G - C)$ plus or minus random noise. We are interested in the first condition: the time it takes to find a needed memory structure that has been correctly recalled. The architecture assumes a best-first serial search of memory. This means that if we knew the rank of a particular memory structure with respect to its need odds, then we would have a good idea of how long it should take to find it. Anderson and Schooler's (1991) method for estimating rank depends on the assumption that need odds, like so many other things, are ordered according to Zipf's law.

Zipf's law is a general observation concerning the ranked frequencies of various alternatives, proceeding from most frequent to least frequent. It states that the distribution of these ranks will be shaped like a backward *J* or highly skewed. The assumption here is that when memory structures are ordered in terms of their need odds, there will be a mass of structures with need odds near zero with a tail of a few structures with high need odds. Ijiri and Simon (1977) suggest that power functions describe distributions that conform to Zipf's law:

$$f(x) = ax^{-m}, \tag{4}$$

where $f(x)$ is the number of items with measure x (e.g., word frequency, need odds, etc.) and a and m are constants. Larger values of m correspond to distributions that are more tightly massed around 0. Larger values of a are associated with greater total numbers of memory structures. Figure 2A plots the power distribution described in Eq. (3) with the values of a and m set to 5 and 2 respectively. If need odds were distributed according to this version of Zipf's law, 50 memory structures would be expected to have need odds between .05 and .1.

Estimating the rank of a memory structure with need odds n depends on an estimate of the number of memory structures with need odds greater than n. If memories are distributed according to Zipf's law, then this implies that the rank of a memory structure would be a power function of need odds,

$$R = \int_n^\infty ax^{-m}dx = \frac{a}{m-1}\, n^{-(m-1)}. \tag{5}$$

Figure 2B plots rank against need odds with the values of a and m used in Figure 2A. The rank of a memory structure with need odds .05 would be 100 with the current values of a and m.

If we assume that each memory structure takes time t to consider, this implies

$$\text{Time for Recall} = An^{-b}, \tag{6}$$

where $A = at/(m - 1)$ and $b = m - 1$. Thus, for the architecture we have described the time to retrieve a particular structure should be a power function of need odds.

Summary of Optimal Retrieval

We started this section by sketching a rudimentary cognitive architecture. We proposed that the processes underlying retrieval search memory for the structures that are most needed in the current situation. We next argued that

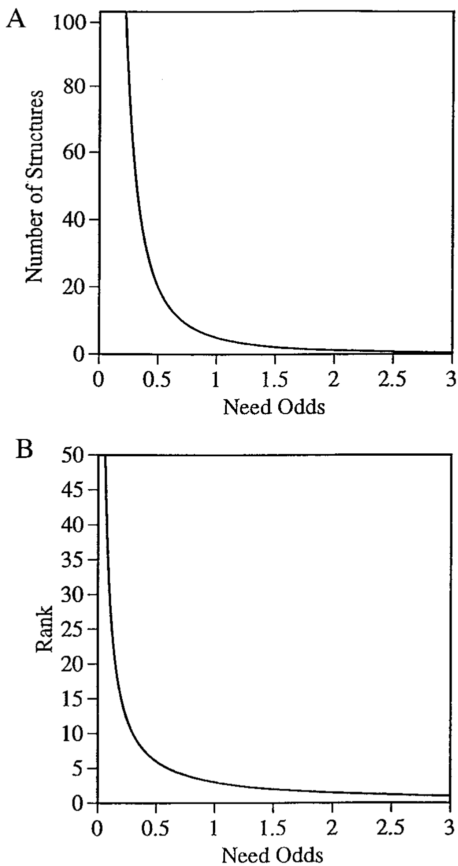

FIG. 2. (A) A plot of Eq. (4) which shows the number of structures which share a particular level of need odds. This captures the intuition that there are relatively few memories that are needed often and many memories which are needed infrequently. (B) A plot of Eq. (5) which shows the number of structures with need odds that exceed a particular level of need odds.

optimal retrieval processes would most efficiently use the limited capacity of the system by considering memory structures in order, according to their need odds. In this framework need odds is a critical determinant of behavior. We argued that in an optimal system power functions plausibly relate the odds that a particular memory structure is needed to both the odds that it would be retrieved and the latency of its retrieval. This leaves the critical question remaining as to how to estimate need odds, the topic of the next section. If we could do this we could use need odds to make predictions about recall odds and latency.

ESTIMATING THE ODDS THAT A MEMORY STRUCTURE WILL BE NEEDED

The rational analysis of memory borrows heavily from work in information retrieval (Anderson, 1990). Research in information retrieval focuses on areas as diverse as modeling patterns of book borrowings at libraries and file accesses on computers. A simple time-saving feature common to many word processors can help highlight the parallels between memory and these other systems. When a user goes to open a document file, the program presents a list of recently opened files to select from. When the sought after file is included in the list, the user is spared the effort of searching through the file hierarchy.

Clearly, the word processor does not know the goals of the user, nevertheless the files on the list are ones the system ''decides'' the user is likely to need. The word processor uses the heuristic that the more recently a file has been opened, the more likely it is to be opened now. Though this simple rule does a surprisingly good job of putting the appropriate files on the list, it ignores other obvious factors that could predict whether a file is likely to be needed. For example, it might be sensible to move more frequently opened files toward the top of the list, independent of when they were last used. Besides the history of a file (e.g., how frequently and recently it has been opened), the contexts in which a file has been needed in the past could be predictive of its current use. For example, when working on the body of a paper, the probability of needing the file that contains the associated tables and figures increases. Therefore, even though the table file may not have been opened recently, it should be moved toward the top of the list when the paper file is opened. It is not obvious how these various factors trade-off. That is, how should historical and contextual information be combined into a single measure (i.e., need odds) that can be used as the basis to sort the list of files? More importantly how should the need odds of a memory structure be assessed?

A Bayesian Estimate of Need Odds

Anderson (1990) formalized intuitions about how to estimate a memory structure's need odds, the odds that a memory structure will be needed to achieve a processing goal. He offered two factors that could be exploited to estimate need odds. The first is the history of the memory structure. A structure's history is the frequency and recency of (and to a lesser degree the spacing between) episodes in which the structure was previously needed. The history factor captures the odds that a particular memory structure will be needed now, given that it has displayed a particular history of being needed in the past. The second is the context factor which measures the strength of the association between a memory structure and elements of the current context. Here context is taken to be particularly salient elements of the environment. The context factor approximates the odds that the elements of the current context would be found, given that a particular memory structure is needed.

Within this framework, need odds is the product of the history and context factors:

$$n = \frac{P(S|H_S \ \& \ Qs)}{P(\bar{S}|H_S \ \& \ Qs)} \cong \frac{P(S|H_S)}{P(\bar{S}|H_S)} \times \prod_{q \in Qs} \frac{P(S|q)}{P(S)}, \qquad (7)$$

where $P(S|H_S \ \& \ Qs)$ is the probability of needing memory structure S given it has displayed a particular history of need, H_S, and the set of cues, Qs, constitutes the current context. $P(\bar{S}|H_S \ \& \ Qs)$ is the probability of not needing a particular memory structure given a particular past history and context.

Contextual strength is measured in terms of associative ratios that approximate the likelihood ratios common in Bayesian statistics. The denominator of this ratio, $P(S)$, is the base rate probability of needing a particular structure. The numerator, $P(S|q)$, is the conditional probability of needing the structure given the presence of some cue. The larger this ratio the better the cue, the greater the associative strength. The overall strength of the context is taken to be the product of the associative ratios linking each individual cue in the context to that structure. Equation 7 rests on assumptions of independence that are common to models of information retrieval. In particular, the associative strength linking a cue to any particular structure is (1) independent of the absence or presence of any other cue and (2) independent of the history displayed by a structure.

We are now in a position to provide accounts of some memory phenomena. First we will review an account based on the history factor (Anderson, 1990; Anderson & Schooler, 1991; Schooler, 1993) before returning to new accounts involving the interaction of context and history.

THE RETENTION FUNCTION

We will start by reviewing the analysis of retention interval—the effect of the interval since the last exposure to a to-be-remembered stimulus on the ability to recall the stimulus. The retention function is studied typically by exposing subjects to an item and then testing performance at various lags. Squire (1989) studied a retention function that was closely tied to subjects' day-to-day experience. He presented subjects with the potential names of TV shows. They had to decide whether the show had aired. Figure 3A plots subject performance as a function of the number of years since the show's cancellation.

Wickelgren (1975) and more recently others (Rubin, 1982; Wixted & Ebbe-

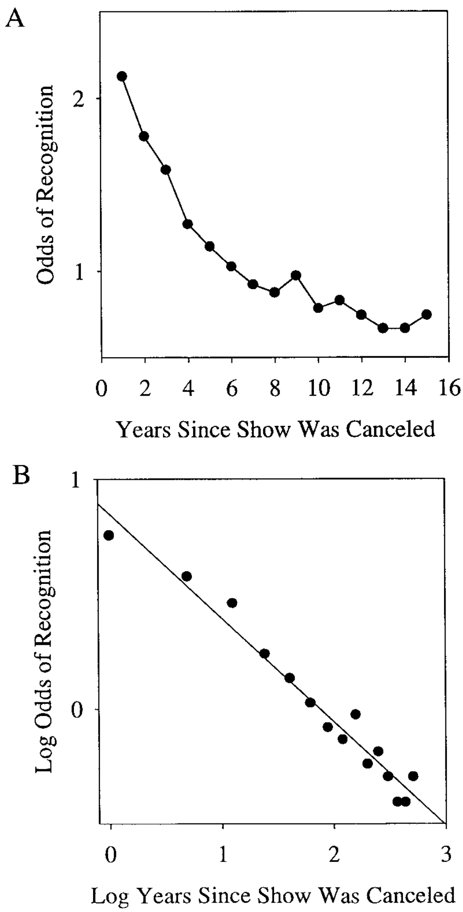

FIG. 3. Recognition for television shows. Retention function from Squire (1989), adjusted for guessing, in (A) standard coordinates and (B) log–log coordinates. A straight line in log–log coordinates indicates power function relation, $R^2 = .97$.

sen, 1991; Anderson & Schooler, 1991) have proposed that power functions describe well such negatively accelerated retention curves. Here a power relation means that performance, Y, is a function of the retention interval, L (for lag), raised to an exponent $-d$, the decay rate:

$$Y = IL^{-d}, \tag{8}$$

where I equals the initial level of performance at a test lag of 1 time unit. Since power functions are unbounded above, Y cannot be a bounded performance measure, such as probability of recall. Therefore it is often appropriate to use odds of recall, a standard measure of performance that is unbounded above like a power function. Equation 8 can be turned into a linear relation by taking the natural logs of each side:

$$\ln Y = \ln I - d \ln L. \tag{9}$$

These transformations allow the use of linear regression to determine whether a power function adequately captures the relation between retention interval and performance. The transformed power equation implies that if performance and retention interval are in a power relation with each other, then when log performance is plotted against log retention interval the result should be a straight line. Figure 3B plots Squire's data in this scale and it does appear relatively straight ($R^2 = .97$).

A Rational Account of the Retention Function

Theoretically, it was shown in Eq. (3) that odds of retrieval, could plausibly be a power function of need odds, n (see Appendix). Empirically, it has been observed that the relation between performance and retention interval takes the form of a power function such as Eq. (8). Substituting odds of recall from Eq. (3) for Y in Eq. (8), setting Eqs. (8) and (3) equal to each other, and isolating n yields

$$n = DL^{-g}, \tag{10}$$

where $D = (I/J)^{1/f}$ and $g = d/f$. Thus, the power function relation between recency and performance leads to Eq. (10), a prediction about the relation between how recently a structure has been needed and the odds of needing it now.

We will review the line of reasoning ending in Eq. (10). The first step was to derive Eq. (3) which showed that a power function plausibly relates the odds of needing a memory structure to the odds of retrieving it. The second step was to observe that power functions relate performance to lag. This was illustrated by Squire's experiment and motivates Eq. (8). The final step was to assume an equivalence between the odds of retrieving a memory structure and performance Y. This assumption enabled us to set Eqs. (3) and (8) equal to each other, and isolate need odds. Solving for n yields Eq. (10), which relates lag to need odds.

Alternatively, one can use Eq. (6) (instead of Eq. (3)), which relates time to recall a memory structure to its need odds, and behavioral data about how latency of recall varies with retention interval, to derive a prediction about how the odds of needing a memory structure varies as a function of the interval since it was last needed. As this behavioral latency function appears to take a power form (Anderson, 1995) (like Eq. (8), but with a positive exponent), the environmental prediction again is that the odds of needing a structure varies as a power function of how long it has been since it was last needed.

Anderson's (1990) account of the retention function lacked the data to address directly the question of how the odds of needing a memory structure relates to how recently it was last needed. Instead a model of book borrowing (Burrell, 1985) was adopted to simulate the informational demands that the environment might make. This model worked quite well, producing the predicted power function relation between lag and need. By augmenting the environmental model slightly, he showed that a Bayesian decision procedure (i.e., the rational analysis) in combination with a plausible environmental model could account for many other empirical findings, including the power law of practice and the exact interactions that occurred in spacing experiments. These results, however, were based on plausible assumptions about the informational demands that the environment places on memory that were embedded within the Burrell model.

Simon (1991) argued that the explanatory power of the rational analysis comes not from the assumption of optimality, but in large part from the auxiliary assumptions underlying the environmental model. The degrees of freedom in the environmental model made the approach difficult, or perhaps impossible, to falsify; for any pattern of behavior, a model of the environment could be constructed in which the observed behavior would be optimal. Anderson's original analysis was underconstrained by a lack of data about the informational demands that the environment makes on people. Replacing this model with statistics from the environment would remove an entire layer of assumptions from the theory, and go a long way in answering this criticism.

An Environmental Analysis of the Retention Function

We provided some of the necessary constraints by characterizing the informational demands that three environments place on people (Anderson & Schooler, 1991; Schooler & Anderson, 1991; Schooler, 1993). Gathering statistics about these informational demands requires detailed records of people's experience in the world. Ideally, we would follow people around, tallying their informational needs. Clearly, it is impractical to study the complete history of the informational demands that the environment places on an individual. Instead we have studied three environmental databases that capture coherent ''slices'' of the environment. Two of these databases, word usage in speech to children and in the New York Times headlines are linguistic in nature, but differ in their time scales. A third involves the daily distribution of authors who sent the second author (J.A.) electronic mail messages. This database captured aspects of his social environment.

Speech to children. Each word a child hears is another demand to retrieve the meaning of that particular word. MacWhinney and Snow's (1990)

CHILDES database is a collection of transcripts of children's speech interactions. We analyzed 25 h of preschool children's verbal interactions donated by Hall & Tirre (1979) that were collected by attaching wireless microphones to the children's clothing. For the analyses based on the Hall and Tirre corpus a word was defined to be needed each time it was mentioned in an utterance.

New York Times. Reading newspaper headlines requires retrieving the meaning of words that make up the headlines. We analyzed 2 years (1986 and 1987) worth of front page headlines. For the analyses based on the New York Times headlines a word was defined to be needed each time it was mentioned in a headline.

Authors of electronic mail. Each time someone receives mail, demands are made on memory to retrieve information about the person who sent it. We analyzed three years worth of J.A.'s mail messages. For these analyses information about an author was defined to be needed each time J.A. received a message written by that author.

Equation 10 makes a prediction for the unobservable odds of needing a memory structure as a function of when it was last needed. Analyses of the environmental databases yield results about the observable odds of encountering stimuli (e.g., words). Conclusions drawn from the environmental analyses require the tacit assumption that there is a one to one correspondence between encountering a word and needing a memory structure corresponding to that word.

With this caveat we come to the critical question underlying the account of the retention function: When an item's environmental need odds are plotted as function of how long it has been since it was last encountered is the resulting curve a power function? That is, can the curve be fit by Eq. (10). We show the results of our analysis of the New York Times in Fig. 4. In Fig. 4A we have plotted the odds of a word being included in the front page headlines as a function of the number of days since the word was last included. Figure 4B shows these same data in log–log coordinates. Here the curve is straight, suggesting that the environmental recency function, like its behavioral counterpart, can be described by a power function (R^2 = .99). Similar results hold for both the analyses of speech to children and the daily distribution of the authors of electronic mail messages.

Consistent with the predictions of the rational analysis, a power function relates environmental need odds and recency. Combined with the processing assumptions of the rational analysis this provides an account of the power function relation between memory performance and recency. We have carried out similar analyses that provide accounts of the power law of practice, the interaction between study spacing and retention interval, and the combined effects of practice and retention interval (Anderson & Schooler, 1991; Schooler, 1993). These results were remarkably consistent with the theory's predictions about the relation between how an item's need odds vary as

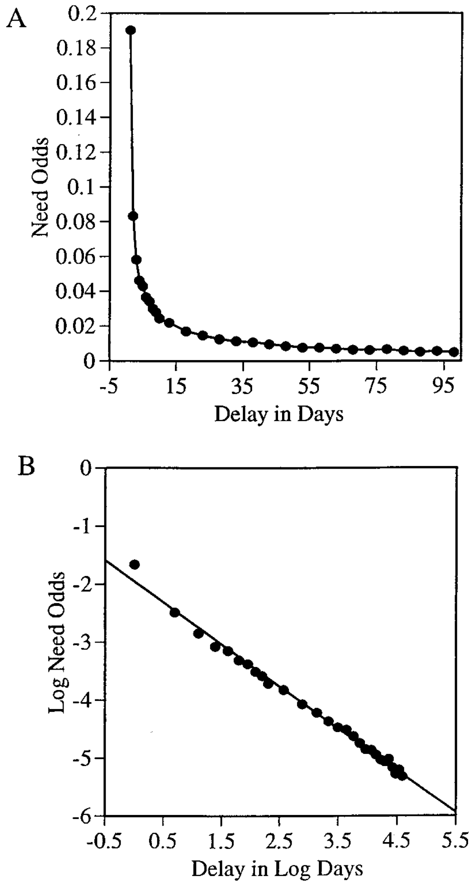

FIG. 4. (A) Environmental recency function which plots the odds of a word being included in a particular days headlines as a function of the number of days since it was last included. (B) The data in (A) plotted in log–log coordinates. A straight line in log–log coordinates indicates a power function relation, $R^2 = .99$.

function of its history. Unlike Anderson's (1990) original analyses, the accounts based on these environmental analyses are not subject to Simon's (1991) criticism because they do not depend on an environmental model.

Our previous environmental analyses have been limited to addressing issues involving the history factor. Next we will focus on the context factor which has previously received little attention.

CONTEXT FACTOR

In Eq. (7) the estimate of a particular item's need odds was taken to be the product of its need odds given its history (e.g., how recently it has been needed) multiplied by the context factor. The context factor gauges the strength of association between elements of the context and a particular memory structure. The critical component of the context factor is the associative ratio, $P(S|q)/P(S)$, that approximates the likelihood ratio common in Bayesian statistics. The denominator of this ratio, $P(S)$, is the base rate probability of needing a structure. The numerator is the conditional probability, $P(S|q)$, of needing a structure in the presence of some cue. The overall strength of the context is taken to be the product of the associative ratios of each of the individual cues in the context (Eq. (7)).

In the same way that we have used the environmental databases to investigate the history factor, we can use them to explore the context factor. Calculating the associative ratios requires estimating the base rate frequencies of the items (i.e., words) as well as the many conditional probabilities of finding one item in the presence of another. In our environmental analyses, the base rate probabilities were taken to be the proportion of all the headlines or utterances in which a word appeared. Estimating the conditional probability of finding a word in the presence of another requires a definition of context. A context was defined to be a headline or utterance; a word's context, then, was the other words that compose the headline or utterance. The top three panels in Table 1 show some words from the headlines along with associates that had particularly high associative ratios. For example, ''AIDS'' was included in 1.8% of all headlines and in 75% of the headlines that included ''virus.'' The associative ratio for the pair is 41 (i.e., .75/.018), or equivalently AIDS is 41 times more like to occur in a headline that includes virus than one that does not. The bottom two panels show examples from the CHILDES database.

CONTEXT AND RECENCY

There has been considerable work on associative memory, which maps onto the context factor. Issues relating to the history factor, such as recency, have received attention since Ebbinghaus. In our day to day lives these factors are not so neatly separated. Navigating through our environment requires responses that are simultaneously sensitive to both temporal and contextual factors. As these responses depend on past experience, memory retrieval, too, should be simultaneously sensitive to these factors. The structure of Eq. (10) contains predictions about how contextual and temporal factors combine. First, we will go through the environmental predictions and tests, before returning to their behavioral counterparts.

We reviewed the results of our environmental analyses which revealed that need odds was in a power relation with how recently an item was last encountered. Evidence for this was the nearly perfect linear relation between log need odds and log recency (Fig. 4B). We can simplify the history factor by tracking only an item's recency. The context factor can be simplified by

TABLE 1

	p(AIDS) = .018	
Associates	p(AIDS/associate)	p(AIDS/associate) / p(AIDS)
virus	.75	41.0
spread	.54	29.4
patients	.40	21.8
health	.27	14.6
	p(trade) = .015	
	p(trade/associate)	p(trade/associate) / p(trade)
imports	.77	49.6
gap	.60	38.7
exports	.44	28.7
deficit	.35	22.7
	p(senate) = .015	
	p(senate/associate)	p(senate/associate) / p(senate)
measure	.48	19.4
veto	.46	18.6
bill	.38	15.3
votes	.35	14.3
	p(play) = .0086	
	p(play/game)	p(play) = .0086 / p(play)
	.41	47.3
	p(made) = .0053	
	p(made/hand)	p(made/hand) / p(made)
	.18	34.5

Note. The probability that ''AIDS'' was in a headline was .018. Given that ''spread'' was in a headline, there was a .54 change that AIDS would also be included. The associative ratio for this pair is 29.4, indicating that spread is a good cue for AIDS. The top three panels are from the analysis of the New York Times, and the bottom two come from the analysis of the CHILDES database.

considering only a single cue. With these simplifications, one can derive a prediction by substituting Eq. (10) into Eq. (7):

$$n = DL^{-g} \frac{P(S \mid q)}{P(S)} . \tag{11}$$

The prediction can be seen most clearly by taking the natural logs of each side of Eq. (11):

$$\ln n = \ln D - g \ln L + \ln \frac{P(S \mid q)}{P(S)} . \tag{12}$$

Thus, the environmental prediction is that when log need odds is plotted as a function of log lag since an item was last encountered ($\ln L$) one should observe linear (slope of $-g$) and parallel functions for each value of contextual association ($\ln P(S \mid q)/P(S)$).

An Environmental Analysis of Context and Recency

Earlier we reviewed the rational analysis account of the retention function. Central to this account were the environmental analyses of speech to children and the New York Times. Next we will extend these analyses to test predictions that the rational analysis of memory makes for the combined effects of context and recency.

Method (New York Times). The aim of this analysis is to estimate recency curves contingent on whether a headline contains a strong associate. A word's strong associates are those words with associative ratios ($P(S \mid q)/P(S)$) that exceed 10. The rest are classified as weak associates. For the analysis of the New York Times all the words that had been last mentioned at a particular lag (say, 10 days) were selected. Each headline on the critical day was checked to see whether it contained one of the words. For a particular word, all the other words in a headline were classified as either strong or weak associates based on their associative ratios. The analysis was restricted to the 226 words that were mentioned at least 20 times and that had at least one strong associate. Estimates were made of the odds of a word occurring in a particular headline as function of the number days since it was last seen and of whether the headline contained at least one strong associate.

Method (CHILDES). The analysis of context and recency in the CHILDES database was restricted to the corpus from a single child, J.A.B., which was about 40% as large as the New York Times headline Corpus. The analysis was restricted to the 175 words that were mentioned at least 20 times and that had at least one strong associate. Here need odds was taken to be the odds of a word occurring in a particular utterance. These odds were estimated as a function of the number of utterances since the word was last mentioned and of whether the utterance contained at least one strong associate.

Results. Figure 5 shows the results in standard (A and B) and log–log (C and D) coordinates. Table 2 summarizes the regression fits to the strong and weak context curves in Figs. 5C and 5D. For both the CHILDES and New York Times analyses the decay parameters (i.e., exponents for A and B and slopes for C and D) for the strong and weak associate curves are approximately equal. In good agreement with the theory, it appears that the resulting curves

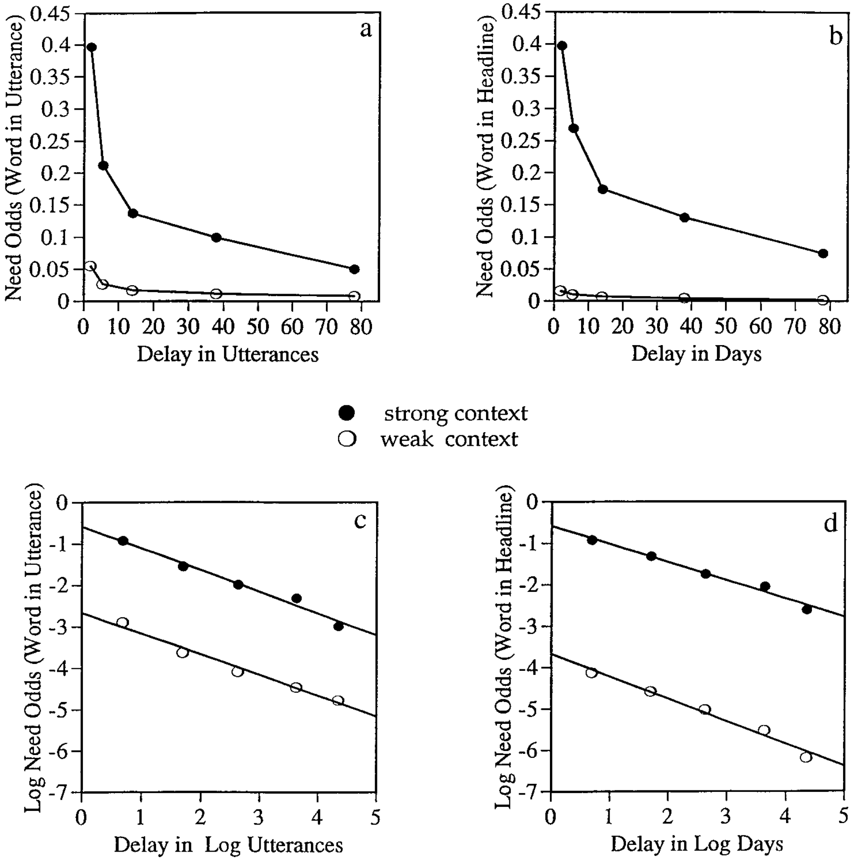

FIG. 5. Environmental recency curves from the analysis of the CHILDES and New York Times database. The left panels show the odds of a word being mentioned in an utterance as a function of the number of intervening utterances since it was last mentioned and whether the utterance included a strong associate (strong context) or did not (weak context). The right panels show the odds of a word being included in a particular headline as a function of the number of days since the word was last included and whether the headline included a strong associate. Parallel lines in (c) and (d) are consistent with environmental predictions of the rational analysis. The results are summarized in Table 2.

are linear and approximately parallel when plotted in log–log coordinates. The effects of association and recency in the environment (or at least in the New York Times and in speech to children) are independent of each other. In sum, the environmental analyses of the New York Times headlines and CHILDES database support the assumption that the rational analysis makes about the independence between the history and context factors. Next we explore whether this independence is reflected in behavioral data.

TABLE 2
R^2's from Power Fits to the Environmental Analyses of Context and Recency

	CHILDES		New York Times	
	R^2	Exponent	R^2	Exponent
Strong	.98	−.50	.98.	−.44
Weak	.98.	−.52	.98.	−.54

Note. The degree to which the strong and weak recency curves can be fit by power functions. The estimated exponents are listed as well.

Context and Recency in Memory

The last section established that both a word's past history of use and the statistical association of the current context to the word contribute to an estimate of the word's need odds—the odds that the word will be included in a particular New York Times headline or utterance. Further, it was found that a word's need odds can be predicted by a multiplicative relation that combines the independent predictions of the history and the context factors. The corresponding behavioral prediction is that when performance is measured in terms of log need odds of recall or in log latency of recall, the retention curves, corresponding to progressively stronger contexts, should be approximately parallel. This prediction can be mapped onto a cued recall task. In such an experiment the cues can be thought of as setting the context for recall. Thus, varying the degree of association between the cue and the target amounts to varying contextual strength. Before reporting the results of a new experiment, we will review some related work that bears on the question of how the influence of retention interval and associative strength between the cue and the target interact.

Mantyla and Nilsson (1988) showed two groups potential target words. One group spontaneously generated words to describe the target, while the second was instructed to generate distinctive words. Subjects were unexpectedly brought back either 1, 3, or 6 weeks later and given a cued recall test in which the cues were the words they had generated. At each retention interval the target words cued with distinctive words were recalled better than those cued with spontaneously generated words. When performance is measured in terms of the proportion of words recalled, Mantyla and Nilsson report a significant interaction between cue strength and retention interval. These results are apparently inconsistent with the predictions of the rational analysis. The predictions, however, are not for probability of recall, but rather for log odds of recall. Measuring performance in terms of log odds of recall removes the interaction. The data support the predictions of the rational analysis and at the same time illustrate the advantages of using the log odds transform to attenuate ceiling effects.

The question of shifting contexts relates to Thomson and Tulving's (1970) encoding specificity principle. In one demonstration of the principle, Thomson (1972) used homophones as potential targets in a recognition experiment. Subjects studied target words in the context of an associate. For instance ''iron'' might have been preceded by ''copper.'' At retention lags of 2, 5, 19, or 62 trials the target was re-presented in either the same context, or a new one like, ''dress''. Using the bounded measures of the probability correctly recognizing targets and the probability of correctly rejecting foils, Thomson reports an interaction between retention interval and changes in context. As with Mantyla and Nilsson's (1988) data, using an appropriate unbounded measure of performance (in this case d') removes the interaction in Thomson's data.

Taken together these experiments support the predictions of the rational analysis, but leave some questions open. First, the predictions are not only for odds of recall, but for latency of recall, data that these studies fail to report. Second, a significant problem with these experiments, for our purposes, is that all the retrieval contexts were strongly associated with the target. In Mantyla and Nilsson (1988), subjects in the spontaneous condition were asked to generate words that described the targets; it is reasonable to suppose that cues generated in this way are likely to be strongly associated with these targets. In Thomson's (1972) paradigm the difference between the strength of association between the cue and target depended not on the general level of association between the two, but on whether the context at study was compatible with the context at test. Further, the results are for recognition, which is outside the scope of the present rational analysis of memory. Finally, in these experiments the targets and cues develop associations within the experiment. A cleaner test of the model would minimize such within experiment associations.

In the following experiment we will address the question whether contextual and historical factors influence recall performance in a way which is consistent with the predictions of the rational analysis of memory and the environmental analyses of context and recency. We do this by testing recall in the presence of cues that are either strongly associated or unassociated to the targets.

EXPERIMENT 1

Experiment 1 used a cued recall task. Subjects studied words and later recalled them after various retention intervals and in the presence of cues that were either strongly associated or unassociated to the target word. The rational analysis predicts the following:

(1) Retention curves should be described by power functions. Though there

is some debate as to the precise parametric form of the retention curve (Rubin & Wenzel, 1996), it is certainly the case that power functions describe well the relation between recency and performance.

(2) Retention curves corresponding to recall in the presence of strongly associated and unassociated cues should be approximately parallel when plotted in a log–log scale.

Method

Subjects. Thirty-three subjects were drawn from the Carnegie Mellon community. Some subjects were paid for their time, and others participated to fulfill class credit.

Procedure. The experiment involved a continuous paradigm: trials in which words were studied were mixed with trials in which recall was tested. In a study trial subjects read a word and immediately had to say that word. In a recall trial the subject had to recall a word they had studied earlier.

The study and recall trials started identically. The first display consisted of a plus sign in the middle of the screen as warning that a trial was to begin. For the second display the plus sign was replaced by a word for 1000 ms, which in turn was followed by a word-stem on the third display. A word-stem consists of the first two letters of a word and underscores to hold the places of the remaining missing letters. When the word from the second display matched the word-stem of the third display, the subject had to respond with the word they had just read on the second display. These trials will be referred to as study trials.

Recall trials started off like study trials. On recall trials, however, the word from the second display was inconsistent with the word-stem on the third display. Subjects had to recall a word they had studied previously that matched the word-stem. The word from the second display was intended to act as a cue for the recall of the previously studied word that matched the word-stem. For half of the recall trials the word stem was preceded by a cue that was unassociated with the target, and for the other half it was proceeded by a cue with strong associations to the target.

Subjects said their answers aloud into a microphone, which tripped a voice-key, and stopped a timer. Immediately after they answered, the missing letters of the word-stem were filled in. If they failed to respond within 6 s, the answer appeared. Subjects scored their answers on the keyboard. Subjects were tested individually.

In Experiment 1 there were two levels of cue strength (strongly associated and unassociated), along with four retention intervals averaging 2.9, 9.3, 22.3, or 59.6 trials, which are approximately equidistant in log transformed scale. Retention intervals were measured in terms of the number of trials intervening between when a word was studied and it was the answer to a word-stem.

There was one warm-up block of 56 trials followed by 12 experimental blocks. Each of the experimental blocks had a total of 8 recall trials and 10 study trials. The study trials alternated with recall trials, and twice in each block one study trial immediately followed another. The additional study trials were added to allow for longer retention intervals. Some study trials were filled with words that were never used as targets.

Materials

Potential targets. We compiled a list of word association norms based on the results of four studies: Kent and Rosanoff (1911); Palermo and Jenkins (1964); Nelson, McEvoy, Walling, and Wheeler (1980); and Perfetti, Lindsey, and Garson (1971). We combined the results of these studies with those compiled by Shapiro and Palermo (1968) which summarizes an additional 20 other studies. For each cue in each study the most common response was recorded. These 1711 cues and 930 responses were used, respectively, as the cues and targets in Experiment 1.

Only those words that were between 4 and 6 letters long were included. Inclusion of some words forced the exclusion of others. Reasons for such exclusions were:

(1) If the words started with the same first two letters. This meant that within the experiment word-stems were uniquely associated with a single target.

(2) If the words shared the same strongest associate. Both ''tennis'' and ''fish'' are strong associates of ''net.'' If ''fish'' was primed with net, then ''tennis'' could not be included, because its strongest cue (i.e., ''net'') would have already been used.

(3) If words were each other's strongest cues. If ''king'' was a target, then ''queen'' would be its cue, and so ''queen'' was no longer eligible to be a target.

The selected words were randomly divided into three groups. The first group served as the associated targets, and the second the unassociated targets, where associated and unassociated refers to the strength of the association between the cue and the target. The third group of words did not serve as targets, rather their strongest associates were donated to act as cues to the unassociated targets.

Results

The results are plotted in Fig. 6A for log odds and Fig. 6B for log latency. Analyses of variances were performed on both performance measures. For log odds of recall there were significant main effects for association, $F(1, 32) = 106.9$, $p < .0001$, and for retention, $F(3, 96) = 61.5$, $p < .0001$. The interaction was also significant, $F(3, 96) = 5.3$, $p < .003$; performance fell more quickly as a function of retention interval, when the cue was unassociated to the target.[1] A similar pattern was found for log latency on correct responses. There were significant effects of association, $F(1, 31) = 86.7$, $p < .0001$, and retention, $F(3, 96) = 29.6$, $p < .0001$. The interaction was significant, $F(3, 93) = 5.6$, $p < .04$; the time it took to recall a target rose more quickly, in log–log scale, when the cue was unassociated with the target than when it was associated.

In carrying out the ANOVA, the Log odds of correct recall was calculated for each subject in each cell. A problem in using analyses of variance on the log-odds data is that the log odds transform is undefined when $p = 0$ (i.e., odds $= 0/(1 - 0)$ and when $p = 1$ (i.e., odds $= 1/(1 - 1)$). Subjects achieved perfect performance ($p = 1$) in 17% of the 264 subject by condition cells (i.e., 33 subjects $\times$ 8 conditions). The analyses reported above used the following substitution rule proposed by Berkson (1953) and recommended by Chatterjee and Price (1977): When $p = 0$, they suggest using $p = 1/2n$, where n is the number of trials. When $p = 1$, they suggest using $p = 1 - (1/2n)$.

Using the Berkson substitution is not wholly desirable, so we should look for alternative methods to test the prediction that the associated and unassociated retention functions are governed by the same decay parameter. The aim

[1] Only 23% of the errors resulted from subjects failing to respond within the 6 s time limit.

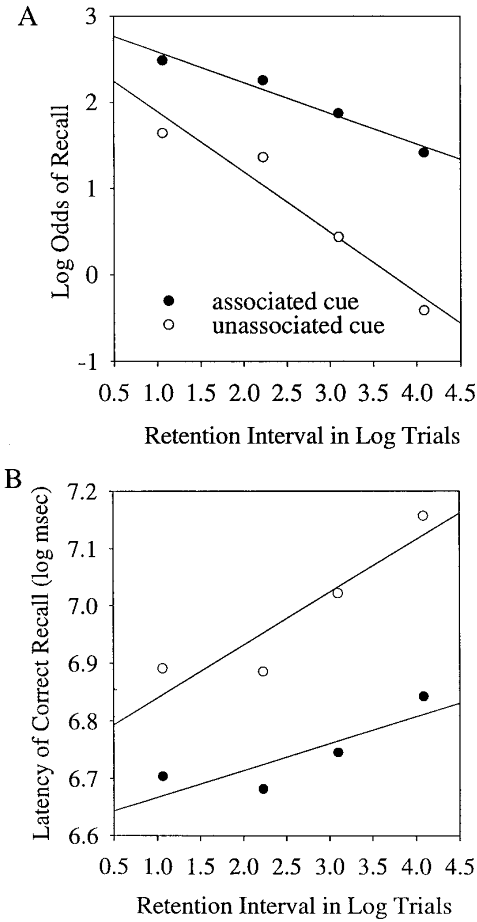

FIG. 6. Recall performance for Experiment 1 measured in (A) log odds of recall and (B) log latency of recall.

of this analysis is to find a procedure for estimating the decay parameters that does not require using the log odds transform. If odds of recall varies as IL^{-d}, where d is the decay parameter, L the retention interval, and I is the initial level of performance at a lag of 1 time unit, then the probability of recall can be written as

$$\text{Probability of Recall} = \frac{IL^{-d}}{1 + IL^{-d}}, \tag{13}$$

which is defined for all probabilities.

Instead of fitting odds of recall, we fit Eq. (13) to probability of recall assuming different I parameters and different decay parameters, d_s and d_u, for the strongly associated and unassociated conditions. Since Eq. (13) cannot be fit with linear regression, we used PRAXIS (Powell, 1964; Brent, 1973)[2] which searches a parameter space on the basis of minimizing the total sum of squared deviations between the predicted and observed values. If the same decay rates govern both the associated and unassociated retention curves, then when d_s and d_u are estimated for each subject individually there should be no difference between the two on average.

We obtained PRAXIS estimates of d_s and d_u for each subject individually. The average decay rates for d_s and d_u were 4.06 and $-.80$, respectively. The estimate for the associated decay rate was biased by a few extreme outliers, so it is perhaps more informative to look at the median values for d_s and d_u, which were $-.29$ and $-.78$, respectively. The results of a repeated measures t-test applied to these estimates of d_s and d_u indicate that they are indeed significantly different, $t(32) = 2.24$, $p < .03$. One can question, however, whether these PRAXIS estimates conform to the parametric assumptions underlying the t test, so we re-analyzed the estimates with non-parametric tests. Both the standard sign test, $p < .04$, and the Wilcoxon signed-rank test, $Z = 2.31$, $p < .02$, indicate that d_u is steeper then d_s.

Though the rational analysis makes no predictions about the shape of the cued recall latency distributions, we should check that the distributions had reached the floor within the 6 seconds in which subjects had to respond. If the distributions had not reached floor, this would complicate the interpretation of both the accuracy and latency results. Typically, it has been observed that latency distributions for memory retrieval tend to have long tails. Such long-tailed distributions have been found for recognition (Ratcliff & Murdock, 1976; Ratcliff, 1978; Nobel & Shiffrin, a, submitted), cued recall (Nobel & Shiffrin, b, submitted), and free recall (Wixted & Rohrer, 1993; Rohrer & Wixted, 1994). Nobel and Shiffrin (submitted-b) gave subjects 5 s in which to respond, and found that the distributions had reached floor within this time. For Experiment 1, excluding the slowest condition (i.e., unassociated cue and retention lag of 60 trials), 99.5% of all correct responses were made within 4 s and 99.8% within 5 s. For the slowest condition (unassociated cue and retention lag of 60 trials) 95.1% of correct responses were made within 4 s, and 98.8% were made within 5 s. Thus, there is little evidence that the latency distributions were being artificially truncated.

In summary, for both latency and accuracy the evidence is that the retention function is steeper in the presence of weaker cues. Although the environmental functions (Fig. 4 and Table 2) were somewhat steeper in the presence of

[2] The program was written by Karl Gegenfurtmer in 1987.

weak cues, it does appear that the behavioral functions show a much greater discrepancy in this direction. Thus, we judge the behavioral data contrary to the strong predictions we derived from the rational analysis.

In contrast to the predictions of the rational analysis, in all the curves in Fig. 6, there is some departure from linearity. When performance is measured in terms of log odds of recall, this departure from linearity is statistically significant for the unassociated retention curve, $F(2, 128) = 3.26$, $p < .05$, but not for the associated curve, $F(2, 128) = .53$, $p < .59$. When performance is measured in terms latency neither the unassociated curve, $F(1, 128) = 1.86$, $p < .16$, nor the associated curve, $F(1, 128) = 1.68$, $p < .19$, depart significantly from linearity.

DISCUSSION

A failed prediction has been identified for the pure rational analysis approach. The predictions of the rational analysis are that the associated and unassociated retention curves should be linear and parallel in log–log scale. The curves approximately satisfy this characterization, but for both latency and accuracy there are significant deviations from linearity and/or parallelism. Next, we will examine how we might repair the problem by loosening the process assumptions of the rational model.

The predictions of the theory depend both on the response functions (Equations 3 and 6) relating need odds to behavior and on the informational demands that the environment places on the system. Earlier versions of the rational analysis relied heavily on a model of the environment to simulate the informational demands. Problems with the theory could be repaired by modifying this model of the environment. However, we have now observed directly how need odds varies in the environment and so are no longer free to accommodate the predictions by adjusting the environmental model. This leaves us with the response functions for more careful consideration. Such considerations raise the issue that as modeled by the rational analysis of memory the basic response functions ignore processing not directly related to retrieval. Such processing might include reading the cue word, deciding to respond, saying the answer aloud, and so on.

A Better Estimate of Latency of Retrieval

Most models of recall latency suppose that overall latency is a function of retrieval time plus an ‘‘intercept’’, or base time, that reflects time to encode the probe and generate the response. The latency predictions derived earlier, however, ignored this, tacitly assuming an intercept of zero. Experiment 1 provides at least a rough estimate of the potential intercept. For a study trial the answer to the word-stem was the word that had just been read. We took the average latency on the study trials (671 ms) as an estimate of the intercept and subtracted these times from the response times in the test trials. These

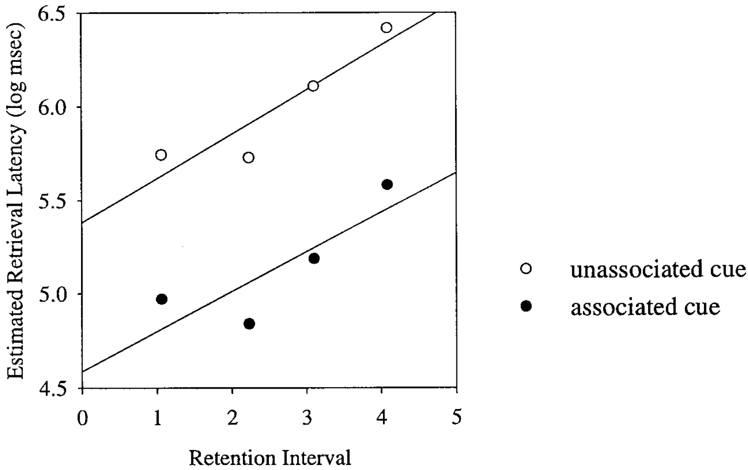

FIG. 7. Estimated time to retrieve memories in Experiment 1. The results are plotted in log–log coordinates.

estimated retrieval latencies are plotted in Fig. 7 in the power form. The exponents for the associated and unassociated curves are respectively .21 (R^2 = .70) and .24 (R^2 = .85). Thus, it does seem that subtracting off this intercept removes the non-parallelism in the latency function.

A Better Estimate of Odds of Retrieval

The distinction we make between retrieval and non-retrieval processes is similar to distinctions made in SAM (Raaijmakers & Shiffrin, 1981), where two events must occur for correct recall. First the appropriate memory structures must be sampled, and then additional processing is required to recover enough information from the sampled structure to make a response. In SAM, the probabilities of sampling and recovering a memory structure are not independent of each other: both probabilities are functions of the strength of the memory structure.

In our analysis, the probability of retrieving a structure, P_V, maps onto SAM's sampling probability, and the probability of successfully carrying out additional non-retrieval processing, P_A, maps onto SAM's recovery probability. The probability of retrieval P_V (derivable from Eq. (3)) depends on the need odds of a structure, much as SAM's sampling probability depends on a memory structure's strength. In keeping with our goal of architectural simplicity, we assume that probability of successful additional processing, P_A, is both independent of the probability of retrieval, P_V, and of the need odds of the retrieved structure. With these assumptions the probability of correct recall P_R equals the product of the two:

$$P_R = P_V P_A. \tag{14}$$

An assumption that went into the earlier predictions was that the retrieval of the appropriate memory structures was sufficient for a correct response, amounting to setting P_A to 1. Undoubtedly, P_A is something less than 1. Thus, a better estimate of P_V would be

$$P_V = \frac{P_R}{P_A}. \tag{15}$$

The question is how to estimate P_A. Experiment 1 provides a possible estimate for the additional processing represented by P_A. For a study trial the answer to the word-stem was the word that had just been read. On such trials the structures needed to respond are likely to be still active. One might take the probability of a correct response on the study trials (.99) as an estimate of P_A. We then estimated P_V by plugging in this estimate of P_A into Eq. (15) and using P_R for the various conditions. Figure 8A plots the odds of retrieval, O_V, resulting from these substitutions. The exponents for both the associated and unassociated curves were respectively $-.47$ ($R^2 = .99$) and $-.61$ ($R^2 = .95$). However, one might well question this as an estimate of P_A. There may be more extraneous factors that can trip up a subject on a test trial than a study trial. The results can be fit by setting P_A to the plausible value of .96. In this case the exponents for both the strong and weak associate curves are $-.67$ with respective R^2's of .99 and .95. Figure 8B plots log-odds of retrieval derived from setting P_A to .96. Note the deviations from linearity attenuate.

RECONCILING THE PROCESS AND RATIONAL APPROACHES

The original goal of the research reported in this paper was to test predictions that the rational analysis makes about the relation between how context and recency combine in the environment and in memory. Answering this question required considering more carefully how non-retrieval processes might influence the behavioral measures we were using to test these predictions. This necessarily moved us away from a pure rational analysis to a more process oriented approach. Though our account of the data appeals to specific processing assumptions, it still retains the flavor of the original analysis. Our account shares with the pure rational analysis the hypothesis that retrieval is tuned to the statistical structure of the environment, and the constraints provided by our empirical analyses of this structure.

It is a relatively mild complication of the pure rational analysis approach to add an intercept parameter for latency and the possibility of errors not due to the retrieval process. However, as one tries to model more and more

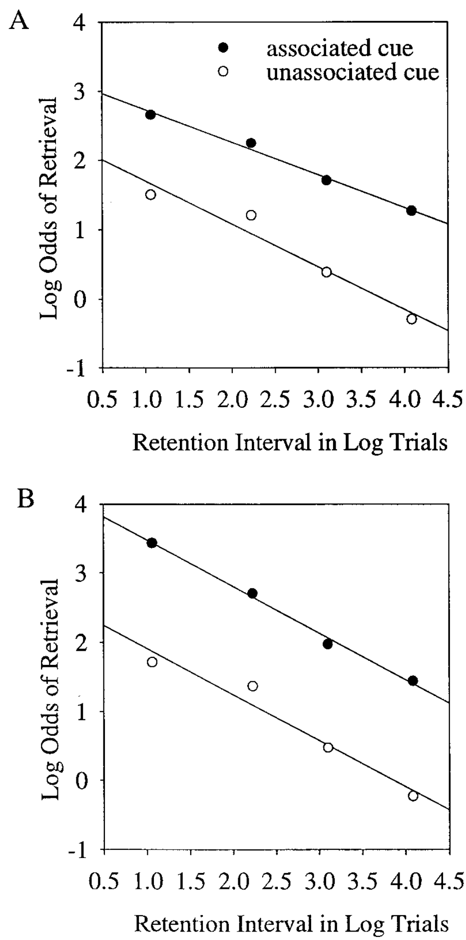

FIG. 8. Estimates of Experiment 1 retrieval odds based on Eq. (15). In (A) the probability of successful additional processing, P_A, was determined empirically to be .99. In (B), P_A was allowed to vary and was estimated to be .96. The results are plotted in log–log coordinates.

complex phenomena, more and more such complicating assumptions will need to be added and the contribution of the pure rational analysis will be proportionately less. Anderson (1991) concluded that even if various cognitive modules like memory were performing optimally it is unlikely that the complex system, involving the interactions of many such modules, could lay any claim to optimality. The ACT-R model of Anderson (1993) is an effort to embed such rational analyses in an overall architecture. Anderson found that rational analyses of individual modules did a lot to guide the design of the component processes but that there were significant issues of integration where the rational analysis provided little guidance.

Our better understanding of the relation between the process and rational approaches leads us to close this paper with a final reflection on the correspondence between the environment (which we described in the first half of the paper) and memory (which we described in the second half of the paper). There is good correspondence between the predictions of the rational analysis, the environment, and the behavioral data. Analyses of the two environmental data sets reveal that the decay parameters governing how environmental need odds varies with how recently an item was last encountered do not depend on the level of association between the item and other elements of the current context. The environmental analyses lead to a novel behavioral prediction that the decay parameters governing how performance varies with retention interval should not depend on the level of association between the to-be-remembered item and the cues used to retrieve it. Analyses of the unadjusted data from Experiment 1 were inconsistent with the predictions of the rational analysis. However, the analyses on the adjusted data that included intercepts for additional non-retrieval processing were in line with the predictions.

A primary goal of the rational analysis is to provide accounts of data with as few processing assumptions as possible. Yet Experiment 1 showed that the predictions were sensitive to assumptions about the dependence between memory retrieval and the additional non-retrieval processing. The experiment certainly gave us an appreciation of the sensitivity of our predictions to the accompanying processing assumptions; a sensitivity that was not anticipated in the original rational analysis of memory.

This raises the issue of the optimality of cognition, which remains a contentious point in the rational analysis. It is the claim of optimality that gives the rational analysis its predictive power. The solution to the optimization problem is a precise behavioral prediction. However, there are reasons for doubting whether the behavior is in all cases optimal. This moves us from being able to formulate strong hypotheses, such as that of no interaction. As a research strategy we think it is better to start with a strong assumption of optimality, and pursue the predictions that it entails. Pursuing the strong prediction of no interaction between recency and context brought us to a better understanding of how adaptive memory embeds in a more complex system.

As we noted at the outset, Estes (1955) admonished us ‘‘to shift the burden of explanation from hypothesized processes in the organism to statistical properties of environmental events.’’ In its most basic form the rational analysis of memory moves in this direction and emphasizes the relation of these internal processes to the structure of the environment. Yet as we have seen this is not sufficient: the theory falters if we fail to consider the relation among these internal processes.

APPENDIX

The derivation of Eq. (3), which relates the odds of retrieval to need odds, starts with a logistic distribution (Fig. 1A) that describes the distribution, $p(x)$, of the observed recall criterion in logged scale,

$$p(x) = \frac{e^{f(k-x)}}{(1 + e^{f(k-x)})^2}, \tag{A.1}$$

where $p(x)$ is centered about k, which equals $\ln(C/(G - C))$, x equals the log of the observed recall criterion, and f is related to the variance ($f = \pi/(\sqrt{3}\sigma)$).

Figure 1B plots the probability that any value of x exceeds the observed retrieval criterion, which in turn is the probability $P(x)$ that a structure with log need odds x will be retrieved. These probabilities are described by a logistic function,

$$P(x) = \frac{1}{1 + e^{f(k-x)}}. \tag{A.2}$$

Equation A.2 implies that odds of retrieval $O(n)$, where $x = \ln n$, will be a power function of need odds:

$$O(n) = \frac{P(n)}{1 - P(n)} = \left(\frac{C}{G - C}\right)^{-f} n^{f}. \tag{A.3}$$

REFERENCES

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

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1993). *The rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1995). *Learning and memory.* New York: Wiley.

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

Brent, R. P. (1973). *Algorithms for minimization without derivatives.* Englewood Cliffs, NJ: Prentice Hall.

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

Chatterjee, S., & Price, B. (1977). *Regression analysis by example.* New York: Wiley.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). ''PsyScope: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers.'' *Behavioral Research Methods, Instrumentation, and Computation.*

Douglas, L. N., McEvoy, C. L., Walling, J. R., & Wheeler, J. W. (1980). The University of South Florida homograph norms. *Behavior, Research, Methods, & Instrumentation,* **12,** 16–37.

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

Hall, W. S., & Tirre, W. C. (1979). *The communicative environment of young children: Social class, ethnic and situation differences.* University of Illinois, Center for the Study of Reading.

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

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

Kent, G. H., & Rosanoff, J. (1911). *A study of association in insanity.* Baltimore: Lord Baltimore Press.

Kintsch, W., & Van-Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review,* **80,** 237–251.

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language,* **17,** 457–472.

Mantyla, T., & Nilsson, L. G. (1988). Cue distinctiveness and forgetting: effectiveness of self-generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning Memory and Cognition,* **14,** 502–509.

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

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping a nonparametric approach to statistical inference.* London: Sage Publications.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review,* **89,** 609–626.

Nelson, D. L., McEvoy, C. L., Walling, J. R., & Wheeler, J. W. (1980). The University of South Florida homograph norms. *Behavior, Research, Methods, & Instrumentation,* **12,** 16–37.

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: Lawrence Erlbaum Associates.

Nobel, P. A., & Shiffrin, R. M. (submitted-a). A model for accuracy and response time in recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Nobel, P. A., & Shiffrin, R. M. (submitted-b). A model for accuracy and response time in cued-recall. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Palermo, D. S., & Jenkins, J. J. (1964). *Word association norms: Grade school through college.* Minneapolis: University of Minnesota Press.

Perfetti, C. A., Lindsey, R., & Garson, B. (1971). *Association and uncertainty: Norms of association to ambiguous words.* Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function in several variables without calculating derivatives. *Computer Journal,* **7,** 155–162.

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

Ratcliff, R. (1976). A theory of memory retrieval. *Psychological Review,* **85,** 59–108.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review,* **83,** 93–134.

Rohrer, D., & Wixted, J. T. (1994). Proactive interference and the dynamics of free recall. *Memory & Cognition,* **22,** 511–524.

Rubin, R. C., & Wenzel, A. E. (1996). 100 years of forgetting: A quantitative description of Retention. *Psychological Review,* **103,** 734–760.

Schooler, L. J. (1993). Memory and the statistical structure of the environment. Ph.D. Dissertation, Department of Psychology, Carnegie Mellon.

Schooler, L. J., & Anderson, J. R. (1991). Does memory reflect statistical regularity in the environment? *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Shapiro, S. I., & Palermo, D. S. (1968). An atlas of normative free association data. *Psychonomic Monograph Supplements,* **20,** 219–250.

Simon, H. A. (1991). Cognitive architectures and rational analysis: Comment. In K. VanLehn (Ed.), *Architectures for intelligence.* (pp. 15–39). Hillsdale, NJ: Lawrence Erlbaum Associates.

Squire, L. R. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **15,** 241–245.

Thomson, D. M. (1972). Context Effects in Recognition memory. *Journal of Verbal Learning and Verbal Behavior,* **11,** 497–511.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology,* **86,** 255–262.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review,* **76,** 1–15.

Wickelgren, W. A. (1975). Single-trace fragility theory of memory dynamics. *Memory and Cognition,* **2,** 775–780.

Wixted, J. T., & Ebbesesen, E. B. (1991). On the form of forgetting. *Psychological Science,* **2,** 409–415.

Wixted, J. T., & Rohrer, D. (1994). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **19,** 1024–1039.

(Accepted October 1, 1996)