JOURNAL OF MEMORY AND LANGUAGE **38,** 341–380 (1998)
ARTICLE NO. ML972553

An Integrated Theory of List Memory

John R. Anderson, Dan Bothell, Christian Lebiere, and Michael Matessa

Carnegie Mellon University

The ACT-R theory (Anderson, 1993; Anderson & Lebiere, 1998) is applied to the list memory paradigms of serial recall, recognition memory, free recall, and implicit memory. List memory performance in ACT-R is determined by the level of activation of declarative chunks which encode that items occur in the list. This level of activation is in turn determined by amount of rehearsal, delay, and associative fan from a list node. This theory accounts for accuracy and latency profiles in backward and forward serial recall, set size effects in the Sternberg paradigm, length–strength effects in recognition memory, the Tulving–Wiseman function, serial position, length and practice effects in free recall, and lexical priming in implicit memory paradigms. This wide variety of effects is predicted with minimal parameter variation. It is argued that the strength of the ACT-R theory is that it offers a completely specified processing architecture that serves to integrate many existing models in the literature.

From our vantage point on psychology it seems that more experiments have been run using the list memory paradigm than any other experimental paradigm (for recent reviews see Healy & McNamara, 1996; Raaijmakers & Shiffrin, 1992). This is a paradigm in which subjects are presented with a list of words and then are tested for their memory of the words. The test may involve an attempt to recall the words in the presented order in which case it is called serial memory, an attempt to recall the words in any order in which case it is called free recall, an attempt to recognize the words in which case it is called recognition memory, or an attempt to do something involving the words but not requiring that the subject recall these words (like stem completion) in which case it is called implicit memory.

The list memory paradigm was the paradigm that Ebbinghaus used in the first experiments on human memory (although he used nonsense syllables). It continued to be used in a great many studies in the subsequent decades. Ebbinghaus and other early researchers usually used serial memory tests. With the rise of cognitive psychology, research on human memory grew in importance and the list memory paradigm seemed to rise with it. The free-recall paradigm was initially of great importance in showing the effects of organizational factors on memory. More recently, recognition memory has become important in discriminating among major theories of memory. The implicit memory research is almost exclusively a phenomenon of the past two decades but has become one of the hottest areas of research in cognitive psychology. The serial memory version of this paradigm has not been forgotten and is currently prominent in the form of tests of immediate or working memory.

Most theoretical accounts one finds address phenomena in just one of these subdomains of list memory and have not tried to provide an integrated account that spans all of the subdomains. Different subdomains involve different aspects of cognition—memory for serial order, free recall strategies, structure of lexical memory, etc. Therefore, it is natural that detailed accounts of specific subdomains should focus somewhat on different aspects of the

Address reprint requests to John R. Anderson, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: ja+@cmu.edu.

We thank Chris Schunn for his comments on this manuscript. This research has been supported by Grants N00014-96-I-0491 from the Office of Naval Research and SBR-94-21332 from the National Science Foundation.

0749-596X/98 $25.00

cognitive system. Still the similarity of the learning experience (studying a list of words) creates the expectation that there should be some way of integrating these accounts. There are some theories that have striven for integrated accounts of list memory (e.g., Todam—Murdock, 1993; Lewandowsky & Murdock, 1989; SAM—Raaijmakers & Shiffrin, 1981; Gillund & Shiffrin, 1984; and now REM—Shiffrin & Steyvers, 1997). This paper will show that the ACT-R theory of cognition (Anderson, 1993; Anderson & Lebiere, 1998) also offers an integrated account. ACT-R is unique in that its basic assumptions were not fashioned to account for list memory but for cognition more generally.

In this paper we will give a brief description of the ACT-R theory sufficient to understand its application to the list memory experiments. Then we will review the list memory experiments in the order of serial memory, recognition memory, free recall, and implicit memory.

THE ACT-R THEORY

ACT-R (Anderson, 1993; Anderson & Lebiere, 1998) is a theory which aspires to provide an integrated account of many aspects of human cognition. It assumes that a production system operates on a declarative memory. It is a successor to previous ACT production-system theories (Anderson, 1976, 1983) and continues the emphasis on activation-based processing as the mechanism for relating the production system to the declarative memory. Different traces in declarative memory have different levels of activation which determine their rates and probabilities of being processed by production rules. ACT-R is distinguished from the prior ACT theories in that the details of its design have been strongly guided by the rational analysis of Anderson (1990). As a consequence of the rational analysis, ACT-R is a production system tuned to perform adaptively given the statistical structure of the environment.

Declarative knowledge is represented in terms of *chunks* which are schema-like structures. Chunks are of different types and each type has an associated set of pointers encoding its contents. To emphasize that ACT-R applies to other domains besides list memory, we will describe an example of its representations from the domain of cognitive arithmetic. Figure 1 is a graphical display of a chunk of the type addition-fact, which encodes that 3 + 4 = 7 with pointers to *three*, *four*, and *seven*. The B_i, W_j, and S_{ji} are quantities relevant to activation computation and they will be discussed in the next subsection.

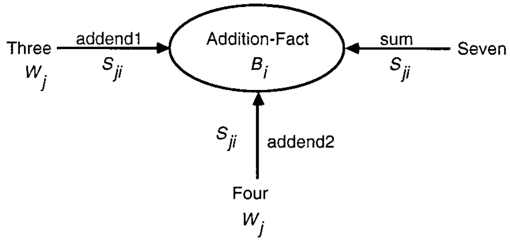

FIG. 1. A chunk encoding the fact that 3 + 4 = 7.

According to ACT-R, procedural knowledge, such as mathematical problem-solving skills, is represented by *production rules* which coordinate the retrieval of declarative information like that in Fig. 1 for purposes of problem solving. For instance, suppose a child was at the point illustrated below in the solution of a multi-column addition problem:

$$\begin{array}{r} 531 \\ +248 \\ \hline 9 \end{array}$$

Focused on the tens column, the following production rule might apply from the simulation of multicolumn addition (Anderson, 1993):

Process-Column.
IF the goal is to process a column containing digits d1
and d2 and d3 is the sum of d1 and d2
THEN set a subgoal to write out d3.

Each production consists of a condition and an action. In ACT-R each condition consists of a specification of the current goal (e.g., ''the goal is to process a column containing digits d1 and d2'') and some number of retrievals from declarative memory (e.g., ''d3 is the sum of d1 and d2''). According to the ACT-R theory, an important component of the time to apply a production is the time to match the elements in the condition of a production. The time to match the goal is not a significant factor in the ACT-R theory because the goal is already in the focus of attention but ACT-R must retrieve chunks from declarative memory to match the rest of the condition and the time to match the condition is the sum of these retrieval times. The times to perform these retrievals will be important contributors to the latency for the production rule and the levels of activation of the chunks will determine these retrieval times. So, in this case, the time to apply this production will be determined by the level of activation of the chunk encoding 3 + 4 = 7 in Fig. 1. The next subsection will explain how activation determines retrieval time. In addition to the retrieval time to match the condition, there are times associated with executing the action. This action latency is minimally 50 ms in the ACT-R theory but can be longer when significant motor actions are involved such as typing or speaking.

Much of the recent development of the ACT-R theory has focused on tasks like mathematical problem solving. However, the ACT theory originated as a theory focused on human memory (Anderson, 1976; Anderson & Bower, 1973). This paper will propose that productions similar to those guiding problem solving in a mathematics domain are guiding recall in list memory paradigms. So, one contribution of this paper will be to show that list-memory experiments can be viewed as problem-solving tasks.

Activation

Activation of declarative structures has always been an important concept in the ACT theories. Basically activation determines how available information will be.[1] The activation of a chunk is the sum of its base-level activation and the activations it receives from the elements currently in the focus of attention. Formally, the equation in ACT-R for the activation, A_i, of chunk i is

$$A_i = B_i + \sum_j W_j S_{ji}, \qquad (1)$$

where B_i is the base-level activation of chunk i, W_j is the salience or source activation of element j in the focus of attention, and S_{ji} is the strength of association from element j to chunk i. For instance, in the context of retrieving the knowledge unit that 3 + 4 = 7 in response to seeing 3 and 4 in a column, the W_j would be the source activations of the elements 3 and 4 in the column and the S_{ji} would be the strengths of association from these elements to the chunk encoding 3 + 4 = 7.

Figure 1 illustrates these quantities in the network encoding of the chunk. It is assumed in ACT-R, in contrast to early versions of ACT (such as in Anderson, 1976) but as in ACT* (Anderson, 1983), that these activation levels are achieved rapidly and that time to ''spread'' activation is not a significant contributor to latency. However, unlike ACT* there is no multilink spread of activation. Rather, activation is simply a direct response to source elements like j. As such, the theory is much like the SAM theory (Raaijmakers & Shiffrin, 1981; Gillund & Shiffrin, 1984), except that activations in ACT-R are like logarithms of SAM familiarities since they add rather than multiply. It is important to keep conceptually separate the quantities A_i and W_j. The former are activations, which control retrieval from declarative memory, while the latter reflect the salience or attention given to the cues. The W_j are referred to as *source activations.*

According to Eq. (1) access to memory is going to vary with base-level activation (the B_i) and associative activation (determined by

[1] According to the ACT-R theory, the activation of a chunk reflects a preliminary estimate of how likely it is to match to a production at the current point in time. More precisely, activation estimates the log odds that the chunk will match to a production.

the W_j and the S_{ji}). The base-level activation of a chunk is a function of its history of use at times $t_1, \ldots, t_n$, where t_j measures how much time has passed since the jth use

$$B_i = \ln(\sum_{j=1}^{n} t_j^{-d}). \qquad (2)^2$$

As developed in Anderson and Schooler (1991), this equation produces both the Power Law of Forgetting (Rubin & Wenzel, 1996), where the strengths of individual experiences decay as power functions, and the Power Law of Learning (Newell & Rosenbloom, 1981), where individual experiences accumulate strength as a power function of number of exposures. The decay effect is produced by the negative d exponent, while the practice effect is produced by the summation across experiences. Throughout this simulation we will use the ACT-R standard of setting the decay parameter d to 0.5. This has emerged as the value of d which gives good accounts of results across a wide range of domains.[3]

The associative activation in Eq. (1) is a product of the weights W_j and the strengths S_{ji}. The typical ACT-R assumption is that, if there are n sources of activation the W_j are all set to l/n. The associative strengths, S_{ji}, reflect the competition among the associations to the cue j. According to the ACT-R theory, $S_{ji} = S + \ln(P(i|j))$, where $P(i|j)$ is the probability that chunk i will be needed when j appears in the context. S is a constant.[4] Basically, this equation makes the strength of association between j and i a function of how likely i is in the presence of j. Built into this equation is the prediction of a fan effect (Anderson, 1974) in that the more things associated to j the less likely any of them will be, on average, in the presence of j. That is, if there are m elements associated to j their average probability will be l/m and $S_{ji} = S - \ln(m)$. This is the simplification that will be used in all the simulations presented in the paper. Thus,

$$S_{ji} = S - \ln(m). \qquad (3)$$

Anderson, Reder, and Lebiere (1996) introduced a new ACT-R assumption motivated by the many errors in algebra that seemed to be due to misretrieving arithmetic facts and algebraic transformations which were similar to the correct ones. Therefore, we extended the pattern-matching facility in ACT-R to allow partial matches between the conditions of productions and chunks in declarative memory. To favor more complete matches we added a mismatch penalty that reflected the degree of mismatch. The goodness of the match M_i of a chunk i to a condition in a production rule is

$$M_i = A_i - P, \qquad (4)$$

where P is a mismatch penalty that depends on the similarity between the chunk and condition. In practice it becomes a parameter to be estimated. Thus, faced with the goal to retrieve the sum of 3 + 4, the chunks 3 + 4 = 7 and 3 + 1 = 4 would have equal activation scores (both are associated to source elements 3 and 4), but 3 + 1 = 4 would receive a mismatch penalty (because the addends 1 and 4 do not match). The chunk retrieved to match a production condition is the one with

[2] The summation in this equation could be viewed as describing a situation where each use resulted in additional synaptic efficacy but which then decayed away (e.g., a new receptor site which decays away with time). As n gets large we sometimes use approximations in ACT-R simulation. One might indeed imagine that the equation is also only an approximate characterization of change in synaptic efficacy in the case of large n.

[3] This equation asserts that the critical variable is time but it is unlikely that this is truly clock time. It might be better read as number of intervening events or some other construct. However, our models will treat the critical variable as clock time because that is usually the variable directly manipulated in studies.

[4] S is essentially a scale constant whose value is reflected in the setting of other parameters in ACT-R—in particular, F and τ—see Eqs. (5) and (6). It can be set in the simulation but if not set it will default to the log of the total number of chunks.

the largest match score. Normally, when a perfectly matching chunk competes with a partially matching chunk, the perfectly matching chunk will be retrieved because it has the largest match score. However, there is noise in the activation values and occasionally a partially matching chunk will be selected over a perfectly matching chunk because the activation noise gives it sufficiently more activation to overcome the mismatch penalty it suffers. When a partially matching chunk so beats out a perfectly matching chunk, there will be errors of commission in retrieval. Only when all chunks fail to reach an activation threshold does retrieval fail completely (errors of omission). Partially matching errors of commission are the cause of intrusions in recall while retrieval failures are the cause of recall blanks.

There now remains the issue of how to relate these match scores to dependent measures. With respect to latency, the time to retrieve a chunk i is related to its match score by the formula

$$\text{Time}_i = Fe^{-M_i}, \tag{5}$$

where F is a time scale factor. Equation (5) only describes the time to perform a retrieval in matching a production. To this we have to add the time for the production's action which is routinely estimated in ACT-R at 50 ms (in line with other production system models—Anderson, John, Just, Carpenter, Kieras, & Meyer, 1995) or more if a physical action is required (e.g., moving visual attention, speaking, and typing).

ACT-R retrieves the chunk i with the highest match score M_i, provided that match score is above a threshold of activation, τ. ACT-R assumes the match scores have noise added to them that is distributed according to a logistic distribution (which is like a normal distribution). Because of the noise there is only a probability of any match score being highest and only a probability of it being above threshold. The actual predictions reported in this paper are obtained by adding random noise to activation (assuming a logistic distribution) and doing Monte Carlo estimations to determine the most active chunk and whether it is above threshold. However, it is useful to have some closed formed descriptions of these probabilities. The probability of a chunk with expected value M_i being above the threshold τ is

$$\text{Prob}_i = \frac{1}{1 + e^{(M_i-\tau)/s}}, \tag{6}$$

where s is related to the variance in activation, σ^2, by the formula $\sigma^2 = \pi^2 s^2/3$.

Equation (6) misses one important contribution to memory error which is retrieval of the wrong chunk through partial matching. Because of the mismatch penalty (Eq. (4)), a partially matching chunk is usually less active than a perfectly matching chunk but sometimes the ordering can reverse because of random fluctuations in activation levels. The following equation describes the probability of retrieving chunk i as a function of its match score M_i:

$$\text{Probability of retrieving } i = \frac{e^{M_i/t}}{\sum_j e^{M_j/t}}, \tag{7}$$

where the summation is over all chunks and t is related to the variance in activations, σ^2, by the formula $\sigma^2 = \pi^2 t^2/6$ and is related to s in Eq. (6) by $t = \sqrt{2}s$. This is the same as the Boltzmann equation (Ackley, Hinton, & Sejnowsky, 1985; Hinton & Sejnowsky, 1986) in which context t is called temperature. Note that both Eqs. (6) and (7) are approximate descriptions of the system. Equation (6) ignores partial matching and Eq. (7) ignores the effect of the threshold.

This completes the description of the basic ACT-R theory. The key ideas are captured by each of the seven equations above:

Equation (1): Chunk activation is the sum of a base-level activation and an associative activation.

Equation (2): Base-level activation will show the influence of practice and time-based decay.

Equation (3): Associative activation will depend on how many chunks are associated to a cue.

Equation (4): The match score of a chunk to a production is a function of its level of activation and its degree of match.

Equation (5): Latency in retrieving a chunk is an exponential function of its match score.

Equation (6): Probability of retrieving a chunk is the probability that its noisy match score will be above a threshold.

Equation (7): Probability of retrieving one chunk over others is the probability that its noisy match score will be the largest.

In subsequent sections we apply ACT-R models to simulate various experiments from the list memory paradigms. In all cases ACT-R predictions come from Monte Carlo runs of the computer simulation models. While the basic equations above characterize much of its behavior, there is always the potential for subtle interactions in the simulations which are not captured by the equations. Therefore, we have made all the simulations available on line and they can be reached by following the *Published ACT-R Models* link from the home page: http://act.psy.cmu.edu/. It is possible to change the parameters and run these simulations over the Web. The simulations are also capable of interacting with experimental software that can administer these experiments to subjects. These interactive simulations, which can be obtained by writing to the authors, are completeness proofs that the models specify all the processing involved in an experiment. One of the major goals in the ACT-R project is to achieve fully explicit theories of cognition that will yield the same computer traces that we see from human subjects interacting with the experimental software (Anderson, Lebiere, & Matessa, 1996). This reflects our dissatisfaction with theories (including past versions of ACT-R) that leave implicit aspects about how the theory relates to experimental data.

SERIAL MEMORY

The area of serial memory has had the longest history of research in psychology. It started with Ebbinghaus's interest in relatively permanent memory, evolved into an interest in transfer among lists, and most recently has been focused on theories of memory span. It has seen a fair amount of theory in the last third of this century (e.g., Baddeley, 1986; Burgess & Hitch, 1992; Conrad, 1964; Ebbinghaus, 1885; Estes, 1973; Lewandowsky & Murdock, 1989; Murdock, 1993; Shiffrin & Cook, 1978; Richman, Staszewski, & Simon, 1995; Wickelgren, 1965; Young, 1968). While we think the ACT-R theory is applicable to all types of serial recall paradigms, we will concentrate our presentation on the relatively immediate recall of relatively short lists as this is where most of the recent interest has been. Much of the recent theory has been dominated by Baddeley's use of the phonological loop to account for memory span which assumes that the amount that can be maintained in a memory span is the number of words that can be rehearsed in approximately 2 s. Evidence for this proposal comes from research showing that people can maintain fewer words that take longer to articulate—because the words either have more syllables or have syllables that are longer to articulate. In one very influential study, Baddeley, Thompson, and Buchanan (1975) looked at the number of words (out of five) which could be repeated back as a function of syllable length. Over the range of syllables from one to five, they found that this was approximately equal to the number of words that could be said in 2 s.

We (Anderson & Matessa, 1997) published an application of the ACT-R theory to the memory span task.[5] The models reported there were mathematical approximations to the ACT-R theory while here we will report the results from actual running simulations. This allows us to more adequately deal with the effects of rehearsal strategy and partial match-

[5] The theory was based on a slightly earlier version of the ACT-R theory than the one in this paper. The updated theory reported here also has a sufficiently efficient simulation that we are able to get predictions by Monte Carlo estimation.

ing. We could not always capture their effects in the Anderson and Matessa article with closed-form equations.

The ACT-R theory shares with Baddeley's theory an emphasis on time-based decay (based on base-level Eq. (2)). However, it also emphasizes important roles for associative interference (based on associative strength Eq. (3)) and for confusions among items in a list (based on partial matching Eq. (4)). In fact, there is good evidence for all of these factors as reviewed by Anderson and Matessa. Holding retention time constant, subjects perform worse when they must remember more items indicating associative interference. Confusions among items that are similar sounding (acoustic confusions) or are in similar positions (positional confusions) are a major fact of memory span performance. It is a challenge to be able to integrate these factors. In this section we will show that ACT-R is able to do this. Rather than reporting applications of ACT-R to past experiments as in Anderson and Matessa, we will show this with respect to some new data that we have gathered in our laboratory. These data were collected expressly to provide a powerful test of the predictions of the ACT-R theory about memory span.

An ACT-R Model of Serial Recall

One of the key issues in the history of research on serial memory concerns the nature of the representation of the serial list. Our assumption is that a list is organized as a set of groups and each group is represented as a set of items. Most critically, we assume that there is a chunk for each group encoding its position in the list and a chunk for each item encoding its position in the group. Positional coding, rather than associative chaining, has been advocated by a number of researchers (Burgess & Hitch, 1992; Conrad, 1965; Johnson, 1991; Shiffrin & Cook, 1978; Slamecka, 1967; Young, 1968). Figure 2 illustrates a possible representation for a list of nine digits grouped as 329 714 856. Each oval in Fig. 2 represents an ACT-R chunk. There is one chunk for each group and each element. A group chunk encodes the list the group is in, the size of the group, and its position in the list. Thus, the first group chunk encodes an item in position Group1 of Size3 in the list. This is indicated by pointers from *Group1, Size3,* and *List.* The elements are represented by chunks encoding the position of the element in the group, its group position in the list, the list it is in, and its content. Thus, for example, the first element 3 is encoded by a chunk with pointers to *1st, Group1, Three,* and *List.* Performance is going to depend critically on the retrieval of these chunks. Most critical is the link to the list context. There are so many links to the *List* context in Fig. 2 that we have had to merge them. However, in actual fact, *List* is the most unique index into each chunk.[6] Terms like *1st, Group1,* and *Three,* will appear in thousands of contexts. Thus, fan out of *List* becomes critical.

Retrieval of such chunks encoding list elements is orchestrated by a set of productions of which the most critical is the following:

Get-Next.
IF the goal is to retrieve the *n*th element of the *m*th group of the list
and *x* is the element at position *n* in group *m* in the list
THEN set a subgoal to process *x*
and move the pointer to the ($n + 1$)th position.

This rule assumes that each element is indexed by its position. Each element is then produced by the following production:

Type-Item.
IF the goal is to process an item
and the item is associated with a key
THEN type the key.

This production rule is specific to typing as the output mode since this is the output modality in the experiment to be reported. Similar

[6] The assumption here is that each list will have its own token. Perhaps to avoid ambiguity, we should have called this *List-7136.*

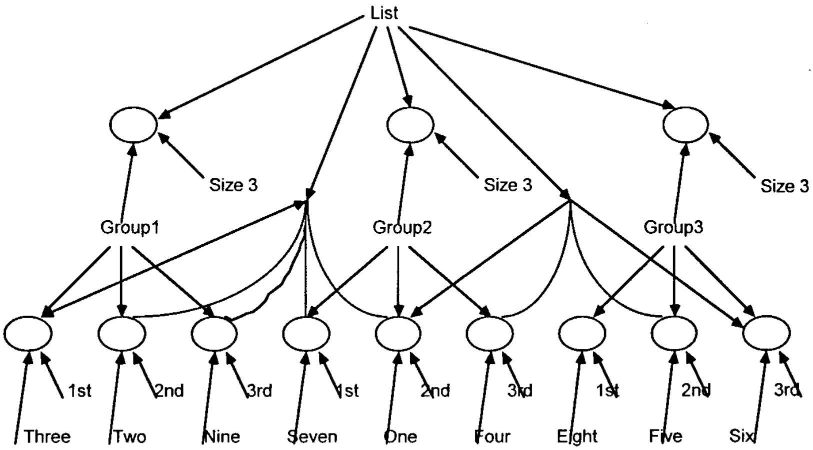

FIG. 2. A network representation of the chunk structure encoding the 9-element list ''329 714 856''.

productions could produce the response by means of written or verbal report.

In addition to retrieving the elements of a group, it is also necessary to retrieve the groups themselves:

Retrieve-Group.

IF the goal is to retrieve the *m*th group of the list
the best *x* is the group in position *m* of the list of size *s*
THEN set as a goal to retrieve the *s* elements of the group starting in position 1
and move the pointer to the $(m + 1)$th group.

The second line of this production retrieves the *m*th group and the size, *s*, of the group (note in Fig. 2 the size is stored with each group). The group size is important because it allows the system to know when to terminate recall of the group.

Note that it is a feature of both Get-Next and Retrieve-Group that they recall in the forward direction. This forward bias to serial recall played an important role in Anderson and Matessa (1997) and will be a critical feature in the experiment reported here.

According to the ACT-R theory, the critical factor determining speed and accuracy of retrieval will be the chunks that encode the group and item information. According to our activation Eq. (1) this activation will be a sum of base-level activation and associative activation. The base-level activation will in turn be determined by the amount of practice (through rehearsal) that these chunks received and how long ago those rehearsals were. As lists get longer the delays will tend to increase, thereby decreasing base-level activations (base-level Eq. (2)). The associative activation will come from the list element. As the list is longer, there will be greater interference because there will be more associations from the list element and less associative activation to any member of the list (associative strength Eq. (3)). Therefore, performance will go down with increased list length both because of increased delay affecting base-level activation and increased interference affecting associative activation.

While we will be presenting the results of Monte Carlo runs of our simulations, it is useful to have an equation which gives the approximate activation values that determine performance. Combining Eqs. (1)–(3) and using the approximation (Anderson, 1993) that

$$\sum^{n} t_j^{-d} \approx \frac{anT^{-d}}{1 - d},$$

where T is total time; we get the equation

$$A_i = \ln\left(\frac{anT^{-d}}{1 - d}\right) + W(S - \ln L),$$

where L is the list length. Collapsing constant factors and expanding, we get

$$\text{Activation} = B' + \ln n - d \ln T - W \ln L,$$

where B' reflects the constant factors, n is the number of presentations and rehearsals, T is time since presentation, L is the length of the list, d is the decay rate, and W is the attentional weighting of the list context. As noted earlier, we will set the decay parameter d to 0.5 throughout. The value of the W in this simulation is 1 since the list is the only useful source of activation. Thus, the effective equation for serial recall becomes

$$\text{Activation} = \ln n - 0.5 \ln \mathrm{T} - \ln \mathrm{L} \quad (8)$$

ignoring the constant factor B'. This equation is only approximate and we will be deriving our predictions from the running ACT-R simulation. This equation states that activation will increase logarithmically with number of rehearsals n and decrease logarithmically with delay T and list length L. The rehearsal and delay effects reflect base-level activation (Eq. (2)) while the list length effect reflects associative strength (Eq. (3)).

There is one additional important aspect to the ACT-R model of serial recall. We assume that the partial matching mechanism can result in positional confusions. Partial matching of the group slot will cause ACT-R to retrieve an element from a corresponding position in another group. Partial matching of the position slot will cause ACT-R to retrieve an element from another position in the current group. We assume that the degree of mismatch (P in partial-matching Eq. (4)) between elements is proportional to their distance apart. Thus, for instance, the mismatch between *First* and *Second, Second* and *Third,* or *Group1* and *Group2,* is *1D* while the degree of mismatch between *First* and *Third* or *Group2* and *Group4* is *2D. D* will be referred to as the scale factor for mismatches. This similarity-based confusion produces a majority of positional confusions between adjacent elements with only occasional confusions at greater distances. The existence of positional confusions within and between groups is well documented (e.g., Aaronson, 1968; Bjork & Healy, 1974; Lee & Estes, 1981; Nairne, 1992). We will not deal with acoustic confusions because the study to be reported involves digits which are not particularly confusable. Acoustic confusions are handled in Anderson and Matessa, again by the mechanism of partial matching.

Note that a positional element like ‘‘first’’ is just a cue for recall of an item just like ‘‘Montana’’ is a cue for answering the question, ‘‘Who is the governor of Montana?’’ Partial matching produces positional confusions in serial recall just as partial matching might result in confusion with the governor in Wyoming. In ACT-R the probability of such confusions is a function of the similarity of the cues (*first* versus *second,* or *Wyoming* versus *Montana*).

We will describe in more detail our exact model of serial recall and its consequences for base-level activation and associative activation. However, first we describe the experiment that we performed which will be the target of our modeling efforts.

A Study of Backward and Forward Recall

Like the Baddeley theory, the ACT-R theory claims that timing of the recall is important to memory performance and that with the passage of time memory chunks can decay to the point where they are no longer available for recall. However, in addition it claims that the state of activations of the memory chunks have a strong influence on the timing of recall. Thus, there is a feedback loop between timing and activations with higher activations yielding shorter retrieval times and shorter times yielding higher activations. Since there has not been research that has really delved into this interaction between timing and recall, we decided to perform an experiment that focused on this issue. We decided to look at memory span for digits, presented at the typical rate of 1 per s. We looked at list lengths from 3 to 12 digits to get a good range of performance and we measured both the timing and the accuracy of recall. We looked at both forward and backward recall to manipulate the delay between presentation and recall. In forward recall the first-presented digits are recalled first while in backward recall the last-presented digits are recalled first. This creates very different delays for the recall of digits in the same serial position during input. Also we decided to control the grouping of the digits by presenting them as visually segregated into units.

Method

Participants. Seventy-two subjects were recruited from the Carnegie Mellon University undergraduate population in an experiment that lasted about an hour. They were paid according to their performance in the experiment and they earned about 20 dollars. We wanted to see whether there would be any effect of our visual grouping. So 10 of the subjects were assigned to a control group for whom we did not try to control the grouping. The remaining 62 subjects had an indicated grouping and we will be mainly concerned with an analysis of their data.

Procedure. Participants went through 10 blocks of recall. During each block they studied and recalled one list of each length from 3 to 12 in the forward direction and one list of each length in the backward direction. Thus, they recalled 20 lists in each block. The order of the 20 lists within a block was randomized.

In the grouped condition we used the following groupings for the various list lengths:

3—one group of 3
4—one group of 4
5—two groups of 3 and 2
6—two groups of 3 and 3
7—two groups of 3 and 4
8—three groups of 3, 3, and 2
9—three groups of 3, 3, and 3
10—three groups of 3, 3, and 4
11—four groups of 3, 3, 3, and 2
12—four groups of 3, 3, 3, and 3.

Thus, we tried to keep the group size close to three and varied the length of the last group from two to four to accommodate various list lengths.

In the grouped condition a series of boxes would appear on the screen, one box for each group. The number of items that would appear in each box was made obvious by the number of spaces in the box. Thus, the subject knew immediately the number of and structure of the items to be studied. However, while they were studying them they did not know whether they would be tested for forward or backward recall.[7] The items were presented one at a time in their respective boxes in the appropriate spaces within the box. When one digit appeared, the other disappeared so that there was always just one digit visible. As soon as the last digit disappeared a signal appeared telling subjects the direction in which to recall the digits. The cursor would either move to the first slot in the first box (for forward recall) or the last slot in the last box (for backward recall). As each digit was typed the cursor moved to the next slot. The subject could skip over positions by typing a space or terminate the recall by hitting the return key. However, the subject could not back up and change recall of a digit. The procedure was the same in the ungrouped condition except that the items were presented and recalled in one large box with enough spaces for the list. Note in the grouped condition there was no temporal structuring to the grouping—it was purely a matter of visual organization.

Subjects were paid to motivate them to recall well. They were given a penny for each digit that they could recall in correct position. This was done to motivate them to recall as much as they could even when they could not recall everything. To give them extra motivation to try to recall the whole list they were given a bonus in pennies equal to the length of the list if their recall of a list was perfect. Their points were displayed at all times in the experiment.

[7] We wanted to test the same memory structure in different orders.

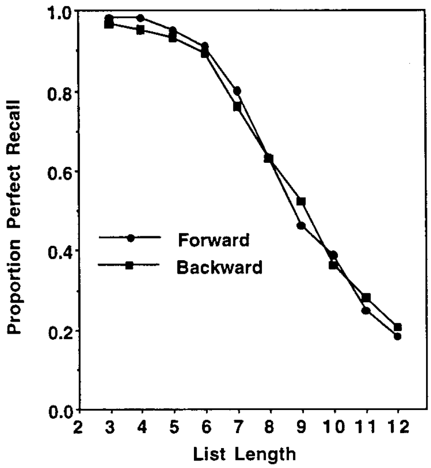

FIG. 3. Probability of perfectly recalling a list as a function of list length for forward and backward recall.

Results

First, we did an analysis of variance on the dependent measure of the percentage of perfect recalls of the lists as a function of length, of whether recall was forward or backward, and of whether the participant saw grouped lists or not. There was only a significant effect of list length ($F(9,621) = 128.05$; $p < .001$; $MSE = 439$). No other effects were significant and, in particular there was no interaction with whether the lists were grouped. Figure 3 displays the drop off in perfect recalls as a function of list length for forward and backward recall. The drop with list length is quite typical and it is not an artifact of averaging over subjects. Individual subjects also showed gradual drop-offs with list lengths. Our subjects were displaying about 50% recall for lists of length 9. This is rather higher than usual and may reflect their high levels of motivation because of the payoff. We find nearly equivalent recall in forward and backward direction, a result which has been obtained in at least some studies (e.g., Li & Lewandowsky, 1995). There is a slight hint of an interaction in the data to the effect that performance is better in the forward direction for short lists while performance is better in the backward direction for long lists. The length-by-direction interaction is marginally significant ($F(9,621) = 1.83$, $p < .10$, $MSE = 111.04$). A contrast comparing the short lists (3, 4, 5, 6, and 7) versus the longer lists (8, 9, 10, 11, and 12) with respect to this difference is significant ($t_{621} = 2.97$, $p < .005$). Subsequent analyses will look only at the grouped recall data.

Figure 4 shows the serial position curves for the various list lengths for the subjects

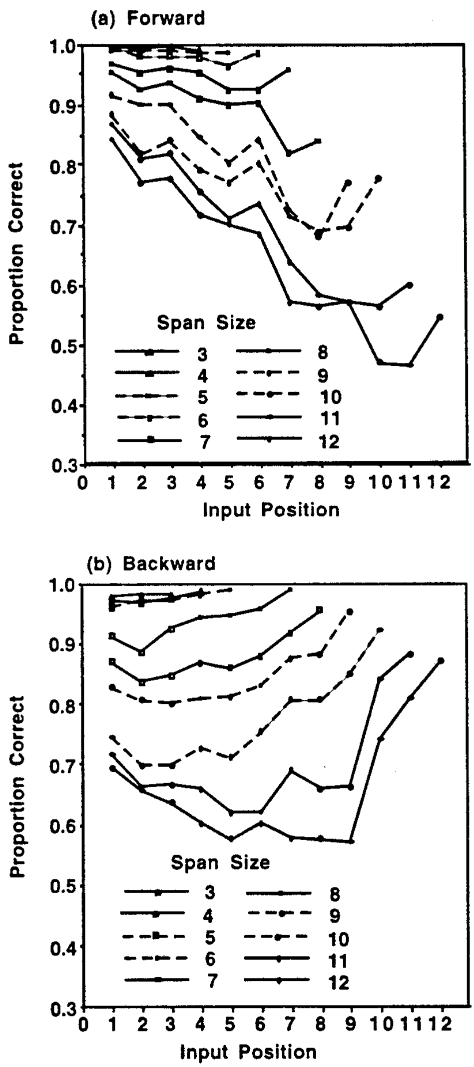

FIG. 4. Probability of correct positional recall for items recalled in the forward direction (a) and in the backward direction (b).

who experienced grouped lists.[8] Both Figs. 4a and 4b plot the probability of correctly recalling a digit in position as a function of serial position in input. For forward recall (Fig. 4a) the output order is the same as the input order but for backward recall (Fig. 4b) the output order is the reverse of the plotted input order. The forward recall curves are quite typical showing decreased accuracy with serial position. There is an upturn for the last item indicating a weak recency effect. Henson, Norris, Page, and Baddeley (1996) have argued that this one-item recency effect is basically due to decreased positional confusion regarding the last item. The backward curves largely show a weak primacy effect and a stronger recency effect spanning many items. Since input and output are reversed, this indicates better recall for the first items recalled just as in the forward recall. Such contrasting serial position curves for backward and forward recall are typical (e.g., Hinrichs, 1968; Metcalfe & Sharpe, 1985; Li & Lewandowsky, 1995). These curves also show some effect of the group structure. There are steep drops in the forward recall curves at positions 3 and 6 which are group boundaries. Thus, subjects show significant drops in recall from group 1 to group 2 and from group 2 to group 3. The group structure is less apparent in backward recall but there are precipitous rises from positions 9 to 10 in the lists of length 11 and 12 which correspond to the boundary defining the first group recalled.

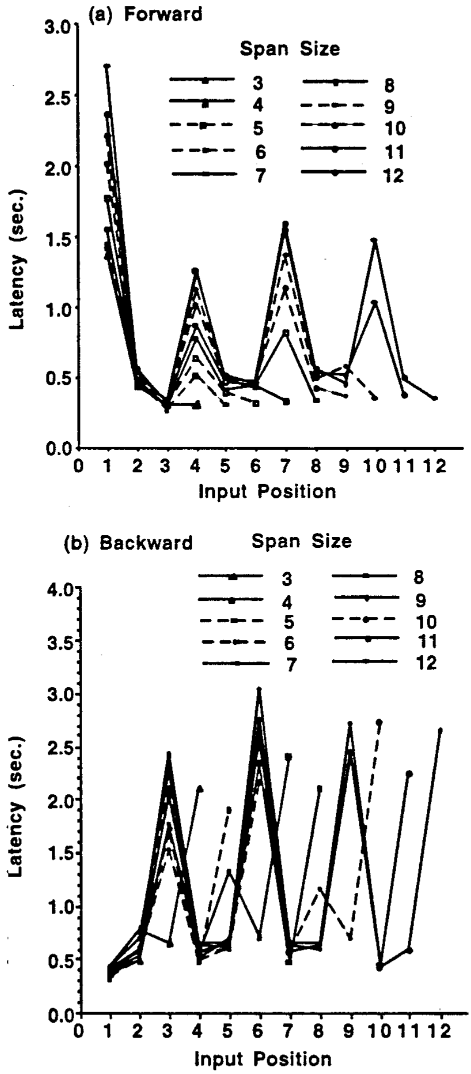

FIG. 5. Time to recall digits as a function of serial position and list length for (a) forward recall and (b) backward recall.

Figure 5 shows the times to recall the digits as a function of input positions. These are the means of the mean correct recall times for each subject. Here the group structure shows through very clearly. There are large spikes in the latency curves whenever subjects must begin recall of another group. In the case of forward recall the latency associated with recalling the first group is much longer that the other latencies, whereas in backward recall all group boundaries have comparable latencies. This suggests that in backward recall subjects start all over again with each group, whereas in forward recall they maintain some initiating information from group to group. Also in both forward and backward recall, the sizes of these spikes are very much a function of list length with longer lists resulting in larger spikes. In contrast, there is relatively little difference in within-group latency as a function of list length. This suggests to us that subjects are doing all of their recall for a group before

[8] The exact numbers for all these figures are available at the ACT-R web site indicated in the introduction.

typing any of the items in the group. In other research (e.g., Cowan, 1992; Sternberg, Monsell, Knoll, & Wright, 1978) it has been found that there is increased latency for individual items as a function of list length, but this research has not tried to control group structure. This may reflect the fact that different subjects used different structure and the data were averaged over different structures. Our failure to find much within-group effect of list length may also reflect the relatively high incentives given to our subjects which may have encouraged them to get each group straight before outputting it.

This one experiment has brought together many of the powerful effects documented in the memory span literature. The study indicates a rich pattern of data reflecting factors on like direction of recall, list length, serial position, and grouping. Overall level of recall was not very different in the forward versus backward direction. However, when we look at the serial position data in forward and backward directions (Figs. 4 and 5) we find strikingly different patterns. The data in these figures should serve as a substantial challenge to any theory including the ACT-R theory.

The ACT-R Simulation

We developed an ACT-R simulation of this task. Table 1 gives a trace of the simulation studying and recalling the nine-element list whose representation is given in Fig. 2. Part (a) illustrates the study process, part (b) illustrates the forward recall, and part (c) illustrates the backward recall. During study, ACT-R is interleaving study and rehearsal. The productions Attend-Start and Attend are responsible for encoding the digits as they appear on the screen. The first production, Attend-Start, encodes a digit at the beginning of a group, creating a chunk for both the group and the item. The second production, Attend, deals with the digits within a group.

The rehearsal strategy illustrated in Table 1a is one in which the system starts at the beginning of a list and keeps rehearsing until it comes to the end of the list. As the list keeps growing it takes longer to complete a rehearsal loop each time and usually it does not get to the end of the list in the last loop. Along with this linear rehearsal it interleaves rehearsal of the current item. Rehearse-Start initiates recall at the beginning of the list and Rehearse-Reset reinitiates recall at the beginning of the list when the current end of the list has been reached. The production Rehearse-Item is responsible for stepping through the items in serial order while Rehearse-Current is responsible for rehearsing the current item. These two productions compete and either is equally likely to fire next. This is a rehearsal strategy which is biased to rehearse the beginning of the list but has some probability of rehearsing all the members of the list. Rehearse-Abort stops rehearsal when a new item appears so that this new item can be encoded. Rehearse-Next-Group, which first appears on cycle 17, fires when one group has been rehearsed and switches rehearsal to the next group.

In forward recall (Table 1b), productions Retrieve-Group and Start-Group (a variant of Retrieve-Group for the first group) retrieve the group chunks. A production Dispatch-Three-Groups sets subgoals to retrieve the groups in the order encoded. For each group, the productions Get-Next and Get-Next-Start (a variant of Get-Next) retrieve the individual item chunks and Dispatch-Three-Items sets subgoals to type each in the order encoded. Then the production Type-Item types the individual digits. Note this scheme retrieves all the items in a group before typing any. This corresponds to the data we saw that indicated that our subjects tended to retrieve all the items in a group before typing any.

For backward recall, the rehearsal and encoding processes have identical structure since the subjects do not know how they will be tested. The structure of the recall (Table 1c) is the same except with respect to the productions that dispatch the subgoals. The production Dispatch-Three-Items-Backward is like Dispatch-Three-Items except that it sets the subgoals to type the items in opposite order. On the other hand, the productions for dispatching groups in backward order only subgoal one group at a time. In contrast the for-

TABLE 1

Study and Recall of ''329 714 856''

(a) Study

Cycle 0 Time 0.000: attend-start	Study 3
Cycle 1 Time 0.200: rehearse-start	
Cycle 2 Time 0.334: rehearse-item	Rehearse 3
Cycle 3 Time 0.879: rehearse-reset	
Cycle 4 Time 0.929: rehearse-item	Rehearse 3
Cycle 5 Time 1.462: rehearse-abort	
Cycle 6 Time 1.512: attend	Study 2
Cycle 7 Time 1.712: rehearse-current	Rehearse 2
Cycle 8 Time 2.246: rehearse-abort	
Cycle 9 Time 2.296: attend	Study 9
Cycle 10 Time 2.496: rehearse-current	Rehearse 9
Cycle 11 Time 3.105: rehearse-abort	
Cycle 12 Time 3.155: attend-start	Study 7
Cycle 13 Time 3.355: rehearse-item	Rehearse 2
Cycle 14 Time 3.891: rehearse-item	Rehearse 9
Cycle 15 Time 4.518: rehearse-abort	
Cycle 16 Time 4.568: attend	Study 1
Cycle 17 Time 4.768: rehearse-next-group	
Cycle 18 Time 4.942: rehearse-item	Rehearse 7
Cycle 19 Time 5.573: rehearse-abort	
Cycle 20 Time 5.623: attend	Study 4
Cycle 21 Time 5.823: rehearse-item	Rehearse 1
Cycle 22 Time 6.455: rehearse-abort	
Cycle 23 Time 6.505: attend-start	Study 8
Cycle 24 Time 6.705: rehearse-current	Rehearse 8
Cycle 25 Time 7.212: rehearse-abort	
Cycle 26 Time 7.262: attend	Study 5
Cycle 27 Time 7.462: rehearse-current	Rehearse 5
Cycle 28 Time 8.019: rehearse-abort	
Cycle 29 Time 8.069: attend	Study 6
Cycle 30 Time 8.269: rehearse-item	Rehearse 4
Cycle 31 Time 8.872: rehearse-next-group	
Cycle 32 Time 9.052: rehearse-abort-last	Rehearse 6

(b) Forward Recall

Cycle 1 Time 9.000: start-group	
Cycle 2 Time 9.782: retrieve-group	
Cycle 3 Time 9.905: retrieve-group	
Cycle 4 Time 9.992: dispatch-three-groups	
Cycle 5 Time 10.042: get-next-start	
Cycle 6 Time 10.244: get-next	
Cycle 7 Time 10.383: get-next	
Cycle 8 Time 10.644: dispatch-three-items	
Cycle 9 Time 10.694: type-item	Recall 3
Cycle 10 Time 11.198: type-item	Recall 2
Cycle 11 Time 11.701: type-item	Recall 9
Cycle 12 Time 12.208: get-next-start	
Cycle 13 Time 12.357: get-next	
Cycle 14 Time 12.656: get-next	
Cycle 15 Time 12.808: dispatch-three-items	
Cycle 16 Time 12.858: type-item	Recall 8
Cycle 17 Time 13.359: type-item	Recall 7
Cycle 18 Time 13.866: type-item	Recall 4
Cycle 19 Time 14.369: get-next-start-skip	
Cycle 20 Time 14.701: get-next	
Cycle 21 Time 14.986: get-next	
Cycle 22 Time 15.220: dispatch-three-items	
Cycle 23 Time 15.270: skip-item	Skip
Cycle 24 Time 15.770: type-item	Recall 5
Cycle 25 Time 16.276: type-item	Recall 6

(c) Backward Recall

Cycle 1 Time 9.000: start-group	
Cycle 2 Time 9.604: retrieve-group	
Cycle 3 Time 9.920: retrieve-group	
Cycle 4 Time 10.237: dispatch-three-group-backward	
Cycle 5 Time 10.287: get-next-start	
Cycle 6 Time 10.458: get-next-skip	
Cycle 7 Time 10.620: get-next	
Cycle 8 Time 10.837: dispatch-three-items-backward	
Cycle 9 Time 10.887: type-item	Recall 6
Cycle 10 Time 11.392: skip-item	Skip
Cycle 11 Time 11.896: type-item	Recall 8
Cycle 12 Time 12.401: start-group	
Cycle 13 Time 12.942: retrieve-group	
Cycle 14 Time 13.165: dispatch-two-group-backward	
Cycle 15 Time 13.215: get-next-start	
Cycle 16 Time 13.442: get-next	
Cycle 17 Time 13.634: get-next	
Cycle 18 Time 13.755: dispatch-three-items-backward	
Cycle 19 Time 13.805: type-item	Recall 4
Cycle 20 Time 14.306: type-item	Recall 1
Cycle 21 Time 14.809: type-item	Recall 9
Cycle 22 Time 15.315: start-group	
Cycle 23 Time 15.842: dispatch-one-group-backward	
Cycle 24 Time 15.892: get-next-start	
Cycle 25 Time 15.988: get-next	
Cycle 26 Time 16.241: get-next-skip	
Cycle 27 Time 16.493: dispatch-three-items-backward	
Cycle 28 Time 16.543: type-item	Skip
Cycle 29 Time 17.047: type-item	Recall 3
Cycle 30 Time 17.553: type-item	Recall 2

ward productions subgoaled all the groups. Therefore, when the group is completed in backward recall, the simulation must scan through all the groups in the list up to the to-be-recalled group. Thus, the structure for backward recall of the list in Fig. 2 is: recall group1, recall group2, recall group3, retrieve members of group3, recall group1, recall group2, retrieve members of group2, recall group1, and retrieve members of group1. This restarting is what produces recall latencies at the beginning of subsequent groups which are as long as the latency at the beginning of the first group. The backward protocol in Table 1 also illustrates failure to retrieve a couple of items. These are cases where the activation of the critical items randomly fell below threshold. It is also possible for the activation of the group chunk to randomly fall below threshold in which case the whole chunk will be skipped. The system is able to skip over the missing items and resume recall in place. It can use the visual structure of the recall display to know where to begin the next group.

In both the forward and backward recall it is the forward moving Retrieve-Group and Get-Item productions that are responsible for retrieving the items. Forward or reverse recall is achieved by subgoaling the items to be recalled in either forward or reverse order by different Dispatch productions.

The protocol in Table 1 illustrates another feature of ACT-R's recall which results from the partial matching. In the forward recall (Table 1b) note that the 8 from the first position of the third group is introduced as the first member of the second group. In addition, the 7 which is the first member of the second group is recalled in the second position of the second group. Finally, because the 8 is ''used up'' nothing is recalled in the first position of the third group.[9] In the backward recall note that the 3 and the 2 of the first group (recalled last) are reversed. These kinds of positional confusions are typical of serial recall and are produced by partial matching of the positional information.

The exact timings of the item recalls and their probabilities of success depend on random fluctuations in the activation levels. We ran 620 trials per condition to correspond to the 620 observations that we got from our subjects. This yielded fairly stable predictions. These predictions depend on a number of parameters that we had to set for this simulation:

1. The activation noise level, s, which we set to 0.300 (corresponding to a variance of 0.296)—see probability Eq. (6).
2. The activation threshold, τ, which was set to -0.35—see probability Eq. (6).
3. The time scale parameter, F, for retrievals which we set to 220 ms—see latency Eq. (5).
4. The scale factor, D, for mismatches which was set to 2.5.

Then there were a number of productions which were given nondefault action times (the default is 50 ms). These times were just set to plausible ballpark values:

5. The time to encode an item which was 200 ms. This is the ballpark time that we have established from the simulations of visual attention (Anderson, Matessa, & Lebiere, 1997).
6. The response time to type an item which was set to 500 ms.
7. The time to rehearse an item which was set to 500 ms to reflect speech time.
8. The initiation time to start recall, for the Start-Group production, which was set to 500 ms.

The last four parameters are constant across all the simulations reported in this paper while the first four parameters, s, τ, F, and D were estimated to produce a good fit to this experiment. However, our search for these four parameters was informal and there is no guarantee that we found the ones which produce optimal fits. The D parameter, reflecting positional similarity, is unique to this experiment but the other three s, τ, and F are potentially esti-

[9] Each token of a digit in the list has a tag as to whether it has been recalled or not. A digit token will not recalled again if it is tagged as already recalled. This is implemented as a test in the production Get-Next.

mated anew for each experiment. Table 4 at the end of the paper tracks all of the parameter estimates. At the end of the paper we will discuss the issue of variations in the s, τ, and F parameters across experiments.

Figures 6 and 7 show the resulting simulated recall behavior. The latency profiles in Fig. 7 capture the general structure of the profiles in Fig. 5. The overall R^2 between the two sets of latencies is 0.946. The accuracy profiles do not match quite as well, producing an overall R^2 of 0.906. Nonetheless, the correspondences between the profiles are quite compelling. We think this indicates some of the power of ACT-R to account for a complex data pattern. We are predicting 300 numbers only estimating four parameters and without carefully optimizing our fit.

To summarize what lies behind the ACT-R account of the data: The latency data speak to a very systematic group-by-group recall

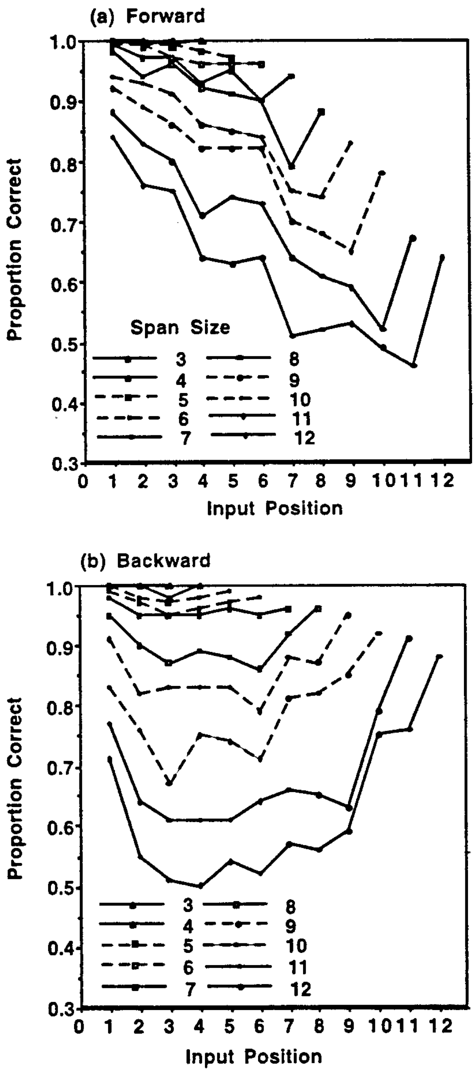

FIG. 6. Results of simulation: Predicted probability of correct positional recall for items recalled in the forward direction (a) and for times recalled in the backward direction (b). Compare with Fig. 4.

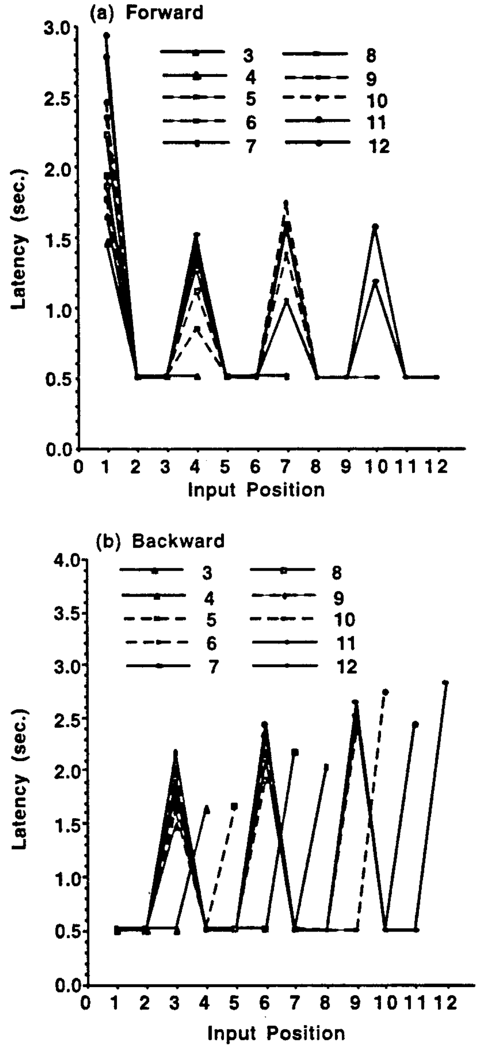

FIG. 7. Results of simulation: Time to recall digits as a function of serial position and list length for (a) forward recall and (b) backward recall. Compare with Fig. 5.

procedure that subjects are using to pace their recall. This is incorporated into the basic production rules that execute the task. This is one of the things we get from the control structure provided by ACT-R's production system. Within this relatively fixed procedure there is considerable variation in latencies at group boundaries as a function of list length. Also, there is considerable variation in recall of items both as a function of list length and input position. ACT-R does not change the fundamental algorithm to predict these variations. These variations reflect the changes in activations of the elements being retrieved. These activations increase with rehearsal (base-level activation), decrease with time (base-level activation), and decrease with list length (associative activation). Also, there are fewer positional confusions (partial matching) at the end of lists. These are all basic processes in the ACT-R theory and they combine to form the behavioral profile that we see.

Both time-based decay and associative interference are required to account for the span limitations. The very different recall profiles for forward and backward recall reflect differences in the time at which the same items are recalled. This difference between items at identical input positions is produced by time-based decay. On the other hand, in the backward data we can look at recall of items which have the same delay between study and test but which vary in list length. For instance, the last item is always recalled first after the offset in its presentation. However, as Fig. 4b illustrates, recall for this item systematically drops reflecting the contribution of associative interference from the other items. This difference between items with identical recall lag is evidence for associative interference.

In addition to associative activation and base-level decay, the rehearsal strategy assumed by ACT-R is critical. The tendency for the earlier items to receive greater rehearsal is one factor that is producing the primacy effect. The other factor is the lower positional confusions among items at the beginning of the list.

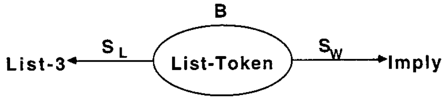

FIG. 8. A chunk encoding that the word ''imply'' has occurred in List-3.

RECOGNITION MEMORY

A different way of testing memory involves simply showing subjects the words in the list and asking them whether they recognize the items when they are mixed in with distractors (or foils) that they have not seen. Our model for this task is basically the one that Anderson and Bower (1972, 1974) developed 25 years ago where it is assumed that a memory trace is set up which encodes that an item occurred in a particular list. Thus, ACT-R records memory of the words in the list by means of chunks like the one illustrated in Fig. 8 which encodes that the word *imply* occurred in *List-3.* This is the same representation used in serial memory.[10] Recognition of a word is achieved by productions like

Recognize-a-Word.
IF the goal is to judge whether the word occurred in a context
and there is a trace of seeing the word in that context
THEN respond yes

This is a very straightforward model which views recognition memory as basically a simple process. The memory trace just consists of two items—the word and the list context. In recognizing a word, a subject has access to both sources of association (in contrast to free recall where the subject has only access to the list context). Thus, based on activation Eq. (1), the activation of a memory trace can be written

$$A = B + W_w S_w + W_L S_L$$

[10] In Fig. 8 we do not show encoding of position. As will be discussed, we will continue use of positional information for the short Sternberg lists but not for the longer lists.

based on activation Eq. (1), where W_w is the weighting given to the word, S_w is the strength of association from the word to the trace, W_L is the weight of the list context, and S_L is the strength of association from the list context to the trace. While the word becomes an important additional source of activation the W_wS_w term will remain constant across conditions. As in the case of the serial memory Eq. (8), we can expand the base level to show the effect of rehearsal time, decay, and list length

$$A = B' + \ln n - d \ln T - W_L \ln L,$$

where B' reflects constant effects including W_wS_w. Thus, just as in serial recall the critical variables remain the amount of rehearsal n, delay time T, and the list length L.

We can use the prior settings of the decay parameter, d, to 0.5 and assume an equal division of source activation between word and list context so that $W_L = 0.5$. Ignoring the constant, our equation becomes

$$\text{Activation} = \ln n - 0.5 \ln T - 0.5 \ln L. \quad (9)$$

This will be the critical activation equation for this section. This is identical to the serial memory equation [14] except that the list length is weighted by 0.5 reflecting the division of source activation (the Ws) between the list and the word. Again, this only gives an approximation to the results of the simulation and we will present data from simulation runs.

An additional relevant factor in recognition memory involves partial matching to either the word or the list context. Partial matching to the word will produce false alarming for similar words. There are ample experiments that show effects of distractor similarity on recognition memory (e.g., Anisfeld & Knapp, 1968; Underwood & Freund, 1968). Similarly, subjects are likely to false alarm for a word if it occurred in a similar list (e.g., Anderson & Bower, 1974).

In this section we want to focus on three results which have proven important in the past 25 years of research and theory on recognition memory. These are latencies to recognize items, particularly in the Sternberg paradigm, and the relationship between list length and list strength, and the Tulving–Wiseman function. One subsection will be devoted to each topic.

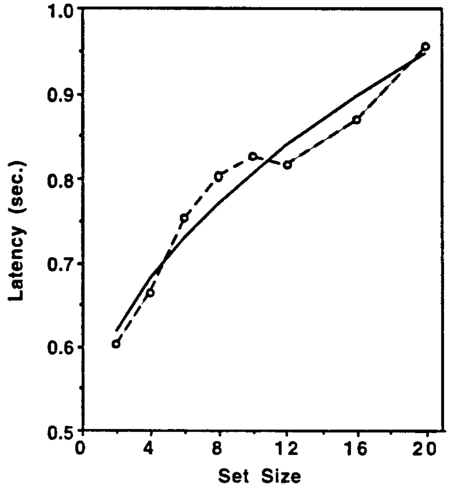

FIG. 9. Observed (dashed lines) and predicted (solid lines) latencies for recognizing probes as a function of set size. From Burrows and Okada (1975).

The Sternberg Paradigm

The Sternberg paradigm is one in which subjects see a relatively small list of items and then are presented with a single item and have to judge whether that item is from the list. As the result was originally described and is still described in many textbooks, the claim is that there is a linear relationship between the number of items in the memory set and time to make this judgment. In fact the relationship is more typically curvilinear and extends out to lists as long as 20 items in length (Briggs, 1974). Figure 9 shows some data from Burrows and Okada (1975) illustrating this relationship and a fit of our ACT-R model which we will describe below.

Since Burrows and Okada do not give us the details of presentation timing or practice, we simulated in ACT-R only the recognition judgment task and not the study. For this recognition judgment task we used a similar model as that which was used for our model of sentence recognition in studies of the fan effect (Anderson & Reder, in press) which seems appropriate for time-pressured recognition judgments. Upon presentation of a probe the subject retrieves the most active item from the list:

Retrieve-a-Candidate.
IF the goal is to recognize whether a word is in the list
and x is a word in the list
THEN consider x as the retrieved word.

This is a variant of Recognize-a-Word given earlier and its latency is determined by the same activation quantity (given by recognition memory Eq. (9)). If the probe is a word in the list, that candidate word will be most active since it receives activation from the probe. Thus, it is only necessary to retrieve one candidate. The subject then checks whether the retrieved item matches the probe and responds yes if it does:

Match-Word.
IF the goal is to recognize whether a word is in the list
and it matches the retrieved word
THEN say yes.

In the case of a foil, some list member will be retrieved only to be rejected as mismatching the probe by Mismatch-Word:

Mismatch-Word.
IF the goal is to recognize whether a word is in the list
and it does not match the retrieved word
THEN say no.

For its success, this scheme relies on the fact that if the word was studied, the chunk encoding its occurrence in the list will be more active than chunks encoding the occurrence of other items. This will be the case because this chunk will receive activation from the probe word.

Since the model for the Burrows and Okada data is so simple we can develop a simple mathematical equivalent to display its predictions by adapting the recognition Eq. (9). Since we do not have the details to model the study process, the number of rehearsals, n, and delay, T, in that equation become irrelevant and the recognition equation becomes

$$\text{Activation} = 3.76 - 0.5 \ln L,$$

where L is the length of the list and 3.76 is the level of activation of an element in a list of length 1 in the ACT-R simulation. The 3.76 reflects the default activation that elements get in this ACT-R simulation. This activation value and the F parameter trade off such that there is only one degree of freedom in fitting the data. Thus, we left the activation value at its default and estimated F at 4.15. Then using retrieval time Eq. (4) our prediction for latency is

$$\begin{aligned}\text{Time} = I + Fe^{-A} &= 0.5 + 4.15e^{-3.76}L^{0.5} \\ &= 0.5 + 0.097L^{0.5},\end{aligned}$$

where 0.5 is the fixed intercept I (reflecting encoding and response generation that will be used throughout this paper), F is the latency factor estimated at 4.15, and $0.097 = 4.15e^{-3.76}$.[11] The fit of this one-parameter model is basically identical to that of a logarithmic equation given by Burrows and Okada (which has two parameters) and is slightly worse than their bilinear model (which has four parameters). While this does involve the estimation of one parameter, it does make the parameter-free prediction that latency should increase as a function of the square root of list length. The data do correspond closely to this predicted form with an R^2 of 0.957 for the square root function as compared to 0.908 for a linear function.

To have better tests for ACT-R it would be useful to have data with better specification

[11] While we estimated F at 4.15, it would have been mathematically equivalent to use the equation above and estimate .097 as our free parameter.

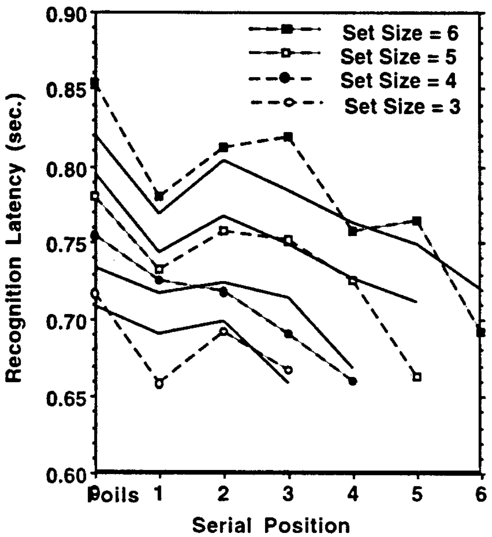

FIG. 10. Data (in dashed lines) from Raeburn (1974): Time to recognize an item for lists of various lengths as a function of serial position. The predictions of the ACT-R theory are in solid lines.

of the presentation timing and information about the effect of serial position on latency. Raeburn (1974) reports such an experiment in which items were presented at the rate of 1.5 s per item and tested at a delay of 1.2 s after the presentation of the last item. Figure 10 presents his data as a function of serial position of the targets plus the mean performance for foils.

We developed a running ACT-R simulation of these data. In modeling these data we want to carry over as much of the representations and processes as we can from our model in the previous section for serial memory given that the lists are of similar length and the timings are similar. We used the same productions for encoding and rehearsal and only changed the productions for making the recognition judgment. These recognition judgment productions are the same ones that we used for the Burrows and Okada simulation. As in the case of the Burrows and Okada data, we estimated the parameter, F, in fitting the data. This was estimated at 2.29 s. Also, only in this experiment did we have to estimate a nondefault intercept parameter—0.6 s rather than 0.5 s.

The quality of the fit is quite good with a mean error of prediction of 19 ms and an R^2 of 0.886 with just two parameters estimated. The data and the theory both reflect strong effects of target versus foil, set size, and serial position. It is worth reviewing what produces these effects in the ACT-R theory. The effect of target versus foil is due to the lower activation in the case of a foil for a trace retrieved by Retrieve-a-Candidate (because the probe does not provide source of activation to the retrieved trace). The effect of serial position reflects the effects of extra rehearsal of the beginning of the list and shorter recency of the end of the list. These are the same factors operating in the memory span except that time to recall in memory span is also affected by order of output in span tests. Finally, the effect of set size is due to the combined effects of decreased associative activation and increased decay of base-level activation since on average there will be longer delays between presentation and test.

Long-List Recognition Memory

The Sternberg task is one where lists tend to be short, accuracy is nearly perfect, and speed is the critical issue. As lists become longer the issue shifts to recognition accuracy. Errors on targets (false negatives) can be simply attributed to lower activation levels in ACT-R which result in failed recall of the target chunk. In the Anderson and Bower framework, false alarms would be attributed to subjects retrieving similar contexts to the target contexts and partially matching these contexts.

Our ACT-R simulation of long-list recognition tasks involves a number of changes from the simulation we used for the Sternberg recognition memory task. First, given the very long length of the list and given that this was not a serial memory test, we no longer tried to represent the serial order of the list and just stored the fact that the word occurred in the list (see Fig. 8). As a consequence we could not use serial position as a basis for rehearsing items but rather just rehearsed each item that came to mind (i.e., the item that is momentarily most active) a maximum number of times which we arbitrarily set to 2.[12] Thus, the critical rehearsal rule was

Rehearse.
IF the goal is to study the words in a list
and there is a trace of a word which has occurred in the list
and has not been rehearsed two times
THEN rehearse it one more time.

We used the following two rules for accepting targets and rejecting foils:

Accept.
IF the goal is to recognize if a word occurred in the list
and there is a trace of the word in the list
THEN accept it.

Reject.
IF the goal is to recognize if a word occurred in the list
THEN reject it,

where the rejection rule was rated lower in ACT-R's conflict resolution[13] (which means it will only fire if the acceptance rule fails). Thus, the Reject rule has to wait for the Accept rule to time out before it fires. We used this rather than the reject rule in the previous simulation which retrieved some trace because (a) latency was not critical, making the longer time-out latency for Reject acceptable and (b) sometimes no chunk was above threshold because of the longer list structure.

False alarms occur when the Accept rule partially matches to a memory of the word in another context. False negatives occur when no trace involving the word, including memory of it in the list, is above threshold and the Reject rule applies.

List Strength and List Length Effects

Recently there has been a considerable stir caused by what are called the list strength and list length effects on recognition memory (e.g., Ratcliff, Clark, & Shiffrin, 1990). Recognition memory for individual items deteriorates as the list has more items (the list length effect). It also gets better as items are studied more (the list strength effect). The interesting question is what happens when some items of the list are studied more often or longer and others are not. On analogy to the list length effect, one might imagine that if some items are studied more (this is analogous to making the list longer), the remaining items in a list would suffer greater interference. Just these effects occur with mixed lists in free recall where extra study for some items make others less available. However, there is no effect of amount of study of other items in recognition memory. Ratcliff et al. (1990) proclaimed that no extant theory of memory could accommodate these results. Since that time a number of theories (e.g., McClelland & Chappel, 1994; Shiffrin & Steyvers, 1997) have been modified or proposed to accommodate the result. It turns out that ACT-R is in this list of theories although we have to say this result was far from our mind when we proposed the ACT-R theory in 1993 and we have only recently realized that it explained these effects.

A representative experiment which captures these results is Experiment 4 reported by Ratcliff et al. (1990). In five conditions they had subjects either study 10 items 4 times (the 10-PS condition for ''pure strong''), 16 items 4 times (the 16-PS condition), 16 items with half presented 4 times, and half presented 1 time (the 16-M condition for ''mixed'' condition which contains 8 strong items, designated 16MS, and 8 weak items, designated 16MW), 16 items each presented once (the 16-PW condition for ''pure weak''), or 40 items each

[12] Also as a consequence, the *D* parameter for positional confusions is no longer relevant.

[13] ACT-R's conflict resolution theory is a mechanism for selecting a production rule when more than one is applicable. Different rules have different values and the highest valued rule applies first. If the selected rule fails to retrieve the needed information, as in the case above, the next highest rule will be tried and so on until one matches. ACT-R's conflict resolution theory plays a minor role in the simulations of the paper, essentially allowing us to order rules. Elsewhere it does play a critical role in models of strategy selection and choice (e.g., Lovett & Anderson, 1996).

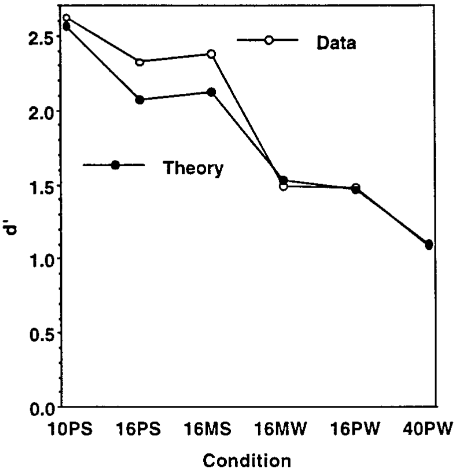

FIG. 11. Effects of list length and list strength from Ratcliff, Clark, and Shiffrin (1990). Data are the open circles and theory the filled circles.

presented once (the 40-PW condition). The results are displayed in Fig. 11 measured in terms of d' with the strong and weak items from the 16M condition plotted separately. It can be seen that for otherwise comparable lists there is a length effect, worse performance for longer lists. It can also be seen within the 16 item lists that there is a strength effect—performance is worse for words that are only presented once (*W* words are worse than *S* words). However, there is effectively no difference between items that come from mixed or pure lists. That is, holding list length and the strength of the target items constant, there is no effect of the strength of the other items in the list.

Figure 11 also presents the results from an ACT-R simulation of this experiment using the production set described in the previous section. In addition to the parameters that are held constant in all the simulations, we preset the latency factor F to be 2.00 which is comparable to its value in the Raeburn experiment.[14]

[14] In experiments where latency is not a dependent measure prediction, our predictions are only weakly dependent on the setting of F. Therefore, in these experiments we satisfied ourselves with ballpark setting and did not search for a good fitting parameter.

We encoded with each word three nonlist contexts which could serve as sources of false alarms. We estimated only the three parameters which influenced accuracy:

1. The activation threshold, τ, which was estimated to be 1.8.
2. The parameter, s, in Eqs. (6) and (7) which controls the noise in the activation values. This was estimated to be 0.55.
3. A P parameter for the partial matching penalty between list contexts (see partial matching Eq. (4)). This was estimated to be 2.0.

In general the correspondence between the theory and the data is quite good. The recognition memory Eq. (9) directly implies that list strength and list length will have an effect on recall. More study results in increased practice producing greater base-level activation. Longer lists result in greater fan producing less associative activation. In addition to these two major factors, there are two minor factors. First, longer lists also result in longer delays. Second, there is rehearsal borrowing in the mixed list conditions such that weak words get rehearsed at the expense of strong words. Rehearsal borrowing is weak in our simulation but real: There are an average of 2.00 rehearsals for 16PS words, 1.99 rehearsals for 16MS words, 1.96 rehearsals for 16MW words, and 1.61 rehearsals for 16PW words. These rehearsal effects are weak compared to the 1 versus 4 presentation difference. There has been a tendency to reject rehearsal borrowing in the literature. The best test of the rehearsal borrowing hypothesis is by Murname & Shiffrin (1991, Experiment 3) who compared mixed with pure conditions, controlling delay until test. The effects are all in the predicted direction of rehearsal borrowing but with nonsignificant ts of 1.08, 1.44, 0.86, and 1.39. We do not regard these weak effects as inconsistent with the weak degree of rehearsal borrowing that is occurring in the ACT-R simulation. It needs to be emphasized that the major factors at work are associative interference which produces list-length effects and differential practice (dominated by number of presentations not rehearsals) which produces list-strength effects. Both differential decay and differential rehearsal are much weaker effects in the simulation. ACT-R explains the differences among the conditions as follows:

10PS versus 16PS: The 10 PS condition is at an advantage because of lower associative interference and shorter lag to test.

16PS versus 16MS: These two conditions are equated on the important factors of list length (associative interference) and number of presentations (differential practice). The 16MS condition does suffer a very little from rehearsal borrowing but has a somewhat shorter delay until test.

16MS versus 16MW: The 16MS condition has advantages of number of presentations, differential rehearsal, and slightly shorter delays until test.

16MW versus 16PW: These two conditions are again equated on the important factors of associative interference and number of presentations. The 16MW condition has slightly more rehearsals but also slightly longer delays until test.

16PW versus 40PW: The 40PW items are at a disadvantage with respect to associative interference and delay until test.

Again we emphasize that, while differential decay and rehearsal borrowing are weak factors at work, the length–strength results really fall out of the ACT-R assumptions about increased associative inference with length and increased base-level activation with practice.

The Tulving–Wiseman Law

The theory we have described of recognition is basically the theory of Anderson and Bower (1972). It has been claimed that this theory was disconfirmed by various demonstrations of the failure to recognize recallable words (Tulving & Thompson, 1973). The typical version of this demonstration presents words for study in the context of weakly associated cues, tests recall of the words to these cues, and tests recognition in isolation. The phenomenon is that subjects cannot recognize some of the words they can recall. Far from this result contradicting the theory, ACT-R predicts that some items recalled in one memory test will not be recalled in another test. The basic explanation of this turns on the noise in base-level activations of memory chunks. Some of this noise is what Anderson and Lebiere (1998) call transient noise which varies from moment to moment. The other is permanent noise which is a one time perturbation of the base-level activations. For many applications the distinction between permanent and transient noise is not critical (and we have to this point just used a single aggregate noise) but the distinction becomes critical when one looks at repeated tests. To the extent that noise is permanent the same performance will be obtained on repeated tests but to the extent to which it is transient the performance will be uncorrelated.

We decided to investigate whether ACT-R could predict the Tulving–Wiseman Law which involves this relationship between recall and recognition (Flexser & Tulving, 1978; Tulving & Wiseman, 1975). This regularity involves plotting the relationship between $p(Rn|Rc)$, the probability of recognizing an item conditional on it being recalled to a cue, and $p(Rn)$, the probability of its unconditional recognition. Nilsson and Gardiner (1993) present data from 302 experiments reported by many researchers. This is reproduced in Fig. 12 along with the original function proposed by Flexser and Tulving:

$$p(Rn|Rc) = p(Rn) + 0.5[p(Rn) - p(Rn)^2].$$

To obtain ACT-R predictions about the relationship between $p(Rn|Rc)$ and $p(Rn)$, we need to specify four quantities relevant to the probability Eq. (6):

1. The activation (M_i in Eq. (6)) of the memory chunk encoding the word in the list. By varying this we can map out the recognition function in Fig. 12. We decided to vary this from -3 to 3, assuming a threshold, τ, of 0. With this threshold, an activation of 0 would map into 50% recall.

2. The permanent noise which we gave an s of 0.5.

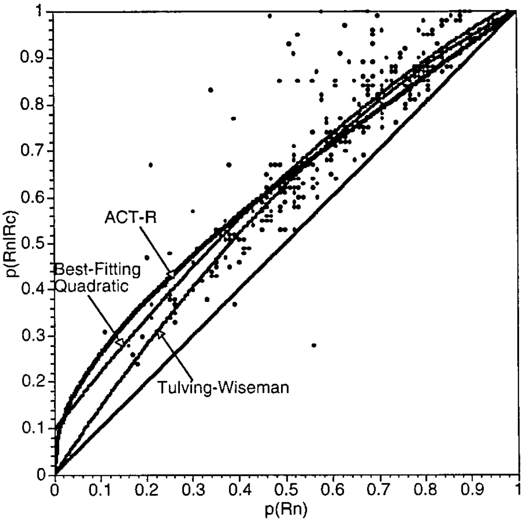

FIG. 12. Probability of recognition, $p(Rn)$, as a function of probability of recognition conditional on recall, $p(Rn|Rc)$. Each data point is a case from Nilsson and Gardiner (1993). Also plotted are the original Tulving-Wiseman function, predictions of ACT-R, and the best-fitting quadratic function.

3. The transient noise which we also gave an s of 0.5.
4. We had to model the effect of the different recall cues which would be differentially effective in prompting recall of the target. We decided to model this by adding another transient s of 0.5 for the recall condition.

The total noise s is the square root of the sum of the s's squared. Thus, recognition has an s of $\sqrt{0.5}$ and cued recall an s of $\sqrt{0.75}$.

These base-level activations and noise values were passed through the probability Eq. (6) to give probabilities of recall and recognition. Because of the common permanent activation these probabilities would be correlated but far from perfectly correlated because of the transient noise. The actual predicted function in Fig. 12 was obtained by Monte Carlo simulation (1 million trials per point) varying activation from -3 to 3 in 0.25 increments.

In general the function does fairly well in fitting the data. It deviates from the Tulving–Wiseman function in its not being symmetric around $p(Rn) = 0.5$ and being above the function for low values of $p(Rn)$. It is as much as 10% above in the range below $p(Rn) = 0.3$. However, there have been relatively few experiments which produce data points in this range and they do, in fact, tend to be above the Tulving–Wiseman curve. We have also reproduced the best-fitting quadratic equation to the data in Fig. 12 which is $p(Rn|Rc) = 0.095 + 1.284\, p(Rn) - 0.363\, p(Rn)^2$. As can be seen this best-fitting equation is raised above the Tulving–Wiseman function in the range of low $p(Rn)$ just as the ACT-R function. In fact, both ACT-R and the Tulving–Wiseman function have equivalent fits to the data with mean derivations of 0.085 as compared to 0.081 for the best-fitting quadratic function.[15]

Nilsson and Gardiner identify two classes of exceptions to the Tulving–Wiseman Law, both of which are associated with much higher-than-predicted values of $p(Rn|Rc)$. First, there are ''retrieval exceptions'' where the recall cue is present or effectively present at the recognition test as well. We would predict this class of deviations because in these cases the noise due to cue variability (quantity 4 given earlier) would be shared by the recall and recognition conditions. Second, there are the ''encoding exceptions'' when the cues are not associated to the target. Again we would predict this class of exceptions because there would be no cue variability to distinguish recognition and recall.

FREE RECALL

Free recall is an experimental paradigm where the subject is allowed to recall the items of a list in any order. The removal of the constraint of recalling in serial order may seem to simplify the task from serial recall. However, in fact it complicates the task substantially because it frees the subject to choose among a wide variety of strategies for studying items and recalling them. Some subjects repeat groups of words over and over again,

[15] We have not estimated an optimal ACT-R but neither is the original Tulving and Wiseman function an optimal fit.

other subjects look for associative or categorical relationships among the words, and still other subjects make up stories involving these words. It has been shown that the more organizational structure a subject tries to impose on the material the better memory they display (e.g., Mandler, 1967).

The generate-test model described in Anderson and Bower (1972) and realized in the FRAN simulation model (Anderson, 1972) was an attempt to extract the essence of these strategies. The basic assumptions of that model were:

1. The subject maintains a small set of about four items from the list which they rehearse and among which they try to find relationships. When a new item is encountered it enters this buffer and an old item is removed. This is basically the buffer model of Atkinson and Shiffrin (1968) with the added assumption that subjects search for semantic relationships among the items in the buffer.
2. At time of recall subjects try to generate candidate items, using among other things, the associative relationships that they have laid down.
3. Every time subjects generate a word they would then try to recognize it. Thus, the recognition process we discussed in the previous section was embedded as part of the recall process.

The SAM model of Raaijmakers and Shiffrin (1981) was another attempt to achieve an abstract characterization of this process. In that model traces were generated according to strength of association to the context and to the last retrieved item. There was a second stage in which an attempt was made to recover the word from the trace. The probability of this second stage was determined by overall activation rather than retrieval of contextual information as in the Anderson and Bower model.

In developing an ACT-R simulation of free-recall experiments we adopted a version that is simpler than either SAM or FRAN.[16] We assumed that subjects had a buffer of four elements for rehearsal and that, when a new element came in, the simulation randomly replaced a member of the buffer. This buffer was implemented by storing the four elements in four slots of the goal. After that the system randomly chose items to rehearse from the buffer as time allowed. At time of recall, the subject would first dump the members of the buffer, if it could,[17] and then recall items whose activation was above a threshold. Thus, if P_B is the probability the item is still in the buffer, the probability of recalling the item is approximately

$$\text{Probability of Recall} = P_B + (1 - P_B)P(\text{Activation} > \tau),$$

where τ is the threshold. The activation for an element would vary with the number of times it had been rehearsed, the length of time since those rehearsals, and the fan out of the list node (determined by list size). While we will present the results of actual simulations, it would be useful to have in hand an equation approximately giving the activation levels. In this case with a single source of activation (the LIST—the word is not presented), the operative equation is identical to the serial memory equation (8).

$$\text{Activation} = \ln n - 0.5 \ln T - \ln L, \quad (10)$$

where n is the number of encodings and rehearsals, T is the time since encoding, and L is the list length.

Serial Position Effects

One of the basic results about free recall is the serial-position curve which is a function giving probability of recall as a function of position of the item in the input sequence. Figure 13 shows some data gathered by Murdock (1962) looking at recall of lists that var-

[16] We do not deny the complexities. It is just that they are irrelevant to some of the principal effects in the literature that we will address here.

[17] Items in the buffer still have to be retrieved to produce their name and there is a small probability that they will not be sufficiently active for this retrieval to succeed.

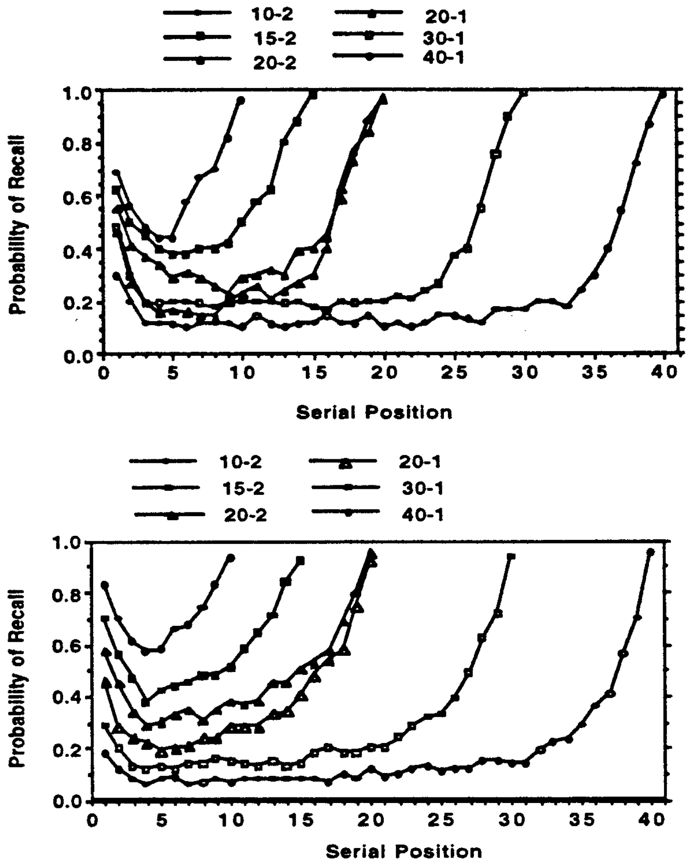

FIG. 13. Probability of recall of lists of various lengths (10, 15, 20, 30, 40) and amount of study time (1 or 2 seconds) as a function of serial position. (a) Data from Murdock (1962). (b) Predictions of the ACT-R theory.

ied in length from 10 to 40 words, presented at 1 or 2 s for each word. These results show the classic recency effect which is the high level of recall at the end of the list and the primacy effect which is the somewhat higher level of recall at the beginning of the list. The performance level is somewhat flat in intermediate positions with levels higher for shorter lists or for lists with more study time. Figure 13b shows the corresponding predictions of the ACT-R model. We preassigned the latency scale parameter F to have a value of 2.0 s for this experiment. The estimated parameters for this simulation were $\tau = 3.2$ and s (activation noise) $= 0.70$. The recency effect in ACT-R is produced by the probability that the item is still in the buffer plus the short delay in recall while the primacy effect is produced by the extra rehearsals given to the target item. The overall correspondence is quite good with an R^2 of 0.923 predicting 135 points.

Under this analysis the striking recency effect is due to both the decay of base-level activation and the tendency to rehearse and recall first the last few items of the list. In experiments where interfering activity is given after the list to wipe out the buffer, the advantage of the last few items disappear and they often show poorer recall, the so-called negative recency effect (Craik, 1970; Gardiner, Thompson, & Maskarinec, 1974). In experiments where an effort is made to eliminate rehearsal the primacy effect goes away and so does the negative recency effect when tested after a de-

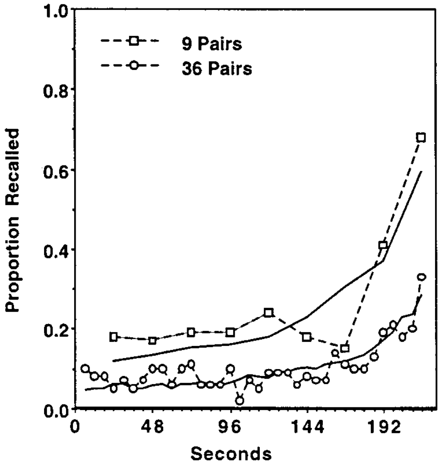

FIG. 14. Data from Glenberg, Bradley, Stevenson, Kraus, Tkachuk, Gretz, Fish, and Turpin (1980) in which interspersed arithmetic was used to eliminate use of a rehearsal buffer. The dashed lines are the data and the solid lines are the predictions of the ACT-R theory.

lay (Baddeley, 1986). In such studies, where subjects are prevented from forming a rehearsal buffer and forced just to process the item under study, one also sees a diminished positive recency effect. Performance tends to drop off continuously from the end of the list. These studies are perhaps the cleanest studies of free recall because they both eliminate the buffer and differential practice.

Figure 14 shows some data from an experiment by Glenberg, Bradley, Stevenson, Kraus, Tkachuk, Gretz, Fish, and Turpin (1980) which is of this kind. Subjects studied pairs of words for 2 s. In one condition they studied 36 such pairs (for 72 items) each preceded by 4 s of distracting activity while in the other condition they studied 9 such pairs (for 18 items) each preceded by 22 s of distracting activity. The experiment was designed so that subjects spent 216 s studying the list in both conditions. The effect of the distraction was to prevent any cumulative rehearsal and force subjects just to attend to the presented items. In both conditions there was 20 s of intervening activity before recall. As can be seen the recency effect is reduced (no longer are subjects recalling the last item 100% of the time) and there is no primacy effect to speak of.

We ran an ACT-R simulation of that data. We took a very extreme interpretation of the intervening activity and assumed that it eliminated all rehearsal. Thus, we assumed each item had a single study when presented and no rehearsal. Recall began 20 s after the last presentation. We assumed that the subject had no rehearsal buffer and that all recall was determined by the activation of the chunks in declarative memory. Figure 14 presents the predictions of the ACT-R model with $s = 0.60$ and $\tau = 1.4$. The fit is quite good with an R^2 of 0.860. Note ACT-R shows a large recency effect in this experiment even though it has no buffer to dump. Thus, the recency function is not just a consequence of buffer dumping. It is more fundamentally a consequence of the decay built into base-level Eq. (2).

The Ratio Rule

Our fit of the Glenberg et al. (1980) serial position curves raises the issue of whether ACT-R can predict the regularity noted by Glenberg, Bradley, Kraus, and Renzaglia (1983) in the serial position curve. This regularity is that recency varies roughly as a function of the logarithm of the ratio of the inter-item presentation interval (IPI) to the retention interval (RI). In Fig. 13 note that the serial position curve is steeper for the 9 pairs condition where the IPI/RI ratio is 24/20 than the 36 pairs condition where it is 6/20. The measure of recency proposed by Glenberg et al. is the slope from the last to the third-back item in a list. They showed that this is roughly a linear function of the logarithm of the IPI/RI ratio. This has been used to argue that the serial position curve reflects a temporal discrimination process and not a simple decay process.

It turns out that ACT-R does predict a systematic relationship between recency and the IPI/RI ratio. To see this relationship, it is useful to consider the odds form of Eq. (6)

$$\text{Odds of Recall} = e^{(M_i - \tau)/s}.$$

The M_i can be replaced by our free recall Eq. (10) in which case we get

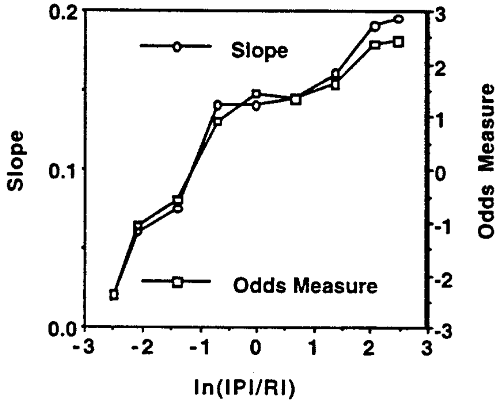

FIG. 15. Data from Nairne, Neath, Serra, and Byun (1997): Two measures of recency as a function of ln(IPI/RI).

$$\text{Odds of Recall} = C*T^{-r},$$

where T is delay, $r = .5/s$, and C is the constant part that does not depend on delay $C = e^{[\ln(n)-\ln(L)-\tau]/s}$.

ACT-R makes a prediction about the ratio of odds of recall at various serial positions. If we compare the final (f) item and two-before-final ($f - 2$) item we get the following prediction:

$$\text{Odds-Ratio} = \frac{\text{Odds}_f}{\text{Odds}_{f-2}} = (T_{f-2}/T_f)^r,$$

where T_f is the delay for the final item and T_{f-2} is the delay for the two-before-final item. If we replace T_f by the retention interval RI and T_{f-2} by $RI + 2*IPI$. We can transform the above rule into

$$\text{Odds-Ratio} = [1 + 2(IPI/RI)]^r.$$

Thus, the odds ratio is predicted to be a function of the *IPI*/*RI* ratio.

The regularity predicted by ACT-R involves a different measure than the slope proposed by Glenberg et al. and is a different function of *IPI*/*RI* ratio. Nonetheless, it does predict the importance of the same critical quantity. Moreover, if we assume $s = 0.5$ (a common value), then $r = 1$ and we can convert the above into a linear function of ln(*IPI*/*RI*), which is the independent measure advocated by Glenberg et al.

$$\ln \frac{\text{Odds-Ratio} - 1}{2} = \ln\left[\frac{\text{Odds}_f - \text{Odds}_{f-2}}{2\ \text{Odds}_{f-2}}\right]$$
$$= \ln(IPI/RI).$$

It would be instructive to compare the two dependent measures, Glenberg's slope measure and the odds measure, $\ln[(\text{Odds}_f - \text{Odds}_{f-2})/2\ \text{Odds}_{f-2}]$, in fitting data reported by Nairne, Neath, Serra, and Byun (1997). These data involve wide variations of the *IPI*/*RI* ratio and are quite reliable. Figure 15 offers this comparison and it makes the point that the two measures are highly correlated ($r = .989$). The slope measure is more intuitive and statistically better behaved. For instance, if it should happen by chance that performance on $f - 2$ is better than performance on f (as in the case in one condition in Glenberg et al., 1980), then the odds measure is not defined whereas the slope measure is just negative. On the other hand, the odds measure, being theoretically derived, is more sensible. The linear relationship proposed between log(*IPI*/*RI*) and slope would imply negative slopes for small enough ratios and slopes greater than 1 for large enough ratios. It also implies that level of performance should be the same for any *IPI*/*RI* ratio, independent of absolute duration. In contrast, Nairne et al. show, as ACT-R would imply, that absolute performance goes down with longer intervals, reflecting the greater forgetting.

Nairne et al. (1997) propose a combination of Estes' (1972; Lee & Estes, 1977) perturbation model and Neath's (1993) dimensional distinctiveness model to predict their data. In their model positions move with time and distinctiveness is lost. Thus, we hardly want to imply ACT-R is the only theory capable of accounting for this data. However, it is important to note that it is embedded within an existing theory and is a direct consequence decay factor in one of the basic equations (base-level Eq. (2)) of that theory.

List Length and Study Time Effects

We have discussed list length and study time effects in recognition. In free recall researchers have also examined how memory for list items increases with the length of the list or the amount of time per item. Figure 16 shows some data from Roberts (1972) displaying how number of words recalled increase as the list length increases from 10 to 40 items and as study time increases from 0.5 to 8 s per item. As Fig. 16b shows the ACT-R model (same simulation as in the Murdock experiment—see Fig. 13), it does a good job in accounting for this pattern of data. The parameters in this model fit were $\tau = 2.9$ and $s = 0.85$. The correspondence is particularly good and the overall $R^2 = 0.990$.

The simulation is accounting for the full pattern of data. As study time increases rehearsal increases and memory improves. As number of words increase, fan out of the list increases and so proportions of recall decrease. These are the standard list length and study time effects. However, there is a subtlety concerning total study time. As list length increases, total number recalled increases even as proportion recalled decreases. This reflects the benefit of increased study time and the cost of increased interference. While the relationship is not perfect, total recall tends to be strongly related to the total study time. ACT-R reproduces all of these effects.

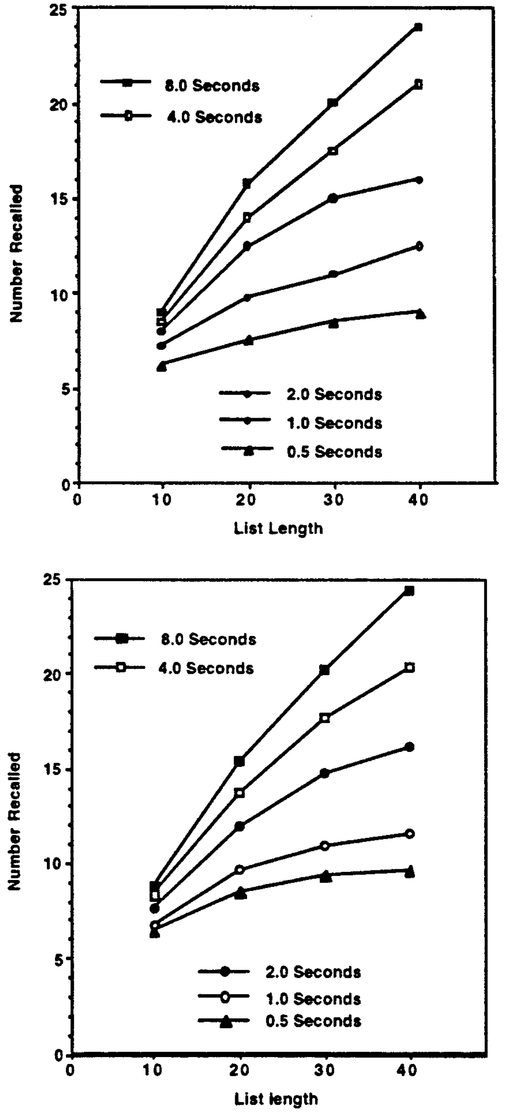

FIG. 16. (a) Data from Roberts (1972) showing how number of words recalled increases with list lengths and study time; (b) predictions form ACT-R.

IMPLICIT MEMORY

The most recent domain of interest in list-learning experiments has been implicit memory. In general, such experiments involve demonstrations that subjects are facilitated in their memory for words in ways that they do not realize. Many of these demonstrations involve perceptual facilitation. For instance, subjects may be able to read faster words that they have studied in a list even though they may not be able to remember seeing these words (Feustel, Shiffrin, & Salasoo, 1983; Johnston, Dark, & Jacoby, 1985; Johnston, Hawley, & Elliott, 1991; Watkins & Gibson, 1988). Other work (e.g., Hayman & Tulving, 1989; Jacoby, Toth, & Yonelinas, 1993) on implicit memory involves word fragment completion. For instance, subjects might study a word like HARE. Then, later they will be tested with a fragment like H_R_ and be asked to complete it as a word. They are more likely to complete it as HARE after seeing the word than other possible completions like HIRE or HURT. Sometimes they will show this tendency even when they are explicitly instructed not to complete the fragment with the word they have studied. They make these ''errors'' because they do not explicitly remember seeing the word but their system has been implicitly affected by seeing the word. One of the reasons for excitement about this research is that some types of amnesic patients show normal levels of implicit memory but show almost no explicit memory for the words (Graf, Squire, & Mandler, 1984).

There have been a number of attempts to account for these results in terms of activation-based network models of memory (Bower, 1996; Reder & Gordon, 1996; Reder, Nhouyvanisvong, Schunn, Ayers, Angstadt, & Hiraki, 1997; Reder & Schunn, 1996). The fundamental idea is that declarative network structures can be left in heightened states of activation as a resulting of processing. These heightened states can facilitate later processing. The ''implicit'' memory is these heightened states of activation. In contrast, explicit memory requires adding and retrieving new declarative network structures rather than just priming existing ones. This basic idea of other researchers can be easily incorporated into the ACT-R system and in the next few pages we will describe how this can be achieved. This idea is that recent access to knowledge will increase its base-level activation and so make it more accessible the next time. While we could extend this idea to priming of conceptual information we will focus on priming of lexical information. Typical tasks involve naming a word, judging whether a word is correctly spelled, or completing a word fragment. In all cases the subject must get access to information about word spelling.

It should be appreciated that the decay function for base-level activation (Eq. (2)) produces extremely slow decay at long delays. Thus, the claim that priming is due to increases in base-level activation is not incompatible with demonstrations of priming at long delays (e.g., Sloman, Hayman, Ohta, Law, & Tulving, 1988).

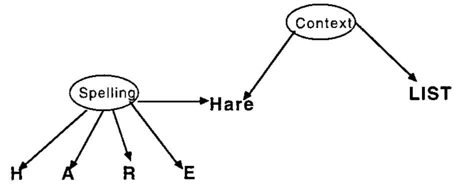

FIG. 17. A representation that encodes lexical information in a spelling chunk and list information in a context chunk.

Relationship between Word Naming and Word Recognition

Figure 17 shows the basic ACT-R representation for a word, its spelling, and its occurrence in a list. The letters are stored as part of a spelling chunk and the word is stored as having occurred in the list as part of a context chunk. There would be a separate context chunk for each context the word occurred in. If we were to deal with conceptual priming we would have to elaborate this representation to include a distinction between words and concepts (as is done by Bower, 1996), but this is not done in Fig. 17 for the sake of simplicity.

Basically, reading the word requires retrieving the word from the letter representation. The operative production is

Read-Word.
IF the goal is to read a word consisting of the letters L1, L2, L3, and L4
and L1, L2, L3, and L4 spell Word
THEN say Word.

Reading a word will strengthen the base-level activation of the chunk encoding the word's spelling. As a consequence, the next time that word is presented or a fragment of the word is presented the subject will be faster and more likely to access that chunk. Reading a word will also cause a chunk to be formed encoding that the word occurred in the list. This new chunk will serve as the basis for a recognition memory judgment.

We attempted to model the experiment of Johnston et al. (1991) as one test of ACT-R's theory of word reading. Subjects in their experiment studied 96 four- or five-letter words at the rate of one word per 2 s. The first and last four words were buffers but the middle 88 were critical. Subjects were then tested with 206 words which consisted of a buffer of 30 words followed by the 88 target words mixed in with 88 foils. The words were presented for recognition camouflaged in dots which either disappeared at a slow rate or fast rate. The subjects were to read the word as fast as they could. In our simulation the reading of the word was governed by the Read-Word production above. The actual timing consisted of three components. There was the time to encode to the letters (estimated at the default of 200 ms during study, 500 ms during test in the fast uncovering condition, and 700 ms during test in the slow uncovering condition), the time to retrieve the word in the Read-Word production, and the time to say the word (given a standard estimate of 500 ms). After reading the word, ACT-R recognized the word using the same productions as used in the recognition section to model the length–strength effect. ACT-R simulated 32 subjects, the same number as in the Johnston et al. experiment. The nondefault parameter settings were the encoding times (set longer than usual to reflect the difficult encoding conditions), the latency scale parameter F set at 1.3 s, the activation noise parameter s set at 0.65, and the activation threshold parameter τ set at 0.9.

Table 2 presents the data and the simulation results broken down according to whether the word was a target or a foil, whether the word was recognized or not, and whether the uncovering was fast or slow. The correspondence between data and theory is good. The R^2 between simulation and data are 0.921 in the case of the percentages and 0.910 in the case of latencies. There are two relevant effects apparent in the data and the simulation. First, ACT-R is faster at reading words that it has seen before. Second, it is faster at recognizing words that it thinks it has seen. Note that, even if the model cannot recognize the word, it is faster at reading it if it has seen it. This is the dissociation of implicit from explicit memory. This occurs because the spelling chunk supports the reading process while the context chunk supports the recognition judgment. The context chunk can be low in activation without the spelling chunk being low in activation.

TABLE 2

Reading Time and Recognition Proportions for Words: Data and Simulation

	Fast uncovering	
	Recognized	Not recognized
Target	Hits 1313 ms (1324) 73% (75%)	False negatives 1521 ms (1446) 27% (25%)
Foil	False alarms 1455 ms (1531) 39% (37%)	Correct rejections 1670 ms (1724) 61% (63%)
	Slow uncovering	
	Recognized	Not recognized
Target	Hits 1481 ms (1531) 72% (72%)	False negatives 1737 ms (1658) 28% (28%)
Foil	False alarms 1789 ms (1718) 35% (36%)	Correct rejections 1876 ms (1918) 65% (64%)

Note. Data from Johnson, Hawley, and Elliot (1991). Simulation is in parentheses.

However, it is also interesting to note that ACT-R is slower to read words which it does not recognize. It is approximately as fast at reading a word that it has seen before but it thinks it has not seen (a false negative) as a word which it has not seen before and thinks it has (a false alarm). The overall ordering of conditions in terms of reading times is Hits < False Negatives, False Alarms < Correct Rejections. It has been argued that this is evidence that subjects use perceptual fluency as a basis for performing word recognition, i.e., that they tend to say that they recognize words which they can read more rapidly. This claim has come in for considerable dispute. For instance, Poldrack and Logan (1997) argued that, while there is a difference in the latencies for targets and foils, there is so much overlap in the latency distributions that subjects would be very poor in the recognition judgments if they used their reading times as a basis for making recognition judgments. Watkins and Gibson (1988) have argued that the correlation between identification time and recognition performance is due to item selection effects such that items which are more easily identified also tend to be more recognizable.

It turns out that ACT-R's explanation for why the reading latencies are shorter for yes's than no's is a variation of the item selection argument. To understand what is happening in ACT-R it is important to appreciate the contribution of associative strengths to the reading times and to recognition accuracy. Since the letters are part of the goal, associative activation from the letters influence both the reading times and the recognition judgments. Reading time will be determined by the S_{ji}s from the letters to the word chunk and the recognition judgment will be influenced by the S_{ji}s from the letters to context chunk. Both of these associations involve the same sources (js, the letters) but different chunks (is, either word or context chunk). Basically, because of the fan effect (associative strength Eq. (3)) some sources will have weaker associations than other sources. That is, some letters appear in relatively few words and so will have strong associations to both these words and the word contexts. This means that words that are read faster are also more likely to produce recall of contexts (resulting in hits if it occurred in the study context and false alarms if occurred only in other contexts).

It should be noted that in Experiment 5 Johnston et al. tested identification and recognition in two separate tests and, in contrast to their data in Table 2, they failed to find a relationship between latency and judged status. This failure has been used to argue against the item selection hypothesis since that hypothesis does not require that identification and recognition be tested together. However, the procedure in their Experiment 5 used a ''mock subliminal paradigm'' in which no words were actually presented for study. Thus, there is no memory chunk to be primed by the letters and ACT-R would not predict a correlation between latency and judged status.

TABLE 3

Proportion of Fragments Solved and ACT-R Predictions Are in Parentheses

	Studied words		Nonstudied words	
	Test 1	Test 2	Test 1	Test 2
Same fragments in Test 1 and Test 2				
Conditional	.32 (.34)	.17 (.16)[a]	.20 (.19)	.09 (.08)
Control		.36 (.33)		.18 (.19)
Different fragments in Test 1 and Test 2				
Conditional	.35 (.36)	.34 (.35)[a]	.19 (.18)	.19 (.19)
Control		.37 (.32)		.17 (.17)

[a] Test 2 responses are Conditional on failure in Test 1.

Fragment Completion

Hayman and Tulving (1989) reported a study on fragment completion which shows some of the subtlety of implicit memory tests. Subjects studied lists of 160 words. These included 16 primacy fillers, 64 critical words, and 80 recency filters. Subjects performed two fragment completion tests. The first test involved 32 words which the subject had studied plus 32 words not studied. The second test involved the other 32 studied words, 32 new nonstudied words, plus whatever words the subjects had not successfully completed in the first test. These repeated words were either tested with the same fragment or a different fragment. For instance, if the original word was aardvark the two fragments were either a--d--rk or -ar-va--. Note the two are complementary. The results and our simulation of these results are displayed in Table 3. On Test 1 subjects show better memory for studied words than new words which is a standard priming effect. On Test 2 they show poorer memory for the words that they failed to complete in Test 1 when they are retested with the same fragment but not when retested with a new fragment. Performance with a new fragment is as good as the control performance on words that had not been tested in Test 1.

The simulation modeled the data assuming subjects tried to retrieve the words by means of productions like:

Complete-Fragment-1-3.
IF the goal is to complete a word consisting of the letters L1_L3_
and L1, L2, L3, and L4 spell Word
THEN say Word,

where L*i* refers to the *i*th letter. Probability of firing this production will depend on the base-level activation of the word plus the strength of association from L1 and L3 to the word. Both of these will be increased by recent exposure to the word. Thus, there will be a higher probability of completing the word if it had been studied. We set the latency scale parameter F to 0.5 in fitting this data and estimated s at 0.3 and τ at -0.45.[18] The other parameters were kept constant. The correspondence between theory and data is clearly quite good with an overall R^2 of 0.962.

There are basically two results in the data. First, subjects and simulation are almost twice as good at completing words they have studied. This shows the strengthening effect of that experience. Second, there is an effect of how subjects are retested on words they missed. They are worse when retested in the same way with the same string on which they have failed. This is because one is conditionalizing on items which are likely to have low base-level activations and associative strengths. However, when tested on different fragments, subjects recall conditional on failure is no worse than their completion of new fragments. The conditional performance on the different fragments reflects a plus–minus effect. That is, as in the case of the same fragment condition, we are conditionalizing on lower base-level activations which should result in lower recall. On the other hand, we are testing with new letters which might have stronger associations. Indeed, since the two fragments are complementary only one will have the first letter which is probably the best cue for the word. We represented this in our model by setting the strengths of association from the first letter much higher than that from other letters (4.0 versus 0.4). Our setting of associative strengths was basically determined to produce equal recall of the different fragments in the conditional Test 2 as control fragments in Test 2 (same or different). So our model does not really predict this effect although it does predict better conditional performance on Test 2 for different fragments than for same fragments.

[18] The activation noise was equally partialed between permanent and transient activation noise. See our earlier discussion of these two types of activation noise in the subsection on the Tulving-Wiseman Law.

REFLECTIONS ON THE MODEL FITTING ENTERPRISE

In a recent paper critical of model-fitting efforts Roberts and Pashler (submitted for publication) complained that models are not applied across a range of experiments. We would submit this effort as one of a number exceptions to this assertion. While these are all list memory experiments they reflect a substantial variation in the testing procedures. There are a number of questions one can ask when a single theory fits a range of experiments. Among these are whether the variations in parameter estimates across the experiments are reasonable and whether there are any hidden assumptions in the models. We will address these issues of parameter regularity and hidden assumptions.

Parameter Variation

Table 4 is a summary of the parameters used in fitting the models to the experiments and the proportion of variance accounted for. Except for the s, τ, and F, the parameters have been basically held constant across the experiments. The variations in s, the noise parameter, are rather small. One reason for variation in this parameter is that it will reflect heterogeneity in the population of subjects and variability in items and how subjects attend to them. The more variable they are the larger this parameter which controls variance in ACT-R activations. Better performance on some items will be modeled by higher ACT-

TABLE 4

Parameter Estimates of Various Experiments

	Serial recall	Burrows & Okada[a] (1975)	Raeburn[a] (1974)	Ratcliff, Clark, & Shiffrin (1990)	Murdock (1962)	Glenberg et al. (1980)	Roberts (1972)	Johnston, Dark, & Jacoby (1985)	Hayman & Tulving (1989)
s (noise)	0.3			0.55	0.7	0.6	0.85	0.65	0.3
τ (threshold)	−0.35			1.8	3.2	1.4	2.9	0.9	−.45
F (Latency scale)	0.22	4.15	2.29	2	2	2	2	1.3	0.5
D (Partial match serial position)	2.5		2.5						
P (Partial match list context)				1.5				1.5	1.5
Respond	0.5			0.5	0.5	0.5	0.5	0.5	0.5
Encoding	0.2		0.2	0.2	0.2	0.2	0.2	0.2 0.5 0.7	0.2
Rehearse	0.5		0.5	0.5	0.5	0.5	0.5	0.5	0.5
Intercept	0.5	0.5	0.6						
Time (R^2)	0.946	0.957	0.886					0.910	
Accuracy (R^2)	0.906			0.970	0.923	0.860	0.990	0.921	0.962
Base activation	0.24	2.19	1.1	0.35	0.35	−1.51	0.52	−1.47	−1.07
Associative activation	0.61	1.64	1.9	2.5	0.6	0.35	0.6	4.51	1.4
Average activation	0.85	3.83	3.0	2.85	0.95	−1.27	1.12	3.04	0.33

[a] Since accuracy data are not modeled there was no need to estimate the s or τ parameters for these experiments. There was also no study process modeled in the Burrows and Okada experiment.

R activations and worse performance by lower activations. Thus, the mixture of activations produced by s will tend to mirror the mixture of performances of subjects.

The other two parameters, τ and F, do appear to vary more dramatically from experiment to experiment. These two parameters map activation into performance. Probability of recall is a function of the distance between activation and the threshold τ. Latency is scaled by the parameter F. It turns out that the τ and F parameters are related to the mean activation level of memory chunks. The bottom line in Table 4 shows these activations at the point at which study is completed and before testing begins. In the experiments with multiple conditions these averages are calculated over the conditions. The activations are sums of base-level activations and associative activations when the goal is set to retrieve these items. These base-level and associative activations are also shown. It is apparent that these mean activations fluctuate from experiment to experiment. The exact activation levels in ACT-R are somewhat arbitrary. The activations reflect things like how many chunks are in the system and in any simulation the number is going to be much less than the true number.[19] It turns out that the τ and F parameters serve to map activation levels onto performance. This is particularly apparent in the case of the τ parameter where probability of recall is a function of the gap between τ and activation. Similarly, latency is a function of the gap between the logarithm of F and the activation level. Retrieval time will be proportional to the exponential of the difference between ln F and activation. If one adds a constant to all activation values one will get the same predictions if one adds that constant to τ and ln F.

[19] For instance, unless otherwise programmed into the simulation, the S in Associative Strength Eq. (3) defaults to the logarithm of the number of chunks in the system.

Figure 18 plots τ and ln F as a function of average activation level. We have separately plotted τ from the free-recall experiments (Murdock, Glenberg, Roberts—F does not vary across these experiments). While the relationship is not perfect there is an approximately linear relationship between τ and ln F and average activation over the experiments. Thus, a major reason for the fluctuation in τ and F is to compensate for the arbitrary differences in mean activation levels from experiment to experiment. Activation is an interval scale in ACT-R where absolute differences are meaningful but there is no meaningful zero.

It might seem strange that the curve for the free recall τ is so far above other τ curve in Fig. 18. The reason for this is that in the free recall models a word can be recalled if, on any cycle, its noise brings it above threshold before recall terminated. Thus, there are many opportunities to recall a word in free recall while, in the recognition or serial memory paradigms, there is just one opportunity to recall the word. Because random noise was independent from cycle to cycle, this meant that chances were good for recall unless the threshold was very high. As discussed in Anderson and Lebiere (1998), the solution to this is to have the noise correlated from cycle to cycle in the simulation—something we did not pursue in the free-recall models.

While these studies reveal a general relationship between activation and the parameters, τ and F, there is no reason to believe the relationship should be perfect since there should be experiment and population differences in these parameters just as there should be in the s parameter. Differences in τ correspond to differences in levels of recall and bias. Some subjects in some experiments will show a greater recall and tendency to false alarm. This is captured by lower values of the τ parameter. Similarly, it is reasonable to assume that there will be population differences in retrieval speed (e.g., Salthouse, 1991). Moreover in some experiments subjects will have practiced more and so display higher activations. So all differences in activation levels are not arbitrary and so we should

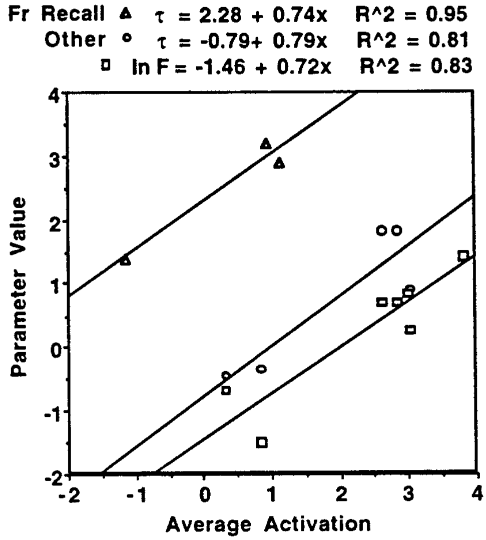

FIG. 18. The relationship between average activation in an experiment and the threshold parameter τ and the logarithm of the latency factor.

not always expect to see compensating changes in τ and F. Nonetheless, Fig. 18 suggests that variations in the average activation levels are a major reason our model fits for the large variation in the parameters, τ and F.

It is remarkable that the three functions in Fig. 18 are nearly parallel. Since τ and ln F both show approximately the same linear relationship to mean activation, they should show a simple relationship to one another. In fact, one can do a fairly good job in predicting F from τ for the nonfree recall experiments by the function

$$F = 0.348e^{\tau}. \tag{11}$$

This accounts for 93.9% of the variance in the F parameter across the nonfree-recall experiments where estimates of both τ and F where obtained.[20] So really in some cases there is only one parameter being estimated per experiment which can be conceived of as τ and the prediction is that retrieval time is about a third

[20] The free recall experiments provide very little constraints for the estimate of the F parameter and it was just arbitrarily set at 2 for all of these experiments.

of a second when the activation is at the threshold τ. This equation also describes a situation where subjects can trade off increased accuracy for increased latency by lowering the threshold.

In conclusion, the parameter variation across experiments is exceedingly regular. The s parameter shows little variation and the F and τ parameters are related by the speed–accuracy equation above. We view this regularity as a major piece of evidence for the underlying ACT-R theory.

Assumptions in Modeling

In a couple of ways the ACT-R models are exceedingly forthright in their assumptions. For instance, one can connect through the worldwide web, test these models, and examine their assumptions. In addition, it is the case that there are versions of these models which will interact with the same experiment-running software that could be used to run the experiments (and these can be obtained by writing to the authors). Thus, there are not any hidden assumptions about how the correspondence is made between the models and the empirical phenomena.

Each experiment simulation requires a number of assumptions (not part of the ACT-R theory) about how subjects approach these experiments. These assumptions are made in terms of the knowledge representations and production rules. We have tried to make explicit these key assumptions. They include things like a hierarchically grouped representation for serial lists, rehearsal assumptions, recognition confusions through partial matching of context, and use of a buffer of items in free recall. The predictions of the theory are dependent on the underlying architecture of ACT-R but also on these auxiliary modeling assumptions.

Interestingly, in almost no case are these auxiliary assumptions novel. They reflect ideas that have been in the field of list memory, often for decades. Indeed, most of these auxiliary assumptions are to be found in Anderson and Bower's (1973) chapter on verbal memory. The Anderson and Matessa serial memory model is basically the model presented there (which in turn reflected work like that of Johnson, 1970) augmented with ideas about positional encoding and Baddeley's ideas of time-based decay and rehearsal. The analysis of recognition and free recall is basically Anderson and Bower (1972) whose free recall component has strong influences from Atkinson and Shiffrin (1968). The implicit memory theory comes from the recent theories of Reder and of Bower which in turn show strong influence of the Anderson and Bower models.

This theory does contrast with the Anderson and Bower models in that it assumes no short-term store. One of the interesting aspects of this modeling enterprise is that it has produced models that seemlessly transition from what are considered short-term tasks like memory span and the Sternberg task to tasks that are considered long-term like recognition and free recall. In fact, these various tasks have long-term and short-term components. For instance, the 12 items that subjects had to remember in some of the conditions of our serial memory experiment or the 20 items required in Fig. 9 go beyond the traditional bounds of short-term memory. The recency function in free-recall is typically within the short-term span but is not in some of the experiments that have studied the *IPI*/*RI* ratio. ACT-R's ability to seemlessly transition among paradigms and to model effects of varying task parameters is produced by the retention function built into the base-level Eq. (2). This produces rapid initial decay and slower later decay. Thus, ACT-R provides evidence for the artificiality of the traditional separation between short-term memory and long-term memory.[21] While the elimination of the short-term store is a clear departure from Anderson and Bower, it is hardly a new idea in the field either.

[21] This is not to say ACT-R denies the existence of transient sensory buffers. It is also the case that its current goal provides a transient abstract memory (which implemented our buffer for free recall).

One might ask what is contributed by the ACT-R theory? One could take the view that these results are predicted by theoretical ideas that predate ACT-R by at least two decades. However, this would ignore that the Anderson and Bower theory could only qualitatively integrate these various domains and only in relatively vague terms. The simulation models offered in this paper establish that all of these qualitative accounts can be woven into a consistent theory which predicts precisely the data obtained in actual experiments. They bring out a substantial systematicity in the underlying parameters across these experiments. They show that phenomena such as the Wiseman–Tulving functions are not inconsistent with the theory as had long been believed. They show that decay assumptions can produce the ratio-rule for the serial-position effect, which had not been suspected. Thus, the ACT-R theory serves the role claimed for it. That is, it integrates existing ideas in the field.

REFERENCES

Aaronson, D. (1968). Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology,* **76,** 129–140.

Ackley, D. H., Hinton, G. E., & Sejnowsky, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science,* **9,** 147–169.

Anderson, J. R. (1972). FRAN: A simulation model of free recall in G. H. Bower, (Ed.), *The psychology of learning and motivation, vol. 5.* New York: Academic Press.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology,* **5,** 451–474.

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

Anderson, J. R. (1983). *The architecture of cognition.* Hillsdale, NJ: Erlbaum.

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

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bothell, D., Lebiere, C., and Matessa, M. (1997). *Published ACT-R models.* [http://act.psy.cmu.edu/]

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review,* **79,** 97–123.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Bower, G. H. (1974). Interference in memory for multiple contexts. *Memory and Cognition,* **2,** 509–514.

Anderson, J. R., John, B. E., Just, M. A., Carpenter, P. A., Kieras, D. E., & Meyer, D. E. (1995). Production system models of complex cognition. In Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society (pp. 9–12). Hillsdale, NJ: Erlbaum Associates.

Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., Lebiere, C., & Matessa, M. (1996). *ACT-R: A working theory of human cognition.* Paper presented at the Psychonomics Society Conference.

Anderson, J. R., & Matessa, M. (in press). The rational analysis of categorization and the ACT-R architecture.

Anderson, J. R., & Matessa, M. P. (1997). A production system theory of serial memory. *Psychological Review,* **104,** 728–748.

Anderson, J. R., Matessa, M. P., & Lebiere, C. (1997). The ACT theory and the visual interface. *Human Computer Interaction,* **12,** 439–462.

Anderson, J. R., & Reder, L. M. (in press). *The fan effect: New results and new theories. Journal of Experimental Psychology: General.*

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology,* **30,** 221–256.

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

Anisfeld, M., & Knapp, M. (1968). Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology,* **77,** 171–179.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2). New York: Academic Press.

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

Baddeley, A. D., Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior,* **14,** 575–589.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior,* **13,** 80–97.

Bower, G. H. (1996). Reactivating a reactivation theory of implicit memory. *Consciousness and Cognition,* **5,** 27–72.

Briggs, G. E. (1974). On the predictor variable for choice reaction time. *Memory & Cognition,* **2,** 575–580.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language,* **31,** 429–460.

Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science,* **188,** 1031–1033.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology,* **55,** 75–84.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior,* **4,** 161–169.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language,* **31,** 668–684.

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

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology* (translated by H. A. Ruger & C. E. Bussenues, 1913). New York: Teachers College, Columbia University.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.

Estes, W. K. (1973). Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning and Verbal Behavior,* **12,** 360–372.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General,* **112,** 309–346.

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review,* **85,** 153–171.

Gardiner, J. M., Thompson, C. P., & Maskarinec, A. S. (1974). Negative recency in initial free recall. *Journal of Experimental Psychology,* **103,** 71–78.

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

Glenberg, A. M., Bradley, M. M., Kraus, T. A., & Renzaglia, G. J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **9,** 231–255.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, 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.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **10,** 164–178.

Hayman, C. G., & Tulving, E. (1989). Is priming in fragment completion based on a 'traceless' memory system? *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **15,** 941–956.

Healy, A. F., & McNamara, D. S. (1996). Verbal learning and memory: Does the modal model still word? In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.) *Annual review of Psychology,* **47,** 143–172.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology,* **49A,** 80–115.

Hinrichs, J. V. (1968). Prestimulus and poststimulus cueing of recall order in the memory span. *Psychonomic Science,* **12,** 261–262.

Hinton, G. E., & Sejnowsky, T. J. (1986). Learning and relearning in Boltzmann machines. In Rumelhart, D. E., McClelland, J. L., and the PDP group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General,* **122,** 139–154.

Johnson, N. F. (1970). The role of chunking and organization in the process of recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 4, pp. 171–247). New York: Academic Press.

Johnson, G. J. (1991). A distinctiveness model of serial learning. *Psychological Review,* **98,** 204–217.

Johnston, W. A., Dark, V., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **11,** 3–11.

Johnston, W. A., Hawley, K. J., J. & Elliott, J. M. G. (1991). Contributions of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **17,** 210–233.

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior,* **16,** 395–418.

Lee, C. L., & Estes, W. K. (1981). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior,* **16,** 395–418.

Lewandowsky, S., & Murdock, B. B., Jr. (1989). Memory for serial order. *Psychological Review,* **96,** 25–57.

Li, S. C., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 837–847.

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive Psychology,* **31,** 168–217.

Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 1, pp. 328–372). New York: Academic Press.

McClelland, J. L., & Chappell, M. (1994). *Bayesian recognition.* Poster given at the 35th Annual Meeting of the Psychonomic Society.

Metcalfe, J., & Sharpe, D. (1985). Ordering and reordering in the auditory and visual modalities. *Memory & Cognition,* **13,** 435–441.

Murname, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **17,** 855–874.

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

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review,* **100,** 183–203.

Nairne, J. S. (1992). The loss of positional certainty in long-term memory. *Psychological Science,* **3,** 199–202.

Nairne, J. S., & Neath, I. (1994). Critique of the retrieval/deblurring assumptions of the theory of distributed associative memory. *Psychological Review,* **101,** 528–533.

Nairne, J. S., Neath, I., Serra, M., & Byun, E. (1997). Positional distinctiveness and the ratio rule in free recall. *Journal of Memory and Language,* **37,** 141–154.

Neath, I. (1993). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language,* **32,** 820–840.

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

Nilsson, L-G., & Gardiner, J. M. (1993). Identifying exceptions in a database of recognition failure studies from 1973 to 1992. *Memory & Cognition,* **21,** 397–410.

Poldrack, R. A., & Logan, G. D. (1997). Fluency and response speed in recognition judgments. *Memory & Cognition,* **25,** 1–10.

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

Raaijmakers, J. G. W., & Shiffrin, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology,* **43,** 205–234.

Raeburn, V. P. (1974). Priorities in item recognition. *Memory & Cognition,* **2,** 663–669.

Ratcliff, R., Clark, S., & Shiffrin, R. M. (1990). The list strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **16,** 163–178.

Reder, L. M., & Gordon, J. S. (1996). Subliminal perception: Nothing special cognitively speaking. In J. Cohen, & J. Schooler (Eds.), *Cognitive and neuropsychological approaches to the study of consciousness* (pp. 125–134). Mahwah, NJ: Erlbaum.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Angstadt, P., & Hiraki, K. (1996). *Modeling word frequency effects in a continuous remember/know judgment paradigm.* Paper presented at the 37th Annual Meeting of the Psychonomic Society. Chicago, IL, November, 1996.

Reder, L. M., Nhouyvansivong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (1997). Modeling the Mirror effect in a Continuous Remember/Know Paradigm. In *Proceedings of the Nineteenth Annual Cognitive Science Conference* (pp. 644–649). Stanford, CA, August 1997.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.) *Implicit memory and metacognition* (pp. 45–77). Mahwah, NJ: Erlbaum.

Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory using EPAM IV. *Psychological Review,* **102,** 305–330.

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

Roberts, S., & Pashler, H. *Data fitting and theory testing.* Manuscript submitted for publication.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review,* **103,** 734–760.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Erlbaum.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not calculate: A source activation confusion (SAC) model of problem-familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **23,** 1–27.

Shiffrin, R. M., & Cook, J. R. (1978). A model for short-term item and order retention. *Journal of Verbal Learning and Verbal Behavior,* **17,** 189–218.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review,* **4,** 145–166.

Slamecka, N. (1967). Serial learning and order information. *Journal of Experimental Psychology,* **74,** 62–66.

Sloman, S. A., Hayman, C. A. G., Ohta, N., Law, J., & Tulving, E. (1988). Forgetting in primed fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 223–239.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning.* New York: Academic Press.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review,* **80,** 352–373.

Tulving, E., & Weisman, S. (1975). Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society,* **6,** 79–82.

Underwood, B. J., & Freund, J. S. (1968). Effect of temporal separation of two tasks on proactive inhibition. *Journal of Experimental Psychology,* **78,** 50–54.

Watkins, M. J., & Gibson, J. M. (1988). On the relationship between perceptual priming and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **14,** 477–483.

Wickelgren, W. A. (1965a). Short-term memory for phonemically similar lists. *American Journal of Psychology,* **78,** 567–574.

Wickelgren, W. A. (1965b). Short-term memory for repeated and non-repeated items. *Quarterly Journal of Experimental Psychology,* **17,** 14–25.

Young, R. K. (1968). Serial learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and behavior theory* (pp. 122–148). Englewood Cliffs, NJ: Prentice-Hall.

(Received June 24, 1997)
(Revision received November 19, 1997)