Psychological Review
1997, Vol. 104, No. 4, 728–748

0033-295X/97/$3.00

A Production System Theory of Serial Memory

John R. Anderson and Michael Matessa
Carnegie Mellon University

A theory is described that provides a detailed model of how people recall serial lists of items. This theory is based on the Adaptive Character of Thought-Rational (ACT-R) production system (J. R. Anderson, 1993). It assumes that serial lists are represented as hierarchical structures consisting of groups and items within groups. Declarative knowledge units encode the position of items and of groups within larger groups. Production rules use this positional information to organize the serial recall of a list of items. In ACT-R, memory access depends on a limited-capacity activation process, and errors can occur in the contents of recall because of a partial matching process. These limitations conspire in a number of ways to produce the limitations in immediate memory span: As the span increases, activation must be divided among more elements, activation decays more with longer recall times, and there are more opportunities for positional and acoustic confusions. The theory is shown to be capable of predicting both latency and error patterns in serial recall. It addresses effects of serial position, list length, delay, word length, positional confusion, acoustic confusion, and articulatory suppression.

In this article we describe our efforts to come to a detailed process understanding of the task involved in reproducing a serial list of items. This is certainly an area that has received a great deal of research, and a great many phenomena have been documented (e.g., Baddeley, 1986; Burgess & Hitch, 1992; Conrad, 1964; Ebbinghaus, 1913/1885; Estes, 1973; Lashley, 1951; Lewandowski & Murdock, 1989; Murdock, 1993; Shiffrin & Cook, 1978; Wickelgren, 1965a, 1965b; Young, 1968). We offer a theory that explains many of these phenomena by specifying the moment-by-moment processes involved in recalling a list. We provide a model within the Adaptive Character of Thought-Rational (ACT-R) theory (Anderson, 1993) of this serial recall process. ACT-R is a theory that naturally addresses the detailed latency patterns and error patterns in any task. The ACT-R theory (Anderson, 1993) comes with a set of independently motivated processing assumptions, none of which were modified or adapted to account for serial memory. This system is capable of simulating the results to the level of detail that it can interact with the same experimental software that we present to participants and can reproduce the same patterns of data (Anderson, Matessa, & Douglass, 1995). However, the ACT-R theory by itself would not be able to account for the results from serial memory. It had to be augmented with a theory of the nature of the representation of a serial list, and, as such, our effort provides an answer to the question of serial representation, which has seemed so simple and yet so elusive (Young, 1968).

Although we have applied our model to a number of tasks, here we concern ourselves mainly with the memory span paradigm: an immediate memory task in which participants are asked to repeat back a list of items. Memory span has been the focus of intelligence tests for decades and has become a recent focus of research. Although the memory span limitation is popularly conceived of as a simple parameter of human memory, the research literature shows that this limitation is complex (a point emphasized relatively early by Watkins, 1977). The reasons for the limitation in the ACT-R theory are correspondingly complex. However, the basic process by which a participant performs a memory span task is relatively simple (i.e., he or she tries to step through a serial list and say every item in order). The ACT-R theory implies that there are many ways that an individual can trip up in making this journey through the list.

Memory Span

Baddeley's (1986) account of the nature of the memory span for verbal items has become extremely influential and finds its way as the correct explanation in most modern textbooks on memory and cognitive psychology, including our own (Anderson, 1995a, 1995b). As Baddeley (1990) described it in his textbook,

> the simplest account might be to suggest that the process of overt or covert articulation involves setting up and running speech motor programs which operate in real time, with the result that the longer the word the longer it takes to run off. If we assume that this process of subvocal rehearsal has the function of maintaining items in the phonological store by refreshing their fading traces, then the faster it can run, the more items will be maintained and the longer the

John R. Anderson and Michael Matessa, Department of Psychology, Carnegie Mellon University.

This research was supported by Office of Naval Research Grant ONR N00014-90-J-1489 and National Science Foundation Grant SBR 94-21332. We would like to thank Marsha Lovett, Chris Schunn, Alan Baddeley, and Stephan Lewandowski for their helpful comments on this article. General information on the Adaptive Character of Thought-Rational (ACT-R) theory can be found at the following Web site: http://act.psy.cmu.edu.

Correspondence concerning this article should be addressed to John R. Anderson, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent via Internet to ja+@cmu.edu.

memory span. If we assume that the memory fades, then the memory span will be determined by the number of items that can be refreshed before they fade away. That number, of course, will depend both on how rapidly the trace fades and on how long it takes to articulate each item and hence refresh each memory trace. Data from studies using English, Welsh, Hebrew, Spanish, Arabic, and Chinese all give results suggesting that trace decay time is approximately two seconds, although as mentioned earlier, rehearsal time, and consequently span vary widely from one language to another. (p. 79)

Thus, according to Baddeley (1990), the amount that can be maintained in a memory span is the number of words that can be rehearsed in approximately 2 s. The strong evidence for this comes from research showing that people can maintain fewer words that take longer to articulate, either because the words have more syllables or have syllables that are longer to articulate. In one influential study, Baddeley, Thomson, and Buchanan (1975) examined the number of words (out of five) that could be repeated back as a function of syllable length. Over the range of syllables per word from one to five, they found that this was always equal to the number of words that could be said in approximately 2 s.

As Baddeley (1986) emphasized, this rehearsal-time hypothesis is only part of the explanation of memory span. There are a number of facts that point to a need to complicate the account of memory span:

1. As a simple hypothesis, it would imply that memory span would be related to articulation rate by an equation of the following form: span = 2 * items per second. However, attempts to regress span on rate often show significant nonzero intercepts (sometimes as much as three items; Hulme, Maughan, & Brown, 1991; Morra, Tressoldi, Mazzoni, Sava, & Zucco, 1991; e.g., see our Figure 9).

2. A literal interpretation of this hypothesis would imply that there was a "drop-dead" length above which participants cannot recall a list perfectly. However, the probability of perfect recall decreases gradually with list length (Crannell & Parrish, 1957). Thus, the basic proposal needs to be embellished with some probabilistic function. For instance, Schweikert and Boroff (1986) suggested that the duration time for an item is a random variable with a mean of about 2 s.

3. There is relatively little effect of either presentation rate of material or inserting a delay between the presentation and recall (Baddeley, Lewis, & Vallar, 1984; Dosher, 1994). For instance, Baddeley et al. manipulated presentation rate from 0.5 to 3 s per digit. Under conditions of articulatory suppression, span dropped only from 7.13 to 6.04 in one study and from 5.79 to 5.75 in another. A 3-s presentation rate should have wiped out memory for the list. Long presentation rates are particularly problematic in studies such as those of Baddeley et al., in which articulation was suppressed. On the other hand, in the absence of articulatory suppression, it might seem possible that a covert rehearsal process could eliminate the effects of slow presentation or the effect of delay between presentation and test. However, even in this case there are difficulties in working out the details of such a proposal so that no item is left unrehearsed for more than 2 s. For example, what if the participant is halfway through rehearsing a 2-s list and receives a command to recall? If he or she starts recalling immediately, the end of the list will not receive rehearsal and it will be a 3-s delay before getting back to the end of the list. Proposals for interleaving rehearsal during a slow presentation would run into similar timing problems.

4. It often takes participants longer than 2 s to recall their lists (Cowan, 1992). Moreover, both the time it takes to initiate list recall and the time to recall each item increase with the length of the list (Sternberg, Monsell, Knoll, & Wright, 1978). Thus, there is not a fixed generation time for an item independent of list length.

5. The rehearsal-time hypothesis needs to be elaborated to incorporate the contribution of confusability to limitations of span. Many of the participants' errors involve positional (Aaronson, 1968; Bjork & Healy, 1974) and phonological (Conrad, 1964) confusions. Baddeley's rehearsal-time hypothesis can be expanded to account for such effects, and Baddeley has made much use of data on acoustic confusions in distinguishing between phonological and articulatory stores (e.g., Baddeley et al., 1984). Any complete explanation must incorporate an interference-based limitation to memory span as well as a time-based component.

6. Participants show a tendency to group a list of items into subsequences (Bower & Winzenz, 1969; Johnson, 1970). This subsequence structure affects both the rate at which they recall the items and the errors that they make.

The challenge, then, is to be able to incorporate Baddeley's insight about the time-dependent factor in memory with all of these additional complications. As Baddeley suggests (personal communication, 1996), this requires going beyond a stated "verbal description" to "a fully specified model." This is the challenge that we try to take up in this article. Moreover, we want to do this within a general theory of cognition, ACT-R, whose basic assumptions have been forged to account for very different phenomena. However, the ACT-R theory is well situated to provide such a model. The complications we listed earlier were all concerned with the issue of how one gets all of the pieces of the puzzle to fit together into a coherent system that is consistent with the data. Because ACT-R is a simulation model that embodies many of the ideas and actually performs the task, it can address questions such as how the timing really works, possible roles of covert rehearsal, and interaction with interference.

We next describe the ACT-R theory and an example of a typical application of that theory. We then turn to the issue of what representational assumptions were made so that ACT-R could perform the serial memory task.

The ACT-R Theory

ACT-R (Anderson, 1993) is a theory of human cognition that assumes that a production system operates on a declarative memory. It is a successor to previous ACT production-system theories (Anderson, 1976, 1983a, 1983b) 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 that determine their rates and probabilities of being processed by the 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.

According to the ACT theories, knowledge is divided into declarative knowledge and procedural knowledge. In ACT-R, declarative knowledge is represented in terms of *knowledge units*,[1] which are schemalike structures. Each knowledge unit is of a particular type and has an associated set of slots encoding its contents. Figure 1 is a graphical display of a knowledge unit of the type addition fact, which encodes that $3 + 4 = 7$. B_i, W_j, and S_{ji} are quantities relevant to activation computation, and they are discussed in the subsequent section.

According to ACT-R, procedural knowledge, such as mathematical problem-solving skill, is represented by production rules that coordinate the retrieval of declarative information like that in Figure 1 for purposes of problem solving. For instance, suppose a child was at the following point in the solution of a multicolumn 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 condition elements. 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 long-term memories to match the rest of the condition. The times to perform these retrievals will be

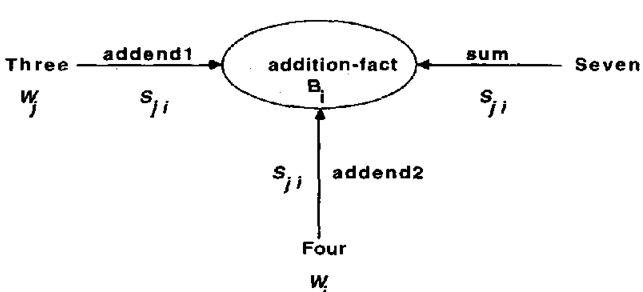

Figure 1. A network representation of a declarative ACT-R knowledge unit. W_j = source activation of element j in the focus of attention; S_{ji} = strength of association from element j to element i; B_i = base-level activation of element i.

important contributions to the latency for the production rule, and level of activation of knowledge units will determine their retrieval time. Therefore, in this case the time to apply this production will be determined by the level of activation of the knowledge unit encoding $3 + 4 = 7$ in Figure 1. In the next section we explain how activation determines retrieval time. In addition to this match time, 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 such as mathematical problem solving. However, the ACT theory originated as a theory focused on human memory (Anderson, 1976; Anderson & Bower, 1973). We propose that productions similar to those guiding problem solving in a mathematics domain are guiding recall in a serial memory task.

Activation

Activation of declarative structures has always been an important concept in the ACT theories. Basically, activation determines how available information will be.[2] The activation of a knowledge unit 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 element i is

$$A_i = B_i + \sum_j W_j S_{ji}, \tag{1}$$

where B_i is the base-level activation of element 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 element 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_js would be the source activations of the elements 3 and 4 in the column and the S_{ji}s would be the strengths of association from these elements to the knowledge unit encoding $3 + 4 = 7$. Figure 1 illustrates these quantities in the network encoding of the knowledge unit. It is assumed in ACT-R, in contrast to early versions of ACT (such as in Anderson, 1976) but as in ACT* (Anderson, 1983a, 1983b), 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 search of associative memory (SAM) theory (Gillund & Shiffrin, 1984; Raajimakers & Shiffrin, 1981), except that activations in ACT-R are like logarithms of SAM familiarities because they add rather than multiply. It is

[1] These knowledge units are called "chunks" by Anderson (1993), but we have repressed this terminology to avoid confusion with the term *chunk* as it is used in a different sense in the serial memory literature (e.g., Miller, 1956).

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

important to keep conceptually separate the quantities A_i and W_j. The former are activations, which control retrieval from declarative memory, and the latter reflect the salience or attention given to the cues. The W_js are referred to as *source activations.*

One aspect of the memory span limitation is time-dependent decay. According to the ACT-R theory, the time-dependent decay occurs in the base-level activations. The base-level activation of a knowledge unit 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:

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

As developed in Anderson and Schooler (1991), this equation produces both the power law of forgetting, where the strengths of individual experiences decay as power functions, and the power law of learning, where individual experiences accumulate strength as a power function of the number of exposures. The decay effect is produced by the d exponent, and the practice effect is produced by the summation across experiences.

With respect to activation, declarative memories can be retrieved only if they are above a threshold of activation τ. Because of noise in the activation levels, there is only a certain probability that a memory will be above the threshold. ACT-R assumes the activation levels are distributed according to a logistic distribution (which is like a normal distribution). This means that the probability of a knowledge unit with expected value A_i being above the threshold is

$$\text{prob}_i = \frac{1}{1 + e^{-(A_i-\tau)/s}}, \tag{3}$$

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

The other dependent measure ACT-R addresses is latency. The time to retrieve a knowledge unit is related to its activation by the following formula:

$$\text{time}_i = Fe^{-fA_i}, \tag{4}$$

where F is a time scale factor and f serves to scale the activation values. Equation 4 describes only 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, such as Anderson, John, et al., 1995) or more if a physical action is required (e.g., speaking, typing).

The exponential functions in Equations 3 and 4 allow for the kind of nonlinear mapping of activation onto behavior required in many activation theories (e.g., McClelland & Rumelhart, 1981). For a justification of these exponential assumptions in ACT-R, see Anderson (1993).

New ACT-R Assumptions: Capacity Limitations and Partial Matching

The assumptions laid out so far are part of the ACT-R theory described by Anderson (1993). However, more recently, as a consequence of our modeling of equation solving (Anderson, Reder, & Lebiere, 1996), we were motivated to elaborate the theory with two additional assumptions. As the ACT-R theory was originally described, there were no necessary limitations on the sources of activation. However, to account for effects of manipulating complexity of the algebra tasks, we added the assumption that there is a limitation on total source activation. Formally, this limitation is

$$\sum_j W_j = 1. \tag{5}$$

This reflects a limitation on the amount of attention one can distribute over source elements.

This resource limitation has some similarity to the ideas introduced by Kahneman (1973) and has a bit of similarity to Just and Carpenter's (1992) controlled activation production system (CAPS) theory, which interprets working-memory limitation as a limitation on the total amount of activation available in a production-system architecture. However, there are differences with the CAPS theory. Activation in the CAPS theory spreads by production firings rather than by associations directly from sources to memory structures. Also, the ACT-R limitation is not directly a limitation on activation but on the sources of activation. The total activation (A_i in Equation 1) is a function of the base levels B_i and strengths S_{ji} as well as the W_j, and consequently there is no fixed cap on the A_is. Finally, and most important, ACT-R's capacity limitation affects retrieval from declarative memory, whereas in CAPS capacity affects the number of times a production must repeat its firing.

The second new assumption was motivated by the many errors in algebra that seemed to be attributable to misretrieving arithmetic facts and algebraic transformations that 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 knowledge units 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 knowledge unit i to a condition in a production rule is

$$M_i = A_i - P, \tag{6}$$

where P is a mismatch penalty that depends on the similarity between the knowledge unit and condition. Thus, faced with the goal to retrieve the sum of 3 + 4, the knowledge units 3 + 4 = 7 and 3 + 1 = 4 would have equal activation scores (both have source elements 3 and 4), but 3 + 1 = 4 would receive a mismatch penalty (because the addends 1 and 4 do not match). The knowledge unit retrieved to match a production condition is the one with the largest match score. Normally, when a perfectly matching knowledge unit competes with a partially matching knowledge unit, the perfectly matching knowledge unit will be retrieved because it has the largest match score. However, occasionally, a partially matching knowledge unit will be selected over a perfectly matching knowledge unit because the activation noise gives it sufficiently more activation to overcome the mismatch penalty it suffers. When a partially matching knowledge unit so beats out a perfectly matching knowledge unit, there will be errors of commission in retrievals. Only when all knowledge units fail to reach the activation threshold does retrieval fail completely (errors of omission). Partially matching errors of commission are the cause of intrusion and transposition errors in serial recall, whereas retrieval failures are the cause of recall blanks.

ACT-R Theory of Serial Recall

Although these are the general assumptions of the ACT-R theory, a theory also is required to describe how knowledge is represented and retrieved in performing a memory span task. Our assumption is that a list is organized as a sequence of groups and each group is represented as a sequence of items. Consider the ACT-R representation in Figure 2 for a list of nine digits grouped as 329 714 856. Each oval in Figure 2 represents a separate knowledge unit. Each knowledge unit encodes the identity of the element and its position in the higher order structure. A stimulus can be perceived at many levels, and each level will have its own encoding as a knowledge unit. Thus, there are knowledge units to encode each group as well as each digit. Note that this representation assumes that each element is indexed by its position. Thus, it is consistent with those theories of serial structure that assume that the effective stimulus for serial recall is positional rather than associations to prior elements (Young, 1968).

This declarative representation for the serial recall task is not sufficient. One also needs a set of production rules that operate on this knowledge representation to generate the recall. The key production rule in our modeling of the serial memory task is the following rule:

Get-Next
IF the goal is to retrieve an element at position p in group g,
  and x is the element at p in group g,
THEN set a subgoal to generate x,
  and change the goal to retrieve the element in the next position.

Thus, if the goal were to retrieve the element in position 2 in group 2, this production would retrieve the knowledge unit in Figure 2 containing the digit "one" and set a subgoal to generate this element. This production rule also would change the goal to retrieve the element in position 3.

This production rule will be followed by a production rule whose responsibility is to generate the item:

Generate-Item
IF the goal is to generate an item,
  and the item is associated with a motor program,
THEN execute the motor program and the goal is satisfied.

The production *generate-item* involves retrieval of an articulatory code (or finger code when typing the answer, as in the pilot experiment to be reported).

Special actions also are necessary when the end of a group has been reached:

Next-Group
IF the goal is to retrieve an element at position p in group g,
  and there is no element at position p in group g,
  and g^* is the next group,
THEN change the goal to retrieve the first element of g^*.

Done
IF the goal is to retrieve an element at position p in group g,
  and there is no element at position p in group g,
  and there is no next group,
THEN the goal is satisfied and stop recalling.

Table 1 provides a trace of the ACT-R production system recalling the following digit string: 329 714 856. Printed after each production is the latency (in seconds) for that production to fire and the cumulative latency so far. We discuss these latencies shortly.

The only productions in this trace that have not been described are start-recall, prepare, and prepare-last, which initial-

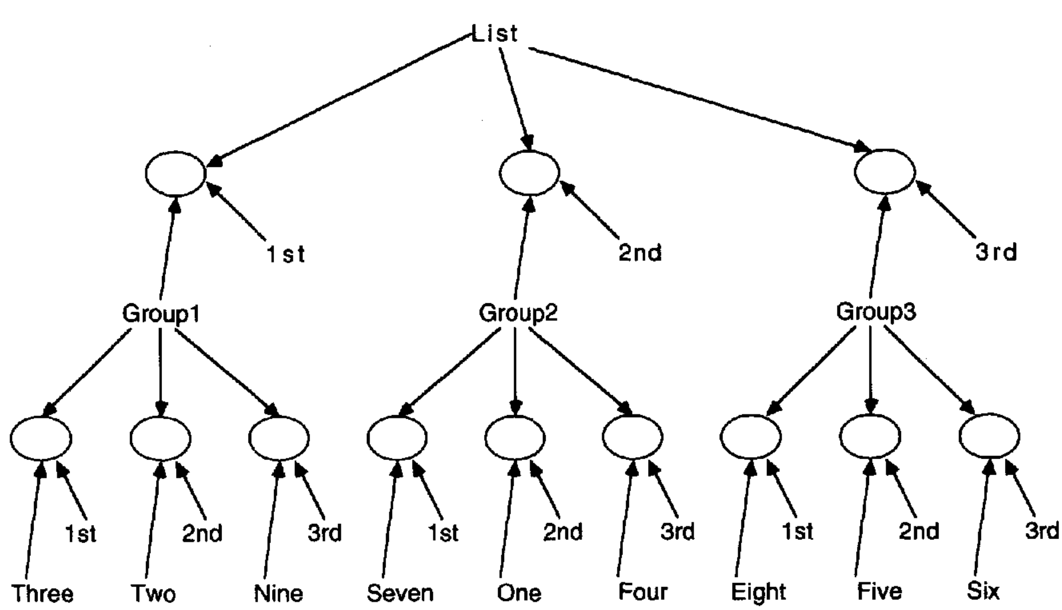

Figure 2. The ACT-R representations of a serial list. Each oval is a knowledge unit.

Table 1
A Trace of ACT-R Retrieving a Digit String 329 714 856

	Production fired	Output	Latency	Cumulative time
1	Start recall		0.690	0.690
2	Prepare		0.288	0.979
3	Prepare		0.214	1.193
4	Prepare last		0.119	1.312
5	Get next		0.352	1.664
6	Generate item	Typing 3	0.207	1.870
7	Get next		0.341	2.212
8	Generate item	Typing 2	0.206	2.418
9	Get next		0.331	2.749
10	Generate item	Typing 9	0.206	2.955
11	Next group		0.269	3.225
12	Get next		0.326	3.551
13	Generate item	Typing 7	0.206	3.757
14	Get next		0.315	4.072
15	Generate item	Typing 1	0.206	4.278
16	Get next		0.303	4.581
17	Generate item	Typing 4	0.206	4.788
18	Next group		0.237	5.025
19	Get next		0.297	5.322
20	Generate item	Typing 8	0.206	5.528
21	Get next		0.284	5.812
22	Generate item	Typing 5	0.206	6.019
23	Get next		0.271	6.289
24	Generate item	Typing 6	0.206	6.496
25	Done		0.097	6.593

Note. Cumulative latencies are in seconds after each production.

ize the recall process. These productions are responsible for setting up a motor output plan (Sternberg et al., 1978) in which the participant identifies the group structure for organizing recall:

Start-Recall
IF the goal is to recall,
and the prompt for recall has been given,
THEN set the goal to prepare the retrieval plan starting with the first group.

Prepare
IF the goal is to prepare the retrieval plan for a group,
THEN prepare to produce that group in the retrieval plan,
and change the goal to prepare the retrieval plan for the next group.

Prepare-Last
IF the goal is to prepare the retrieval plan for the last group,
THEN prepare to produce that group in the retrieval plan,
and set the goal to start retrieval with the first element of the first group.

There is nothing in the ACT-R theory that requires this motor planning because ACT-R does not have a theory of motor output. However, Sternberg et al. (1978) identified a pause before serial production that increases with the length of the sequence. They speculated that this may be caused by preparation of a motor program, and we have extended this speculation by assuming that the length effect is attributable to the number of groups that have to be prepared in the program. This model implies an effect on preparation latency of the number of groups that need to be prepared. Sternberg et al. observed an effect of the number of items but, because they did not control group structure, the number of items would have covaried with the number of subjective groups.

Although these motor planning productions are motivated to account for the initiation latency, the remaining four productions (i.e., get-next, generate-item, next-group, and done) simply specify the traversal of declarative data structure like that in Figure 2. Thus, they do not reflect any "added assumptions" beyond those in the declarative representation. They precisely identify the necessary logic for traversing that structure and identify the three key components to latency: retrieving the groups (get-group), retrieving items (get-next), and generating the response (generate-item). This serves to illustrate the fact that an ACT-R production set really does nothing more than provide a precise and detailed specification of the sequence of declarative retrievals and response generations.

Note that our model addresses in detail only the process of recall. It does not model potential rehearsal processes that are occurring during the input of the list. This is because there are a wide set of potential rehearsal strategies and it would be an enormous complication to model them all. Rather, we assume each item just gets a study on presentation. This might be viewed as assuming the simplest rehearsal strategy, which is not to rehearse. We discuss potential effects of rehearsal processes at the end of this article.

Partial Matching in Serial Recall

Partial matching has important consequences for the execution of the productions *get-next* and *generate-item* given earlier. Partial matching of get-next will produce positional confusions. Rather than retrieving the item at the target position, it can retrieve the item at a close-by position. The probability of a positional confusion will be a function of the similarity between the two positions (the target position and the incorrectly retrieved position). For instance, rather than retrieving the item in the second position of Group 1, it may retrieve the item in the third position. If there were a fixed ability to discriminate among positions within a group, there would be increased confusions as more elements are crowded within a group. The reason for a hierarchical representation may be to minimize the number of positions to be discriminated among. Wickelgren (1964, 1967) has shown that intragroup positional confusions increase when the groups are larger.

Partial matching of the production *generate-item* can lead to acoustic errors. The term *item* in generate-item is the index for retrieving something to match the condition element "and the item is associated with a motor program," just as the term *p* was the index for retrieval of something to match "and *x* is the element at *p* in group *g*" in get-next. If *item* is acoustic, a partial match can occur with a similar sounding item. Because acoustic and positional confusions occur in different productions, they will be independent. As we discuss in more detail later, Bjork and Healy (1974) demonstrated that acoustic and positional confusions are probabilistically independent.

Modeling Memory Span Tasks in ACT-R

Pilot Experiment

To illustrate this model we describe its behavior, fitting some data from one of our pilot laboratory experiments. We then apply the model to some data from the literature. Our laboratory experiment involved having participants recall digit strings of 5, 7, and 9 digits. The digits were presented 1 per second across a computer screen. When one digit appeared, the previous was removed. We tried to induce grouping by introducing visual spaces between the groups. For 5 digits we used a 3–2 grouping, for 7 digits we used 3–4 grouping, and for 9 digits we used a 3–3–3 grouping. Participants typed their answers and we recorded the time of each keystroke plus their success at recalling the item. The purpose of introducing this pilot experiment is to illustrate the theory and to establish some timing parameters. Our main agenda in this article is to fit a set of results from the literature using this theory and the timing parameters obtained from this pilot experiment.

Figure 3 shows two displays of the data that are relevant to understanding the ACT-R predictions but that are not often found in the literature. Figure 3A shows the time to recall each word in the list for those cases in which the list is perfectly recalled. The grouping structure shows through clearly with long pauses at the beginning of each group. We emphasize that the strings were presented only visually segregated without any pausing, and participants were not asked to introduce any pausing into their recall. Figure 3A also shows that latencies are somewhat longer at the same serial position of longer lists. These effects are particularly large at group boundaries. Figure 3B shows the cumulative probability that participants had made an error in their recall by the time they reached a particular serial position. Participants were encouraged to recall as much of the list as they could even if they could not recall it all. Figure 3B shows that the error rate is much higher at the same serial positions of longer lists. The probability of failing to recall each list length is just the terminal probabilities of these curves: 2.3% in the case of 5 digits, 14.4% in the case of 7 digits, and 43.3% in the case of 9 digits. As we will review these error rates are somewhat lower than observed with some other participant populations.

ACT-R can do a reasonable job of accounting simultaneously for this relatively complex pattern of latency and accuracy data. We could have presented simulation results from ACT-R directly, but in an effort to make the theory more understandable, we have chosen to present mathematical models that characterize the simulations. Therefore, we specialize the general ACT-R model in Equations 1–5 to a span experiment. First, we describe the factors that determine the level of activation in ACT-R, describe how this level of activation determines latency for productions to fire, and describe how these factors determine the probability of correctly recalling the list.

Activation. One factor controlling the model's level of recall is the activation of the digits being recalled, which was described earlier in Equation 1. This equation makes activation a function of a base-level activation that is decaying away as a function of time (Equation 2) and an associative activation that must be divided among the elements of the list. Applied to the current context this yields an equation for activation of a digit:

$$A_i = -d*\ln(t) + S/n, \quad (7)$$

where $-d*\ln(t)$ is the base-level activation, d is the power exponent from Equation 2, t is the time since the digit was encoded (measured as the time since presentation ceased), and

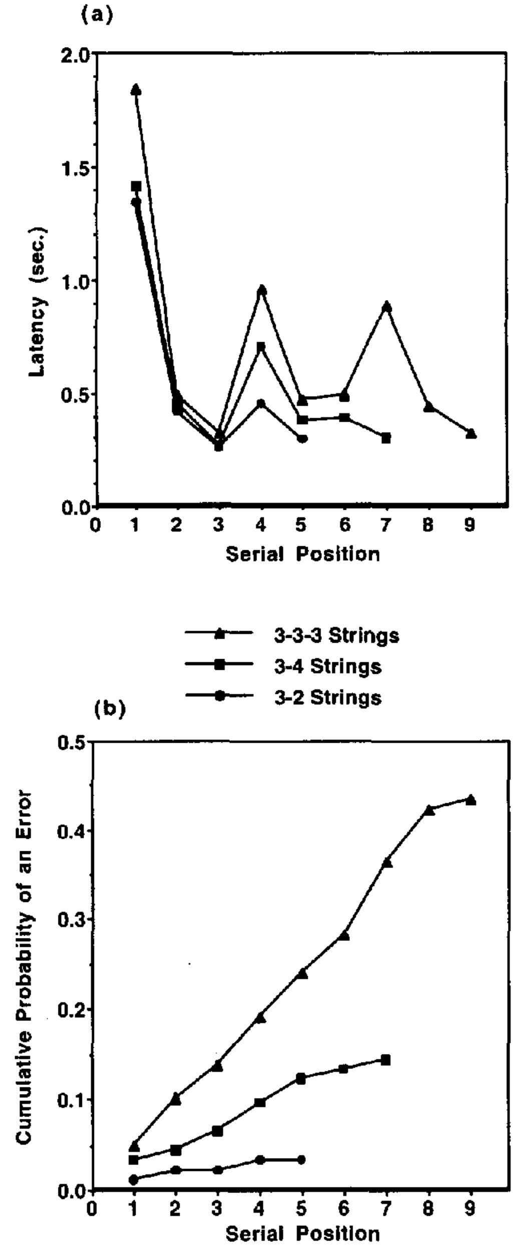

Figure 3. A: Mean time to recall each item of the digit strings as function of serial position. B: Cumulative probability of an error in recall by that serial position.

S/n is the associative strength S_{ji} from the list, j, to the element i[3]. The associative activation component, S/n, amounts to assuming a fan effect (Anderson, 1974). To keep this model consistent with Anderson et al. (1996), we set S to be the strength of self-activation, which, in ACT-R (Anderson, 1993), is the logarithm of the number of knowledge units. In this simulation this turned out to be 3.45. Thus, Equation 7 becomes

[3] The assumption is that there is just one source element and hence W_j (from Equations 1 and 5) is equal to one.

$$A_i = 3.45/n - 0.5 \ln t.$$

The 3.45 becomes a weight for the contribution of list size (n) and 0.5 becomes a weight for the contribution of time (t). Both were set arbitrarily, and we could have explored other possible weightings. However, we got satisfactory performance with this weighting and did not explore other possibilities. Because there are strong parameter trade-offs in ACT-R, we doubt fits would be much improved by using these degrees of freedom.[4]

A critical feature of Equation 7 is that it makes the activation of the list elements a function of both list length and time. Another important feature of the theory is that it does not propose any drop-dead length or time. As either increases there is decreased activation, which will gradually lead to increased errors and increased latency, as we describe later.

Latency. Figure 4A displays the times the simulation took to recall each digit of the various lists. These times are the sum of the times for the productions that fire between recall of digits (see Table 1). The time for each production in a sequence like that shown in Table 1 is

$$\text{time} = P + Fe^{-fA_i}, \tag{8}$$

where P is the action time for the production and Fe^{-fA_i} retrieval time based on Equation 4. The time-scale parameter F was estimated to be 0.106 s, and the activation scale parameter f was estimated to be 1.430. Because this is the only experiment for which we have good timing data, we will continue these estimates of F and f throughout all of our simulations of results from the experimental literature. The parameter P was set at the ACT-R minimum of 0.05 s for productions that involved no physical action and 0.20 s for productions that involved typing. The 0.05 value is a standard minimum, and 0.20 is an approximate representation of typing time (Salthouse, 1986). Finally, we estimated an additional time for the production start recall, which represented the time to recognize the recall probe and switch from study mode. This was estimated to be 0.690 s and is also used in the other simulations.

The fits obtained in Figure 4 were obtained by estimating parameters that minimized the deviation between the observed and predicted values. We chose to minimize the quantity

$$\sum[(\hat{T}_i - \bar{T}_i)^2 + 30(\hat{P}_i - \bar{P}_i)^2],$$

where $\hat{T}$ is the predicted time, $\bar{T}_i$ is the observed time, $\hat{P}_i$ is the predicted probability of recall, and $\bar{P}_i$ is the observed probability. This formula is somewhat arbitrary but results in approximately equal contributions of times and errors to the overall deviation measure. We estimated f, F, the start time, and two accuracy parameters (to be described) to minimize these deviations. All other parameters were set on a priori basis. We wrote an Excel program that estimated these parameters to minimize this quantity. This and other Excel programs used for model fitting in this article can be found by following the "published models" link from the ACT-R home page: http://act.psy.cmu.edu. The actual ACT-R simulation also can be found at this site.

As an example of how Equations 7 and 8 are used to estimate time, consider the time for get-next to fire and retrieve the first item (three) of a nine-item list in Table 1. The item was last seen 8 s before recall of the list began and 1.31 s have passed

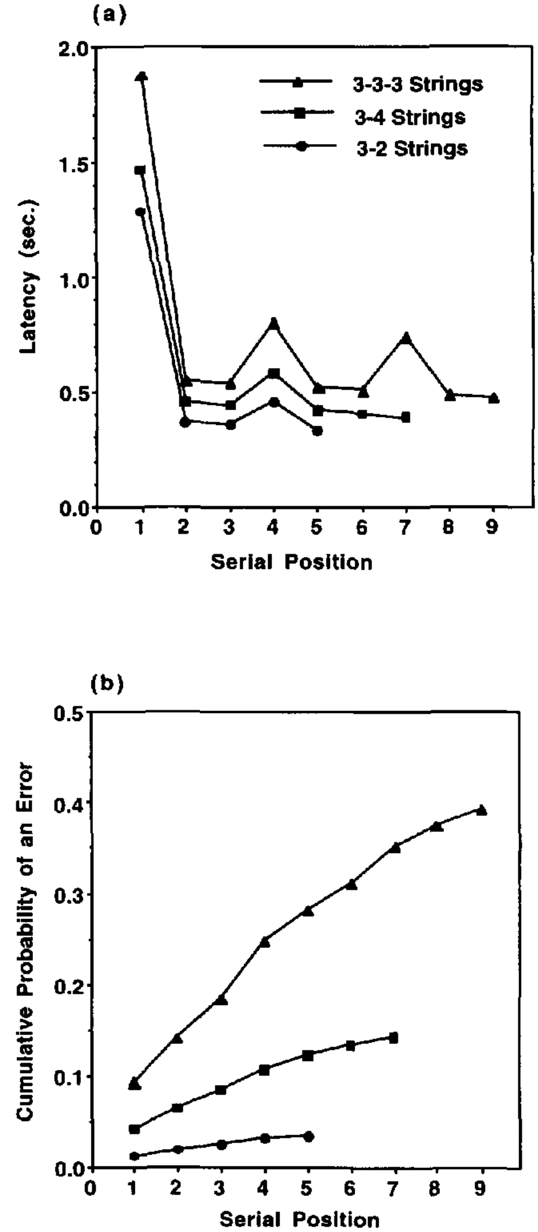

Figure 4. Predictions of the ACT-R theory. A: Mean time to recall each item in the digit strings as a function of serial position. B: Cumulative probability of an error in recall by serial position.

preparing recall. Therefore, t in Equation 7 is 9.31. Thus, the activation of the to-be-retrieved knowledge unit is $-0.5*\ln(9.31) + 3.45/9 = -0.73$. The time for the production to fire (based on Equation 8) is $0.05 + .106e^{-1.43(-0.73)} = 0.352$ s. The retrieval time for the next production, generate-item, is just 0.007, which is much faster because it enjoys the rehearsal advantage from the previous production. The knowledge unit it

[4] For instance, there is only a 0.5% increase in variance not explained if 3.45 is replaced by 1.

retrieves has a recent second exposure and therefore high activation (based on Equation 2, which specifies how to accumulate activation). On the other hand, the typing action time for this production (0.20) means that the total time for the production is 0.207 ms. Everywhere, the latencies are determined by the activation levels of the group or item retrieved.

Note that in Figure 4A, ACT-R produces an effect of list length on retrieval latency. As Sternberg et al. (1978) noted, there is a large effect on initiation time reflecting the time to plan the recall structure. However, each item takes longer to recall because of its lowered activation. Everywhere at similar serial positions (or similar delays), retrieval times are longer for longer lists. The list length effects on latencies are larger at group boundaries because both the next group and the next item must be retrieved.

Probability of errors. Figure 4B shows the cumulative error data from the ACT-R simulation. Compare this figure with Figure 3B. Errors occur because the production *get-next* fails to retrieve an element or the next-group production fails to retrieve a group. Such failures of recall will occur when the level of activation of the knowledge units encoding elements or groups fall below threshold. So, for instance, consider recall of the first element, 3, whose latency we considered earlier. As we noted in the discussion, the expected activation of this element is -0.73. This quantity is substituted in Equation 3 to help obtain the probability of plotted recall in Figure 4B. This requires estimation of the parameters s and τ, which we estimated at $s = 0.32$ and $\tau = -1.63$. Then, according to Equation 3, the probability of successful retrieval is

$$\text{Prob} = \frac{1}{1 + e^{-(1.63-0.73)/0.32}} = .942.$$

A similar calculation for the first group produces a retrieval probability of 0.964. The probability of retrieving the first item is the probability of both retrieving the first group and the first element. Therefore, it is $.964 * .942 = .908$. The error probability plotted in Figure 4B for the first serial position of the nine-item list is 1 minus this quantity, or .092. To find the probability of retrieving the first n elements, we simply take the product of the probabilities of all the get-next and next-group productions that must fire to that point. The reason why we plot cumulative probability of error is that there are multiple strategies for what participants will use after their first error and we wanted to avoid this complication in our initial model. However, we consider strategies on error when we discuss accounting for serial position functions in the published literature.

This completes the description of the basic ACT-R model for serial recall in a spanlike task. We now describe its application to some of the other results in the literature. Table 2 summarizes the fit of the model to this pilot experiment and to the other experiments. At the end of this article, we discuss the variation in parameter values across experiments.

Table 2
Summary of Fits to the Experiment

Parameter	Pilot study	Crannell and Parrish (1957)	Nairne (1992)	Bjork and Healy (1974)	Drenowski and Murdock (1980)	Morra et al. (1991)	Morra et al. (1991) constrained
Activation							
Associative strength (S)	3.45	3.45	3.45	3.45	3.45	3.45	3.45
Decay rate (d)	.50	.50	.50	.50	.50	.50	.50
Time							
Time scale (in ms; F)	**106**	106	—	—	106	106	106
Activation scale (f)	**1.43**	1.43	—	—	1.43	1.43	1.43
Generation time (in ms)	250/digit	250/digit 250/letter 450/word			500/digit	201/1-syl 266/2-syl 352/3-syl 413/4-syl 540/5-syl	201/1-syl 266/2-syl 352/3-syl 413/4-syl 540/5-syl
Time per production	50	50			50	50	50
Start-up time (in ms)	**690**	690			690	690	690
Accuracy							
Activation noise (s)	**0.321**	**0.243**	**0.500**	**0.213**	**0.302**	**0.178**	**0.579**
Activation threshold (τ)	**−1.626**	**−1.007**	**−0.900**	**0.065**	**−0.709**	**0.142**	−0.600
Acoustic confusions		**.013 letters** **.024 words**		**0.148**		**0.196**	**0.180**
Positional confusions			**1.50 times mismatch**	**0.106**			
Strategy							
Probablity of aborting on failure (p)					**0.275**	**0.712**	**0.564**
% of variance explained							
Latency	95.1						
Accuracy	96.2	98.8	93.4	98.6	93.9	97.2	92.9

Note. Numbers in boldface are estimated parameters. Microstructure of the reconstruction process is not modeled. Dashes indicate that parameters were not used; syl = syllable.

List-Length Functions

In the pilot experiment just described, we considered only memory for digits. Digits are special in two ways relative to most other words. First, they can be spoken faster than most other words. Second, they have low phonemic similarity to one another, and so we were able to ignore phonological confusions. Digits contrast with letters on this dimension of phonological similarity, and there is a long tradition of psychological research, such as the work of Conrad (1964), studying acoustic confusions with letters.

A classic study of memory span for different types of material was performed by Crannell and Parrish (1957). Figure 5 shows their results for digits, letters, and words in terms of the mean probabilities of perfectly recalling lists of various lengths. These functions show the gradual drop off typical of list-length functions but are shifted lower for letters than for digits and lower for words than for letters. The reason that letters are worse than digits is presumably their greater acoustic confusability. The reason that words are worse than letters presumably reflects the longer time to say them.

We fit the same model to this data as we did to our pilot data. Typical of the published literature, Crannell and Parrish (1957) did not report generation time, only the proportion of perfect recall. Because recall depends on decay of activation, which depends on time, we needed to make some assumptions about latency of recall. We used the same f and F parameters from the pilot experiment to transform activation values into latencies. For letters and digits, we kept the generation time (for generate-item) at 200 ms. Adding this generation time to 50 ms for the get-next production gave us a minimum reading rate of one digit per 250 ms.[5] For the words we raised the generation time to 400 ms, which gave us a minimum reading time of 450 ms for one-syllable words to correspond to the estimate of Baddeley et al. (1975). Thus, except for the slower generation time for words, we used the same timing parameters as in our pilot experiment. Because there were more lengths of lists, we had to assume more grouping structures. We generalized our grouping structures from the pilot study and assumed a single group of 4 for 4 digits, 3–2 for 5 digits, 3–3 for 6 digits, 3–4 for 7 digits, 3–3–2 for 8 digits, 3–3–3 for 9 digits, 3–3–4 for 10 digits, 3–3–3–2 for 11 digits, and 3–3–3–3 for 12 digits. The restriction to groups of about 3 digits is suggested by a number of experimental results (e.g., Broadbent, 1975; Ryan, 1969; Wickelgren, 1964). However, the results depend little on the exact group structure assumed.[6]

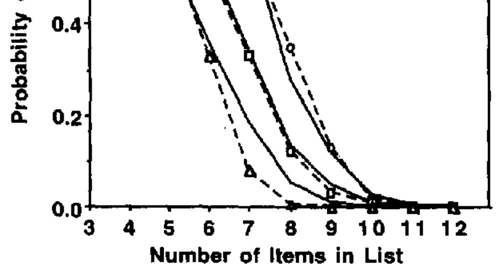

Figure 5. Predicted (solid lines) and observed (dotted lines) probability of reproducing lists of various lengths. Data are from Crannell and Parrish (1957).

To model these results we have to model the effects of acoustic confusion on recall. Our theory ascribes it to retrieval of the wrong item by the generate-item production given earlier. According to ACT-R, this is a result of partial matching controlled by Equation 6. All the lexical items on the list will be active and the one retrieved will be the one with the highest goodness-of-match score. A mismatching item will receive a mismatch penalty P, but, because of noise in the activations, the wrong item can be retrieved. The probability of this is the function of the size of the mismatch penalty, with smaller penalties for more similar items. The probability of confusing another item for the correct lexical item, given that they are equally active, is

$$c = \frac{1}{1 + e^{P/s}}, \tag{9}$$

where P is the degree of mismatch of the wrong item and s is the parameter controlling activation noise. In the model to follow, we only estimated the confusion probability c.

We estimated four parameters to minimize the squared deviations between data and predictions. Two of these were the noise parameter, s, and the threshold parameter, τ. The other two were the probabilities of confusions among letters and among words. We estimated a probability c, separately for each type of material, that there would be a confusion between a pair of items (assuming $c = 0$ in the case of digits). If there are n items, there are $n(n - 1)$ possible confusions. Thus, the probability of no confusion is $(1 - c)^{n(n-1)}$. The parameters estimated were $s = 0.243$, $\tau = -1.007$, $c_{\text{letter}} = 0.013$, and $c_{\text{word}} = 0.024$. For comparison, the values from the pilot were $s = 0.321$ and $\tau = -1.626$. The lower value of the threshold parameter (τ) than in the pilot data reflects the somewhat better digit span performance displayed by pilot participants. Figure 5 displays the fit of the model to the data using these four parameters. With these parameters the model did a good job of accounting for the data. The only possible exception was the 4-digit lists, on which participants did somewhat better than the predictions of the theory. This may reflect an effect of an acoustic buffer at short delays. Elsewhere in this article we comment on potential complications created by such a buffer.

These data confirm the ACT-R assumption that there is not a drop-dead span length but a gradual decrease in performance.

[5] The assumption is that reading time does not involve any significant retrieval time and only the action time contributes to latency.

[6] However, they would be sensitive if we were modeling item-by-item recall time data.

Of course, these are average curves, but in our laboratory research we have found that individual participants also show gradual functions. Individual functions are sometimes a little steeper, but this can be accommodated by lowering s, the noise parameter in our model.

Positional Confusions

In the discussion of Crannell and Parrish's (1957) data, we considered acoustic confusions. To this point we have ignored positional confusions. Such errors probably contribute to all the data, but it is difficult to assess the proportion of positional errors in the data like that considered so far. When participants fail to recall, they may guess and produce what appear to be order errors. On the other hand, when participants misorder an item, they may recognize that their answer is incomplete or out of order and abort the recall effort, producing what appear to be omission errors. One way these ambiguities are dealt with in the literature is to try to eliminate the problem with item recall by giving the participants the items and simply asking them to order the items. Figure 6 contains some data from Nairne (1992) that uses this method.[7] He had participants reproduce 5-item lists either 30 s, 4 hr, or 24 hr after studying the lists. The results are presented in terms of the proportion of items from every serial position placed in each serial position. What is striking about these data is the similarity of the generalization gradients at the various delays. His results are similar to positional generalization gradients obtained in other tests of immediate recall (e.g., Lee & Estes, 1981) but allowed us to test ACT-R's predictions for various delays.

The figure also presents the predictions that he derived from Lee and Estes's (1981) perturbation model. This model assumes that during each unit of time there is a fixed probability of inverting a pair of items and that forgetting is produced by the accumulation of such perturbations over time.

We fit a simple ACT-R model to these data. In contrast to the perturbation model, ACT-R assumes that forgetting is attributable to decay of base-level activation. We used the same decay model as was used in fitting the more immediate memory tasks in the previous experiments. Because this is a reconstruction task, it is difficult to model the exact steps of processing. Therefore, we do not model the step-by-step reconstruction process but the net effect, which will depend on the activation of the knowledge units. Fortunately, because the recall is not immediate it is not critical to model the exact timing of the reconstruction steps. At longer delays the effects of slight differences in timing become insignificant because the logarithmic function for decay (Equation 7) compresses these differences. Therefore, we simply estimated the mean activation of these memory traces after 30 s, 4 hr, or 24 hr.

If participants cannot retrieve the item they cannot retrieve its position, and so we assume they will guess from among the available positions. If they can retrieve the item that came from position v, the probability of misrecalling it in position d is a function of the difference between the two positions. Assuming the partial matching Equation 6, we set the mismatch penalty, P, to be $g*|v-d|$ where g is a scaling of the mismatch. The partial match between positions will be accepted if the normally distributed noise added to the item gives a match value greater than that of the item that should go in position d. Using the logistic approximation to the normally distributed noise, this becomes

$$\text{Prob(confusing position } v \text{ for } d) = \frac{1}{1 + e^{g|v-d|/s}}. \quad (10)$$

The probability of correctly placing the item was set to one minus the probabilities of all the misplacements.

The other complication is that, because the words are presented at recall time they also are sources of activation for their traces. Thus, they contribute to the W_jS_{ji} in Equation 2. Thus, the total activation of a trace is

$$A_i = -0.5 \ln(t) + 3.45/5 + 3.45.$$

The first term ($-0.5 \ln t$) reflects decay with time t, the second term (3.45/5) reflects activation from the list, and the third term (3.45) reflects self-activation.[8] This is the same equation as in other models except for the 3.45 self-activation.

Because of the possibility of misrecalling positions and the guessing of the position for items that cannot be correctly placed, the model has some subtlety in its interactions. The following is a specification of the precise algorithm for item placement: We assume that the items are presented to the participant in random order. For each item we assume that the participant tries to recall it and its position. The probability of recalling it is governed by its activation and Equation 3 given earlier. If the item can be recalled, its position is recalled according to the specification in Equation 10. If the item can be recalled and its position is not already occupied, it is placed in that position. Otherwise (i.e., it cannot be recalled or its recalled position is already occupied), it is temporarily put aside. After an attempt has been made to put all the items into position, any that were put aside are randomly placed into the remaining positions.

This is a complex, interactive procedure, and it was not possible to estimate parameters using a simple Excel program. Rather, we wrote a LISP program and performed a Monte Carlo simulation. We set the activation noise to 0.5 and searched for values of the threshold parameter, τ, and scale factor, g/s, that gave the best fit in terms of minimizing the squared deviation between prediction and data.[9] The parameters estimated were $\tau = -0.900$ and $g/s = 1.50$. The fits of the ACT-R model also are displayed in Figure 6. The ACT-R model does fit better than the perturbation model. Its R^2 is .934 compared with an R^2 of .857 for the perturbation model. We do not want to make too much of this comparison of fits because it is not clear that the optimal parameters were estimated in Nairne's (1992) application of the per-

---

[7] We would like to thank James Nairne for providing us with his data and the fits of the perturbation model.

[8] It was not clear whether self-activation should participate in the capacity limitation assumption (Equation 5). This model assumes not. It is actually inconsequential to the predictions for this experiment because lowered activation can be compensated for by changes in the estimates of τ. The distinction would be important had list size been manipulated in this experiment.

[9] This LISP file is available at the same location the simulations and Excel code are found: by following the "published models" link from the ACT-R home page: http://act.psy.cmu.edu.

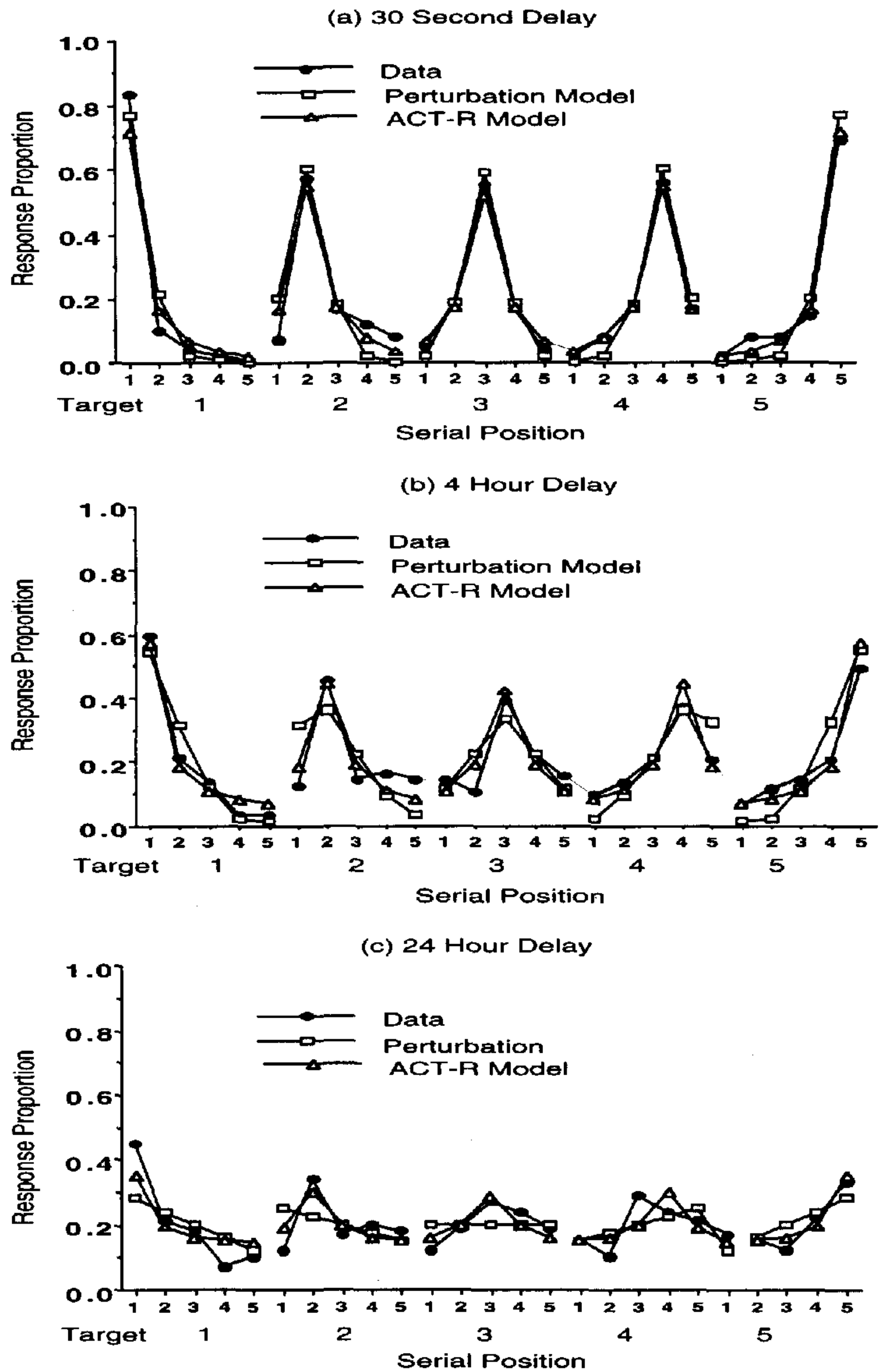

Figure 6. Data from Nairne (1992) and fits of the perturbation and ACT-R models. The data are plotted as a function of the target position and the reproduced position. Data are plotted separately for 30 s (A), 4 hr (B), and 24 hr (C).

turbation model. Nonetheless, we can conclude that the ACT-R partial-matching analysis of positional confusions is competitive with other published models.

One feature of the ACT-R model contrasts significantly with the perturbation model. Forgetting over time in ACT-R is not a function of increased positional confusion but a function of decay of activation. ACT-R does not assume that the probability of positional confusion depends on the level of activation of the trace. A mismatch produces the same penalty at all levels of activation and will be accepted if the noise in the activation reverses the mismatch penalty. That noise is constant for all levels of activation. Thus, ACT-R, unlike the perturbation model, does not predict that positional confusions increase with time. Performance declines because the items can no longer be recalled at all and participants are just guessing. Thus, we predict that positional uncertainty gradients do not change their basic shape with time but only become flatter. Nairne (1992) noted that the perturbation model has difficulty with the data at the 24-hr delay. The only way the model can produce the low level of accurate recall is to assume nearly totally flat positional confusions. By contrast, ACT-R has time-based forgetting, which is independent of positional confusions. According to ACT-R, one cannot confuse items that cannot be recalled. ACT-R is relatively unique in this assumption that positional confusions do not increase with delay (however, see Drenowski, 1980).

Finally, note that ACT-R predicts, as the data show, best performance for items on the end positions 1 and 5 (i.e., the peaks are highest for these positions). More generally, we predict a bowed-shaped serial position effect for this task in terms of accuracy of positional placement. This is because end items have adjacent positions only on one side for confusion. Later in this article we consider serial position curves in recall but not reconstruction. Such recall curves have added features because the position of the items is correlated with temporal order of output. Nonetheless, they also show special advantages for the first and last positions.

Acoustic and Positional Confusions

Previously, we modeled separately acoustic confusions (Crannell & Parrish's, 1957, data) and positional confusions (Nairne's, 1992, data). In this section we combine a model of both types of confusions. First, however, we comment further on acoustic confusions. As noted, there is a long tradition of research studying acoustic confusions. Different levels of recall are obtained as a function of how similar the list items are. We do not model acoustic errors in the detail with which we have modeled positional errors in Nairne's data because this would deviate from our focus on the serial nature of the list memory and require developing a theory of acoustic similarity. Nonetheless, it is important to consider the basic outline of how such effects are to be incorporated in ACT-R because they have played such a major role in the research of the field and are clearly an important piece of the span limitation.

One tradition in the research on short-term memory has been to attempt to account for recall failure entirely in terms of such confusions denying other factors. Conrad (1965) proposed that order confusions were really confusions of acoustically similar items. Perhaps the strongest case of an acoustic forgetting theory is Posner and Konick's (1966) "acid bath" theory, according to which interitem similarity is supposed to create an environment of metaphoric "acidity" in which items slowly dissolve over time. This type of theory leads to a critical prediction that the difference between memory for acoustically confusable lists and memory for nonconfusable lists will increase with time. This was supported in Posner and Konick's experiment contrasting immediate recall with delayed recall, but, to our knowledge, this has never been shown in the contrast of various nonzero delays. In fact, in Posner and Konick's original report, there was a contrast of nonzero delays that failed to show such an effect.

The low levels of acoustic confusions at zero delay may reflect the correcting influence of an echoic buffer. There may well be an acoustic buffer of this sort operating at short delays in many experiments. For instance, Crowder and Morton (1969) proposed the existence of a precategorical acoustic storage that is the source of speech phenomena such as the suffix effect. Our model does not represent such a transient acoustic store.

ACT-R does not predict increased acoustic confusions with delay for the same reason it does not predict increased positional confusions with delay. Acoustic confusions occur because of the retrieval of a similar-sounding word through partial matching (in the generate-item production). The probability of this partial matching does not depend on the activation, but the probability of retrieving the item does. As a consequence, with delay a greater portion of errors will be omissions or random guesses. Thus, ACT-R predicts that a smaller portion of the errors will be acoustic confusions with delay, a prediction that is generally confirmed.

The ACT-R model also makes an important prediction about the relationship between positional and acoustic errors—that they should be statistically independent—which is counter to Conrad's (1965) claim that the positional errors were actually acoustic errors. In ACT-R positional errors are produced by partial matching in the get-next production, whereas acoustic errors are produced by partial matching in the generate-item production. We discuss in detail the experiment by Bjork and Healy (1974), which confirmed this prediction.

In Bjork and Healy's (1974) experiment participants saw lists of 4 letters presented at a rate of 0.4 s per letter. The letters came from a pool of 12 letters, half of which were acoustically confusable and half of which were not (control letters). Participants were to recall the letters after reading an intervening list of digits also presented at the rate of 1 digit per 0.4 s. Figure 7A shows the error data from their experiment as a function of number of intervening digits (3, 8, or 18). The total number of letters misrecalled increased with delay from about 20% to about 60%, with more errors being made in recalling confusable letters. Bjork and Healy also calculated the proportion of the total errors that were acoustic confusions or positional confusions. An error was scored as an acoustic confusion if the letter recalled in a position was acoustically similar to the correct letter for that position. Each confusable letter had two other similar letters, and for each control letter two other control letters were arbitrarily designated as in the confusion set for purposes of scoring. An error was scored as a positional confusion if the letter recalled in a position came from some other position in the list. In this scoring scheme, positional and acoustic errors are orthogonal categories—a given error can be both an acoustic and a positional error, just one, or neither. One noteworthy result in Figure 7A is that, although the number of errors is increasing, the proportion of the errors that are confusion errors is decreasing. Thus, increased forgetting cannot be attributed to increased confusions.

An important aspect of Bjork and Healy's (1974) design was that two of the letters in the recall list of four were acoustically confusable and two were not. This enabled Bjork and Healy to calculate the probability of an acoustic confusion error condi-

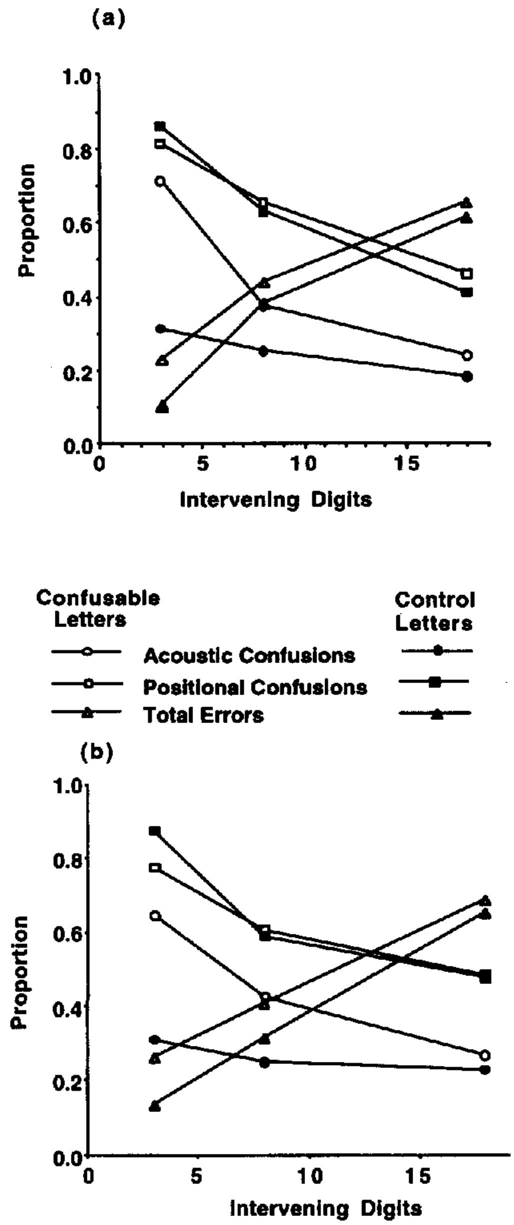

Figure 7. Results (a) and predictions (b) for Bjork and Healy's (1974) experiment. Note that the error curve gives proportion of trials that are errors, whereas the acoustic and positional confusion curves give the proportion of errors that are these types of confusions.

tional on it being a positional error. In contrast to what the Conrad (1965) theory would predict, conditional confusion rate was not much different from the unconditional confusion rate. This supports ACT-R's separation of positional errors from acoustic errors.

We applied ACT-R to predict the results of this experiment, and the results are presented in Figure 7B. Because participants were well aware of the 12 possible letters and were apparently required to generate 4 letters for each recall, we treated this experiment as a reconstruction task (write 4 out of 12) in the same way we treated Nairne's (1992) experiment. That is, we did not model the detailed timing of output but simply used the manipulated delay time. The equation for activation was

$$A = -0.5\ \ln(t) + 3.45/4,$$

which, unlike the activation equation for Nairne's (1992) experiment and like the equations for the other experiments, it has no term to reflect the self-activation from the letter because the letters were not presented. To fit the data we estimated that the probability of recalling a letter out of position (a partial-matching error in get-next)[10] was .106, whereas the probability of recalling an acoustically similar letter (a partial-matching error in generate-item) was .148 in the case of confusable letters. The probability of an acoustic confusion in the case of control letters was assumed to be zero.[11] These two probabilities were treated as independent. To fit the overall forgetting, we estimated the two parameters that control the probability of retrieval of an item to be $s = 0.213$ and $\tau = -0.065$.[12] Finally, we needed to estimate one other parameter to properly model the guessing process. If an item cannot be retrieved and a participant is forced to guess, we assume that he or she will guess some highly active letter. This produces the restriction of guesses to the 12 letters from the experiment but also favors the letters in the current list because they will be the most active. That is, even if the participant could not recall the item and was just guessing from the set of 12, there would be an increased chance of guessing a letter from the most recent list.[13] The probability of guessing a list letter was estimated to be 2.15 times the probability for comparable nonlist letters. This assumption, which is consistent with recent work on implicit priming (e.g., Jacoby, Toth, & Yonelinas, 1993; Reder & Gordon, 1997), is required to fit a number of aspects of the data, including the slower decay of positional confusions than acoustic confusions. Thus, according to this model some positional confusions are produced by a bias to recall items from the current list regardless of any explicit memory of the letter in that list.

The general quality of the fit is good. We not only modeled the data in Figure 7 but the more fine-grained data cross-classified by acoustic and positional errors. The success of this fit largely is a credit to ACT-R's assumptions about the independence of item loss, positional confusion, and acoustic confusion and about the decay process producing item loss.

[10] Because Bjork and Healy (1974) did not report specific position-to-position confusions, we simply estimated an overall probability of a positional error rather than specific positional errors that we modeled with respect to Nairne's data.

[11] Nonetheless, some errors are classified as acoustic in Figure 7 because of random guessing of the paired words for the control set.

[12] The threshold parameter τ is much lower than in previous experiments, but in this experiment the study time was much less. We discuss parameter variations in the Conclusions section.

[13] Explicit recall of the item requires retrieval of the knowledge unit encoding that the letter occurred in the list, which will depend on the unit's activation. Guessing a letter depends on the activation of the letter representation. Thus, the distinction is between token activation (recall) and type activation (guessing).

Serial Position Curves

Another basic description of serial recall is the serial position curve, which is a plot of the probability of correct recall as a function of serial position of the item. There are numerous paradigms for obtaining such curves and numerous definitions of a correct response for a serial position. One of the two common paradigms involves giving participants sheets to write the answer down and not constraining the order of recall. The other common paradigm requires the participant to recite the items in the order they occurred. Sometimes, recall is scored for whether the item is recalled at all. Other times it is scored for whether the item is recalled in the correct serial position. Different procedures and different scoring measures produce different serial position curves. Most procedures yield curves that show a strong primacy effect, with best recall occurring at the beginning of the list. Many procedures yield a somewhat weaker recency effect, with recall improving substantially for the last item.

From the point of view of a process model it is hard to make predictions about serial position curves because the participant has the option for so many different strategies for recall when recall is not perfect. When recall is perfect, it is plausible to assume that the participant simply recalls the items in order. However, when recall is imperfect, there are all sorts of ways for participants to respond to their errors. If the order of recall is not constrained, the participant can skip over difficult intermediate items and recall later items. Even when the order of recall is controlled, the participant has at least three possible options when faced with an item that he or she cannot recall: abort the whole recall at that point, skip to the next item he or she can recall, or guess some item (often the participant is allowed to say something like "blank") and continue recall. That is why in our pilot data we tried to fit points of first error and ignored modeling possible recall after that error.

Figure 8A shows serial position curves for different list lengths (from Drenowski & Murdock, 1980) when the recall was constrained to be left to right and when the data were scored correct only if they were in the correct position. Figure 8B shows the same data when an item was scored correct independent of serial position. All the data show strong primacy effects, but the data in Figure 8B also show clear evidence of a recency effect for the last item for 6- and 7-digit lists. The probable reason for an absence of a strong recency effect for the last item in Figure 8A is that participants in Drenowski and Murdock's experiment were not allowed the option of indicating blank. Thus, if they omitted an item, their recall would be out of order by the end of the list.[14] When participants are allowed an option of a blank response, one generally gets recency in ordered recall.

As we saw with respect to Nairne's (1992) data (see Figure 6), ACT-R does predict an advantage of first and last positions because of decreased serial position confusions. End items can only be confused on one side. However, it is difficult to model mathematically interactions of such serial position confusions with the timing of recall. (In Nairne's task, we ignored timing processes in the serial reconstruction task where it did not matter because testing was delayed.) Because of such complications, we ignore such positional confusions in the model of Drenowski and Murdock (1980) data to follow.[15] Our goal is to determine how much of the data we can account for simply assuming the time decay and associative interference implied by Equation 7. ACT predicts the general decreasing trend in the recall functions because of increased time to get to these items in recall. It also can account for the different height of these functions because of associative interference. The question is whether it can account for the relative magnitudes of these effects.

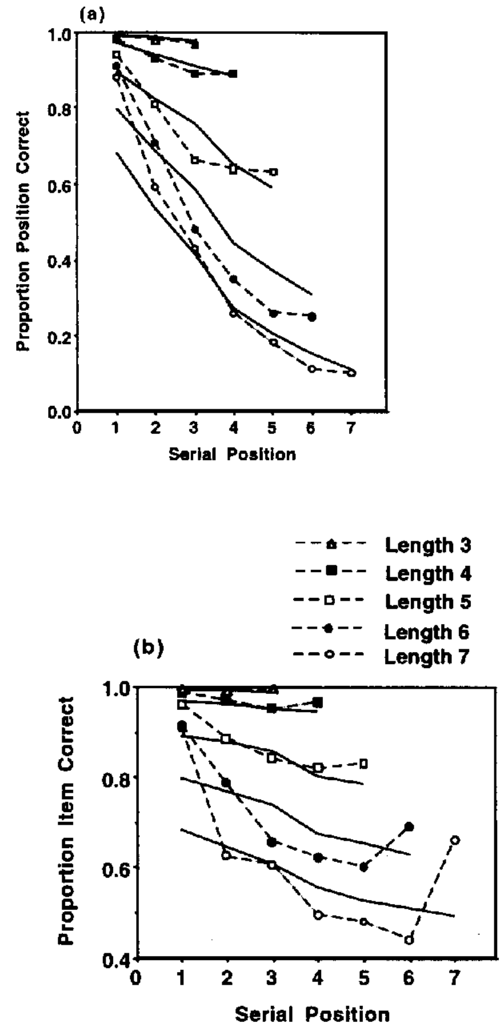

Figure 8. Serial position curves (predicted are solid lines and observed are dotted lines) from Drenowski and Murdock (1980) scored for correct order (a) and for item recall (b).

Figures 8A and 8B also show the fits of an ACT-R model to the data item. We used the same timing parameters as for the pilot experiment except that we assumed slightly longer action times to correspond to the longer reading times for two-syllable words. The action time for generate-item was 450 ms, and the

[14] However, participants were encouraged to guess if they could not recall an item and so had a way to keep positions correct.

[15] More recently, Anderson & Lebiere (in press) we have modeled the effect of positional confusions on the serial position curve.

action time for get-next was 50 ms, resulting in a 500-ms reading time per word for the two-syllable words (consistent with the reading time estimates of Baddeley et al., 1975). We also modified the simulation of the task to incorporate the $\frac{2}{3}$ s presentation rate used by Drenowski and Murdock (1980) rather than the 1-s rate used in the pilot experiment and in Crannell and Parrish's (1957) experiment.

We assumed that in Drenowski and Murdock's (1980) procedure, participants would get an item in correct position only if they could recall it and all of the previous items correctly. This seemed reasonable because participants were not allowed to skip items, but it does ignore the strategy of skipping an item but inserting some guess to keep the serial position correct. To model item recall, we used two alternative strategies. One was that participants would abort their recall when an error occurred, and, as a consequence, the probability of item recall would be the probability of position recall. The other was that participants would skip over the missing item and just recall the next item that they could. Then, the probability of recalling an item was the probability that the item would be available regardless of whether prior items had been skipped. We estimated a probability p of implementing the first strategy of aborting when an error occurred. The three estimated parameters were $p = .275$, $s = .302$, and $\tau = -.709$. Again, our estimates of s and τ were similar to the estimates for the previous experiments. The low value of p is reasonable because participants were encouraged to keep going.

Without incorporating positional confusions, the predicted serial position curves in Figure 8 failed to capture the magnitude of the primacy effect or the recency effect when scored for item recall. Except for these end anchor effects, we captured the shape of the serial position curves (particularly the rate of decline with serial position) as well as the absolute height of the curves for various list lengths. Overall, we accounted for 93.9% of the variance. The importance of this demonstration is that we modeled the combined effect of time-based processes that produced the drop off and associative interference that produced the separation of the curves.

Effects of Word Length on Various Span Measures

Baddeley et al. (1975) used a number of paradigms to study the effect of word length (in syllables) on span. However, their most reported results involve a paradigm in which participants are asked to recall as many words out of five as possible and span is measured as the number of words correctly recalled in position. This has become an alternative measure of span in the literature, in contrast to the maximum number of words recalled perfectly. Morra et al. (1991) provided some data from Italian in which they compared the performance of participants given the more traditional span test (the maximum span a participant can reproduce) with performance under the correct-out-of-five procedure. We chose to model their experiments because they involved a more elaborate set of data. They also examined the effects of word length and articulatory suppression. These results are shown in Figure 9 along with the ACT-R predictions. It is important to be clear about how these data were scored. The conventional span procedure involved starting up from lists of three, increasing length by one each trial, and stopping when the participant had failed to perfectly recall two consecutive lists. Span was measured as the longest list reproduced or two if no lists were reproduced. The correct-out-of-five procedure involved verbal recall, and an item was scored correct only if it was recalled in the correct position. We used the same scoring procedures as Morra et al. in our ACT-R simulation, and this served as the basis for the predictions. As with the model for Drenowski and Murdock's (1980) data, we assumed a strategy mixture of aborting on recall failure with probability p and of continuing and recalling later items with probability $(1 - p)$. In this case, however, participants were allowed to skip to the correct position, and so we assumed that these later items would be correctly recalled when scored for order. We estimated p to be .712, which was much higher than the value for Drenowski and Murdock's data. In addition, we had to model an effect of articulatory suppression. We simply modeled this by assuming articulatory suppression had a fixed probability of interfering with each item in the list. This probability was estimated to be $c = .196$. Thus, if an item had a probability R of being recalled without articulatory suppression, its probability of recall was $(1 - .196)R$ with articulatory suppression.

The effects of articulatory suppression are complex. One thing articulatory suppression does is to inhibit rehearsal during study. Such rehearsal processes have not been modeled in our simulations but are discussed in the Conclusions section of this article. Another effect of articulatory suppression is to prime acoustically similar items that can be retrieved rather than the target items (see our discussion of acoustic confusions). We do not pretend to model in detail the complex interaction of such effects but are simply modeling the net effect as a lowered memory for the items. The question of interest is whether this simple net effect will interact properly with word length and span procedures that we are modeling in detail.

Morra et al. (1991) also collected articulation rates for each

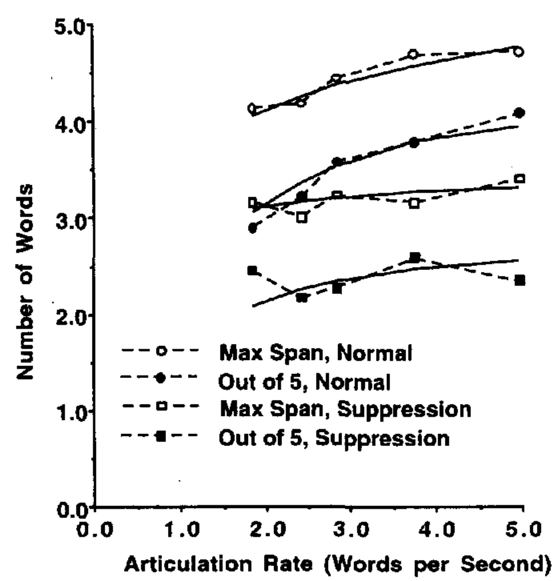

Figure 9. Predicted (solid lines) and observed (dotted lines) span measures as a function of articulation rate. Data are from Morra, Tressoldi, Mazzoni, Sava, and Zucco (1991).

word length and the data in Figure 9 are plotted as a function of articulation rate.[16] These articulation rates were used to set the generation times in the ACT-R model. In addition to the p (the strategy mixture parameter) and c, the probability of interference, we estimated our standard two accuracy parameters: $s = 0.178$ and $\tau = 0.142$.

There are several noteworthy features of the data. For instance, the conventional span measure yields the higher estimate of span. This is just a reflection of the details of the procedures: The out-of-five procedure is bounded above by five, whereas in the conventional span task there are two chances for recall. If either of these features is changed, the ordering of the two measures might reverse. There is no inherent superiority of the conventional span measure. Indeed, the ACT-R theory implies that there is no real concept of a span as a psychological primitive. Nonetheless, it is to ACT-R's credit that it predicts the relative ordering of these two procedures.

Note in Figure 9 that the effect of articulatory rate is greatly reduced under articulatory suppression replicating a frequently obtained result (e.g., Baddeley et al., 1975, 1984; Murray, 1968). The reason for this in our model of this task is because the effect of articulatory suppression is to make the early items on the list less available and so to make it less important how long it takes to get to the end of the list. However, there still are small effects of rate under articulatory suppression. A survey of the literature reveals that suppression usually reduces (often to nonsignificant levels) but does not eliminate the effect of word length. In addition to predicting the relative slopes of the functions in Figure 9, ACT-R also simultaneously predicts their heights.

Memory Span: Overall Evaluation

The literature on the memory span is vast, and it is hard to judge how much of it has been addressed by this ACT-R model. As one measure, we consider the phenomena that Burgess and Hitch (1992) claimed are the important empirical constraints from the human data: (a) declining immediate recall with increasing list length; (b) effects of phonemic similarity, word length, and articulatory suppression; (c) shape of the serial position curve; and (d) types of recall error (transpositions, acoustic confusions).

As we have tried to show in the preceding sections, our model does deal with all these results except the serial position curves. With respect to serial position curves, we were able to show (see Figure 6) that ACT-R predicts an advantage for end positions because of positional confusions and that it does predict decreased performance with position because of recall failure (see Figure 8). However, for reasons of tractability, we were not able to put these together into a single-model fit.

In addition to the results listed by Burgess and Hitch (1992), we would like to emphasize two other results that are basic to the ACT-R architecture. One is the ability of the model to account for the complex timing patterns observed in recall. The other is the prediction about the pattern of confusions over time. As seen in Figures 6 and 7, ACT-R predicts that passage of time leads to loss of information and possibly random guesses. Systematic confusions among similar items (acoustic or positional) do not increase with time. This is because time lowers the activation of all list elements, making them less available for both correct recall or intrusion. Partial matching is produced by random activation noise, which can reorder the match scores (Equation 6), and this noise does not interact with the activation levels.

Conclusions

Table 2 provides a summary of the ACT-R fits to the data. It lists the parameters and the percentage of variance explained. Many of these parameters were set a priori rather than estimated in fitting experiments. The associative strength and decay rate parameters were set a priori and not estimated. In the context of this research, these two parameters essentially serve as a way to weight the contribution of list length and time to the overall memory limitation (see Equation 7). Only our pilot experiment had data available on recall latency, and so it was used to set the time scale (F) and time activation scale (f) parameters for the rest of our model fits. The activation noise parameter (s) and the activation threshold parameter (τ) were the two parameters we estimated anew for each experiment to reflect the participants' overall performance. We also used articulation times from the experiments or published estimates to determine the generation time.

In addition, we estimated other accuracy parameters to be sensitive to factors that were being manipulated in some of the experiments. Therefore, for instance, we estimated probabilities of pairwise acoustic confusions in Crannell and Parrish's (1957) experiment and of acoustic confusions in Bjork and Healy's (1974) experiment and the Morra et al. (1991) experiment. The parameters from these last two experiments appear to be much higher than in Crannell and Parrish's experiment, but when one calculates for Crannell and Parrish's experiment the probability that an item would be confused with any one of the members of the list, the confusion probabilities are approximately equal in all experiments. For instance, the probability that a word from a 7-item list will be confused with some other word in Crannell and Parrish's study is $1 - .976^6 = .136$.

Both Nairne's (1992) and Bjork and Healy's (1974) articles required estimating parameters to control positional confusions. Because we addressed the total confusion gradient in Nairne's experiment, we estimated a mismatch scale parameter, whereas we estimated only an overall probability of a positional confusion in Bjork and Healy's study. However, they do lead to comparable probabilities of confusion. Nairne's parameter implies probabilities of confusion of 18.2%, 4.7%, 1.1%, and 0.2% for items that are 1, 2, 3, and 4 items apart in Nairne's 5-item lists, respectively. If we assumed the items in the 4-item lists used in Bjork and Healy's experiment are proportionately more discriminable (i.e., adjacent items are 1.33 units apart in a 4-item list if they are 1.0 units apart in a 5-item list), the probabilities of confusion would be 11.9%, 1.8%, and 0.2% for items 1, 2, and 3 apart for the 4-item list from Bjork and Healy, respectively. The probability of a positional confusion was estimated as 10.6% for Bjork and Healy's experiment. So this probability is in the ballpark of what would be estimated from the Nairne parameters.

It is true that acoustic confusions probably play a significant

---

[16] The articulation rates are faster than those reported for English.

role in most experiments that do not involve digits. Thus, there were probably acoustic confusions not modeled in Nairne's (1992) experiment, Drenowski and Murdock's (1980) experiment, and intralist confusions in the Morra et al. (1991) experiment. However, we adopted the strategy of not incorporating these into our models when they were not manipulated in the materials. Similarly, we did not explicitly model positional confusions in experiments in which these were not measured. The effects of ignoring these factors would be partially compensated for by shifting values in the threshold parameters.

The threshold parameter, τ, reflects the overall level of performance. All other things being equal, the lower it is the more likely items will be retrievable. There are numerous things that can be reflected in its variation. As we already noted, it can reflect other performance limitations not modeled. It can reflect population differences, as in the case of the differences in digit span performance between our pilot population and Crannell and Parrish's (1957) participants. It also can reflect degree of study of the material. In some cases, what is being measured as τ probably reflects the base-level activation and not threshold, but retrievability will be determined by the difference between base-level activation and the threshold. Thus, actual differences in base-level activation will be reflected in estimated differences in the threshold. For instance, the thresholds were higher in Bjork and Healy's (1974) experiment, in which study time was 0.4 s, and in Drenowski and Murdock's (1980) experiment, in which study times were $\frac{2}{3}$ s rather than the more usual 1 s. Thus, the differences in the threshold parameter might better be conceived of as reflecting the differences between base-level activation and threshold. The one setting of τ that does seem out of line is in the Morra et al. (1991) experiment, in which the positive value is larger than the other experiments. However, the fit to that experiment is not sensitive to this value. We tried fitting the Morra et al. experiment constraining τ to be -0.6. The parameters of that fit are reported in Table 2, and, as can be seen, the variance explained stayed above 90%. Also, the estimate of probability of aborting recall shifted to 0.564 to make it more consistent with the estimate from Drenowski and Murdock.

As Table 2 makes clear, there were two basic models used in fitting the data. For all but Nairne's (1992) and Bjork and Healy's (1974) experiments, we used the model illustrated in Table 1. That is, our model of the task involved simulating how the participant stepped moment by moment through the list. This process simulation was critical in giving us the time parameters that were used in estimating activation levels. On the other hand, because both Nairne's and the Bjork and Healy's experiments were effectively reconstruction tasks and were administered at delays, we simply used the time delays as our time estimates and did not simulate the reconstruction process. Such a simulation would have been complex, and small timing differences at substantial delays are not critical to the predictions.

Nature of the Span Limitation

These data and the models fit to them illustrate the complex system of factors that go into producing the memory span limitations. To review, they are as follows: (a) As the list gets longer, source activation has to be divided among more items. (b) As the list gets longer, base-level activation decays more because it takes longer to get the end of the list. (c) As the list gets longer, there is more opportunity for positional confusions. (d) As the list gets longer, there is more opportunity for acoustic confusions. (e) Successful recall of a longer list requires more things to happen successfully, any of which can go wrong. We hope that we have demonstrated that all these factors are necessary to understand the span limitation.

Comparison to Other Theories

It is worth trying to place this theory within the space of theories for serial memory. With respect to its hierarchical chunk organization of the list, it has strong similarities to the theories of Johnson (1970) and Estes (1973). Johnson examined what he called "transitional error probabilities" (TEPs), which are the probabilities of incorrectly recalling the next item given that the current item is correctly recalled. He showed that TEPs were much higher between chunks than within. The current theory predicts this because two retrievals must be successful at transition points (retrieval of the chunk and next item) rather than just one retrieval (of the item) within a chunk. On the other hand, in contrast to the ACT-R model, Johnson assumed that recall of items within a chunk was all or none, which is not an assumption shared with the current ACT theory. As an empirical matter, there certainly is partial recall of chunks (i.e., within-chunks TEPs are not zero).

In allowing for imperfect recall of a chunk once accessed, this theory is much more like Estes's (1973) hierarchical theory. However, Estes made forgetting a result of positional inversions in a rehearsal process, whereas the current theory makes recall failure a function of many things. Positional confusions are one factor, but these are produced by the partial-matching process and do not depend on the participant's rehearsing or on time. As noted earlier, Estes's perturbation model cannot handle the long-term confusion patterns such as in Nairne (1992). In contrast to Estes, Johnson's (1970) theory was really a theory of long-term memory. The ACT-R theory does not make a distinction between these two types of memory.

In contrast to ACT-R and the models of Estes (1973) and Johnson (1970), which hold that the representation is fundamentally hierarchical, a number of theories propose that serial memory depends on item-to-item associations (Lewandowski & Murdock, 1989; Shiffrin & Cook, 1978). Such theories have difficulties with the effects that have been cited as evidence for hierarchical representations. On the other hand, it is a more subtle issue whether information about chunk order and element order within a chunk is encoded according to position or association. Johnson (1970) and McNichol (1978) provided evidence in favor of the positional conception (and the ACT-R conception) in that participants showed transfer among lists that would preserve position information within the hierarchy but not associative connections.

According to the ACT-R theory, loss of information is both time based and interference based. Thus, ACT-R is different from both Baddeley's (1986) 2-s loop theory (although Baddeley clearly included interference processes in his more general conception) and the interference theories that attribute forgetting to acoustic confusions (e.g., Conrad, 1965; Posner & Konick,

1966) or positional confusions (Estes, 1973). On presentation, items receive an activation boost that decays according to a power law. As noted elsewhere (Anderson & Schooler, 1991), such a power function produces the rapid initial loss and then slow loss, which produces the appearance of a qualitative difference between short- and long-term memory. Interference in the form of positional and acoustic confusions among list items is independent of time because this is produced by a constant noise added to the ordering produced by the match scores (Equation 6). On the other hand, as the activations of list items decrease, they become closer to nonlist items in activation and noise processes are more likely to produce intrusions from outside the list or omissions.

A strong prediction of the ACT-R theory is that within-lists positional and acoustic confusions (corrected for guessing) will not increase with time. This is not what would be predicted by interference theories that attribute increased information loss to increased confusion. A number of experiments have shown fewer such intrusions immediately than later, but no research, to our knowledge, has shown increased (corrected-for-guessing) intralist confusions beyond the first few seconds. Rather, such confusions decrease with time beyond the first few seconds (see Figure 8). ACT-R would predict that corrected-for-guessing[17] confusions would occur because there is a loss of items and so no possibility for confusing them. The immediate effect may reflect a correcting influence of an echoic memory.

Final Evaluation

Readers will draw their own conclusions about the strengths and weaknesses of this modeling effort. Nonetheless, we would like to offer our own self-evaluation. We think the strength of the effort was the relative success of a theory that addresses the moment-by-moment events occurring in the serial recall process. We were specifying events that were happening every few hundred milliseconds and were able to accommodate a wide variety of data. ACT-R's strength in this regard is displayed in its unique ability to simultaneously fit the timing and accuracy of recall (see Figure 4). We think this illustrates the strength of the ACT-R approach to modeling. ACT-R is an instance of a "hybrid" model in that it involves a symbolic component and a subsymbolic component. The symbolic component involves specifications of the rules and knowledge structures required to produce the behavior and so enables the detailed modeling of a complex task that requires a coherent sequence of steps. On the other hand, the subsymbolic activation processes in ACT-R allow us to model the continuous variation in latency and probability of recall with variations in the materials and designs of the experiments. The ACT-R model is unique among the models of span in its ability to account for the moment-by-moment temporal dynamics of the recall process.

The fact that ACT-R gets its predictive power through such a detailed analysis of the recall process makes us particularly sensitive to what we think is the most significant hole in our modeling effort. This is that we have completely omitted any model of the rehearsal processes that are taking place during presentation of the material (however, see Lovett, Reder, & Lebiere, 1996; Anderson & Lebiere, in press). We simply assume that the items are recorded at their moment of exposure and that recall will be determined by the time from presentation to the moment of recall. However, it is apparent to us that our participants were sometimes rehearsing during the study interval. There is more than enough time with the typical 1-s presentation rates for such rehearsal activities. As we noted earlier, the effects of presentation rate are weak presumably because of such intervening rehearsal activities that compensate for slower presentation rates. Our own experience in piloting such experiments is that typical articulatory suppression does not eliminate such rehearsal, although it makes it harder. Also, under articulatory suppression, we find ourselves reviewing the list in a nonverbal manner.

The difficulty with modeling these rehearsal strategies is that they are idiosyncratic to participants. Some participants will rehearse just the current item, others the previous pair, others the current group, and others a past group or the beginning of the list. There is not the same problem with strategy variation in recall because the task demands of serial output (when output is so constrained) take away most degrees of freedom. Perhaps a similar tack would be to constrain the participants rehearsal strategies. However, this has not been done in any of the published research we modeled. Thus, the success of our modeling effort has to be regarded as simply reflecting that our no-rehearsal assumptions approximated the actual situation with compensating parameter changes.

This issue illustrates the strengths and weaknesses of modeling in the ACT-R system generally. When there is a good basis for assuming a consistent processing strategy, then one can embed that processing strategy in an ACT-R simulation and obtain (hopefully) high-quality predictions. However, when this is not possible and different participants adopt different unobservable strategies, one is forced to approximate models. Note, however, that this problem is not really unique to the ACT-R model. Variability in rehearsal strategy is a fact of the data whether one is modeling that data in ACT-R or not. It is just that other theoretical approaches do not represent this level of processing detail and so do not have to acknowledge (perhaps at their peril) its potential influences on the data.

---

[17] The point of this qualification is that if participants forget and guess, they may produce what looks like more systematic confusions. Therefore, the necessary studies are ones (e.g., Bjork & Healy, 1974) that attempt to get baseline numbers to correct for such random guessing.

References

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

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. (1983a). *The architecture of cognition.* Mahwah, NJ: Erlbaum.

Anderson, J. R. (1983b). Retrieval of information from long-term memory. *Science, 220,* 25–30.

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. (1995a). *Learning and memory.* New York: Wiley.

Anderson, J. R. (1995b). *Cognitive psychology and its implications.* New York: Freeman.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Washington, DC: Winston.

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 J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 9–12). Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (in press). *The atomic components of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., Matessa, M., & Douglass, S. (1995). The ACT-R theory and visual attention. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 61–65). Hillsdale, NJ: Erlbaum.

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.

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

Baddeley, A. D. (1990). *Human memory: Theory and practice.* Boston: Allyn & Bacon.

Baddeley, A. D., Lewis, V. J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36,* 233–252.

Baddeley, A. D., Thomson, 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., & Winzenz, D. (1969). Group structure, coding, and memory for digit series. *Journal of Experimental Psychology Monographs, 80*(5, Pt. 2).

Broadbent, D. E. (1975). The magical number seven after fifteen years. In R. A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory.* New York: Wiley.

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

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.

Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *Journal of Psychology, 44,* 319–327.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics, 5,* 365–373.

Dosher, B. A. (1994, November). *Forgetting in STM recall: Presentation rate and output time.* Paper presented at the 35th Annual Meeting of the Psychonomic Society, St. Louis, MO.

Drenowski, A. (1980). Attributes and priorities in short-term recall: A new model of memory span. *Journal of Experimental Psychology: General, 109,* 208–250.

Drenowski, A., & Murdock, B. B., Jr. (1980). The role of auditory features in the memory span for words. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 319–332.

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

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.

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

Hulme, C., Maugham, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short term memory span. *Journal of Memory and Language, 30,* 685–701.

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.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Lashley, K. S. (1951). The problem of serial order in behaviour. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behaviour* (pp. 112–136). New York: Wiley.

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.

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

Lovett, M. C., Reder, L. M., & Lebiere, C. (1996, July). *Modeling individual differences in a digit working memory task.* Paper presented at the Second International Conference on Memory, Padua, Italy.

McClelland, J. R., & Rumelhart, D. E. (1981). An interactive model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88,* 375–407.

McNicol, D. (1978). Initial codes for order in memory. *Quarterly Journal of Experimental Psychology, 30,* 681–691.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Morra, S., Tressoldi, P. E., Mazzoni, G., Sava, D., & Zucco, G. (1991). *The capacity of the articulatory loop is not limited by time rather than by the amount of information or number of events.* Unpublished manuscript.

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

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78,* 679–684.

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

Posner, M. I., & Konick, A. W. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology, 58,* 193–198.

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

Reder, L. M., & Gordon, J. S. (1997). Subliminal perception: Nothing special cognitively speaking. In J. Cohen & J. Schooler (Eds.), *Cognitive and neuropsychological approaches to the study of consciousness* (pp. 125–134). Hillsdale, NJ: Erlbaum.

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of groups. *Quarterly Journal of Experimental Psychology, 21,* 137–147.

Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin, 99,* 303–319.

Schweikert, R., & Boroff, B. (1986). Short-term memory capacity:

Magic number or magic spell? *Journal of Experiment Psychology: Learning, Memory, and Cognition, 12,* 419–425.

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.

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* (pp. 117–152). New York: Academic Press.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition, 5,* 529–534.

Wickelgren, W. A. (1964). Size of rehearsal group and short-term memory. *Journal of Experimental Psychology, 68,* 413–419.

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 nonrepeated items. *Quarterly Journal of Experimental Psychology, 17,* 14–25.

Wickelgren, W. A. (1967). Rehearsal grouping and the hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology, 19,* 97–102.

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 19, 1995
Revision received December 16, 1996
Accepted December 27, 1996 ■

Sternberg Appointed Editor of *Contemporary Psychology (APA Review of Books)*, 1999–2004

The Publications and Communications (P&C) Board of the American Psychological Association announces the appointment of Robert J. Sternberg, Yale University, as editor of *Contemporary Psychology (APA Review of Books)* for a 6-year term beginning in 1999.

Contemporary Psychology has been in existence for 42 years and, for most of the time, has been operating under the same coverage model. The model is a good one, as the current issues edited by John H. Harvey reflect, and the journal has long met the needs of individuals and libraries. The pace of change has increased during the past few years, however, and the P&C Board recently decided that it was time for a new model, one that would reflect the 21st century reader's needs for information about books.

Sternberg, at the request of the P&C Board, will be embarking on a program to make the journal even more timely and interesting during his editor-elect year in 1998. Some of the changes envisioned include fewer but longer and more thoughtful reviews of books, reviews only of "new" books (with a few noteworthy exceptions), comparative textbook reviews at strategic times of the year, and changes in publication frequency and pricing. Sternberg welcomes suggestions for improving the journal and serving reader needs.

E-mail: robert.sternberg@yale.edu

Please note that all reviews are written by invitation. Publishers should note that books should not be sent to Sternberg. *Publishers should continue to send two copies of books to be considered for review plus any notices of publication to*

PsycINFO Services Department, APA
Attn: *Contemporary Psychology* Processing
750 First Street, NE
Washington, DC 20002-4242

As the editorial term of John H. Harvey comes to a close, the P&C Board wishes to express its appreciation for his hard work and dedication as well as that of his staff at the University of Iowa.