A Theory of Sentence Memory as Part of A General Theory of Memory

John R. Anderson, Raluca Budiu, and Lynne M. Reder

Carnegie Mellon University

We describe an ACT-R model for sentence memory that extracts both a parsed surface representation and a propositional representation. In addition, if possible for each sentence, pointers are added to a long-term memory referent which reflects past experience with the situation described in the sentence. This system accounts for basic results in sentence memory without assuming different retention functions for surface, propositional, or situational information. There is better retention for gist than for surface information because of the greater complexity of the surface representation and because of the greater practice of the referent for the sentence. This model's only inference during sentence comprehension is to insert a pointer to an existing referent. Nonetheless, by this means it is capable of modeling many effects attributed to inferential processing. The ACT-R architecture also provides a mechanism for mixing the various memory strategies that participants bring to bear in these experiments. © 2001 Academic Press

Key Words: sentence memory; ACT-R theory; surface information; propositional information; situational information; inferential processing.

In his 1998 book, Kintsch writes: "We don't need a special theory of sentence memory: If we understand sentence comprehension (the CI theory) and recognition memory (the list-learning literature), we have all the parts we need for a sentence recognition model" (p. 263). CI is Kintsch's construction-integration theory (Kintsch, 1988, 1998) and he adopts Gillund and Shiffrin's (1984) SAM model of memory to account for sentence memory. In this article we argue for a conclusion that has a similar spirit—which is that the established results on sentence memory also follow from the ACT-R cognitive architecture (Anderson & Lebiere, 1998). ACT-R bears similarity to SAM but is a more complete theory of cognition because it contains a model of cognitive control. As such we can directly embed in it a theory of sentence comprehension. Because of some of the architectural commitments of ACT-R, the theory of sentence comprehension is somewhat different than

Kintsch's and closer to what is characterized as the minimalist hypothesis of sentence processing (McKoon & Ratcliff, 1992, 1995).

This article demonstrates, even more strongly than has Kintsch, that there is nothing special about sentence memory. An important novel conclusion from this theory is that there are not different retention functions for the three forms of memory that have been postulated to encode information about a sentence (e.g., Fletcher, 1994; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998)—surface code (exact words and syntax), textbase (propositions asserted in the text), and situation model (inferences contributed from long-term memory). A single retention function contrasts with a frequent assumption (e.g., Anderson, 1974, 2000; Brainerd & Reyna, 1995; Kintsch, Welsch, Schmalhofer, & Zimny, 1990) that the superficial surface information is more rapidly forgotten than the propositional information, which is in turn forgotten more rapidly than the situation information. However, we do not challenge the concept of the three levels of representation—although in keeping with ACT-R's minimalist leanings, we offer a somewhat Spartan interpretation of what the situation information amounts to.

In this article we present ACT-R models for a number of sentence memory tasks that empha-

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

size different subsets of these three representations. In each case we present models that actually perform in real time the tasks described in the literature. These models can be run and inspected by going to the *Published Models* link at http://act.psy.cmu.edu. The real-time nature of these models is significant because constraints on processing time force the models in the direction of minimalist encoding. In ACT-R each production rule applies serially and requires a minimum of 50 ms and often more. When we apply ACT-R to sentence processing we find there is just not enough time, at normal reading or listening rates, to do more than a minimal number of inferences.

We chose to model data sets that would directly test two critical aspects of the ACT-R theory—its retention assumptions and its assumptions about the speed of production rules. Some of the data sets (Anderson, 1972, 1974; Reder, 1982; Schustack & Anderson, 1979) that we model are ones gathered from our own laboratories and in these cases the models that we describe are ACT-R implementations of what are essentially the models that we already proposed, prior to the development of ACT-R. In these cases we show that the earlier proposed models are consistent with the general ACT-R architecture. We also model other researchers' data sets (Bower, Black, & Turner, 1979; Zimny, 1987). Although we do not know these data sets as well as our own, they were chosen because they serve to test significant aspects of the theory. This article begins with a description of the ACT-R architecture, a minimal model for sentence processing and representation, and the underlying architectural assumptions that control the behavior of the model.

THE ACT-R THEORY

General Architectural Commitments

The basic assumption throughout the development of the ACT theory (e.g., Anderson, 1976, 1983, 1993; Anderson & Lebiere, 1998) has been that human cognition emerges through an interaction between a procedural memory and a declarative memory. The basic units of knowledge in procedural memory are productions and the basic units of knowledge in declar-

ative memory are chunks. Since we want to make the point that the ACT-R assumptions we are using for sentence memory apply generally throughout cognition, we first illustrate them with respect to mathematics. For instance, consider a student in the midst of solving the following multicolumn addition problem:

$$\begin{array}{r} 336 \\ +848 \\ \hline 4 \end{array}$$

The next production to apply might be:

> IF the goal is to add n1 and n2 in a column
>   and n3 can be retrieved as the sum of n1
>   and n2
> THEN set as a subgoal to write n3 in that
>   column.

This production would retrieve the following chunk from declarative memory encoding the fact that the sum of 3 and 4 is 7:

```
fact
    isa          addition fact
    addend1      three
    addend2      four
    sum          seven
```

and embellish the goal with the information that 7 is the number that should be written out. Then other productions would apply that might deal with things like processing the carry into the column. The basic premise of the ACT-R theory is that cognition unfolds as a sequence of such production-rule firings where each rule can retrieve chunks from declarative memory to transform the goal state.

One of the major trends in the ACT theory development from ACT* (Anderson, 1983) to the current ACT-R (Anderson & Lebiere, 1998) has been a firmer commitment to the temporal grain size at which cognition unfolds. Each production rule in ACT-R takes at least 50 ms to fire and almost never much more than 500 ms. Thus, we have bounded the time scale to an order of magnitude and we will shortly describe the factors that determine just how long a production rule takes in the 50- to 500-ms range. The ACT-R theory is also committed to the proposal that only one production rule can fire at a time.

These commitments to a temporal grain size and serial production-rule firing place severe constraints on a theory of linguistic processing because ACT-R must complete all the steps needed to comprehend a sentence in the short time typically allocated to sentence processing.

Representational Commitments

Another significant constraint on the proposed theory of language processing is that it must incorporate the theory of declarative representation that was articulated in the theory of list memory (Anderson et al., 1998) and was elaborated in the theory of analogy (Salvucci & Anderson, 2001). In the theory of serial memory, declarative chunks are used to encode the position of an element in a higher structure. Thus, a sequence like "392 714 856" would be encoded in the hierarchical graph structure depicted in Fig. 1. Each node and link in this figure is a chunk. While the nodes contain no structural information (e.g., the leaf *3* in the graph is a chunk that encodes the digit 3, with no information about it being part of this list), the links are more complex (for simplicity, in Fig. 1 we only show the structure of two link chunks; the other links are similar, though). As Fig. 1 shows, together with pointers to the nodes that they connect (the *parent* and *child* slots), the link chunks maintain information about the position of the child within the parent group. For instance, *9*, the child of *Group1*, occupies the second position in *Group1* and this information is recorded in the slot *role* of the link that connects *9* and *Group1*. Also, in order to be able to keep track of different lists, it was important to have a *context* slot in each link chunk and in this way identify to which list a given link should be associated. Individual declarative chunks in ACT-R can be forgotten or confused with others, and these chunk-based processes produce many of the error patterns associated with serial memory (Anderson & Matessa, 1997).

Salvucci and Anderson (2001) elaborated and generalized this representation to account for the semantic effects found in the analogy literature. Thus, to model the famous solar system analogy (Gentner, 1983), they represented arguments to a proposition like "The planets revolve around the Sun" with a number of chunks like:

Chunk82
   isa   semantic-chunk
   parent   revolves
   child   Sun
   role   center
   referent   revolution
   context   solar-system.

This chunk encodes the fact that the *Sun* serves the *center role* in that proposition.[1] Another chunk would be used to encode that the planets serve the role of revolving objects. That is, there is a separate link chunk for each argument of the proposition. Note also that Salvucci and Anderson added a new piece of information to the link chunk: the *referent* slot which points to the more general concept of the motion of *revolution*. Sometimes it may be useful to think of the referent as the prototype of the particular instance that is represented. In Salvucci and Anderson's model, the referent served to guide the analogy process. It can be also used to guide metaphor comprehension and other semantic interpretation processes (see Budiu, in preparation). The chunks we use to encode propositional information in sentences are basically identical to the chunks introduced by Salvucci and Anderson. The referent link is important to our theory of situation memory.

It is worth noting that this representation takes what commonly had been thought of as a single proposition (e.g., "the planet revolves around the Sun") or a single group "(3 2 9)" and fragments it into multiple ACT-R chunks. This fragmentation proved useful in list memory to account for phenomena such as transposition errors. It also proved useful in the theory of analogy to explain how a participant analyzes the components of an analogical mapping. In the case of sentence memory, this assumption has implications for fragmentary sentence recall and we test these implications in this article.

Representation of Sentential Information

We propose a representation for the syntactic structure of a sentence (Fig. 2a) similar to the

---

[1] The actual names used to refer to the slots have been changed from those used by Salvucci and Anderson to facilitate current exposition.
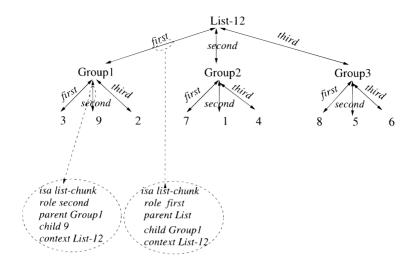
FIG. 1. The encoding of a serial list into a set of chunks from Anderson, Bothell, Lebiere, Matessa (1998). Each link and node in the graph reflects a chunk.
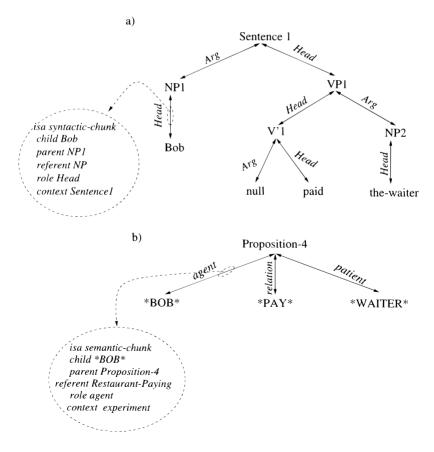

FIG. 2. A comparison of the syntactic encoding of a sentence (a) with its propositional encoding (b).

list representation (Fig. 1) and a representation for its propositional structure (Fig. 2b) that is basically identical to the semantic representation developed in the Saluvcci and Anderson (2001) model. Thus, for the sentence *Bob paid the waiter*, the syntactic representation is an encoding of the actual parse tree of the sentence: The nodes in this tree are either words like *Bob*, *paid*, and *the-waiter* or nonterminals like *NP1*, *VP1*, *V′1*, *NP2*, and *Sentence1*. The *null* element in the verb phrase encodes potential verb auxiliaries. As before, the links are more complex chunks containing structural information. The labels of the links represent the syntactic roles that the children play within the parents (for instance, *Bob* is the head of *NP1*, which is the first argument of *Sentence1*). As in the solar system representation, the link chunks also encode a referent, whose value denotes a more general concept (e.g., the link connecting *NP1* and *Bob* has the referent *NP* to denote that it is an instance of a noun phrase structure). The *context* slot in the link representation keeps track of the current sentence.

Similarly, the semantic structure of the sentence is encoded as a tree whose nodes are concepts or propositions and whose links represent relationships among these concepts (see Fig. 2b). Thus, the link between the concept *BOB* and the chunk *Proposition-4* encodes the fact that *BOB* is the agent of *Proposition-4*. The referent slot records that the relationship encoded is an instance of paying in a restaurantlike script. All the links in the representation of this proposition can have the referent slot pointing to this referent. In general, the referent slot is filled with a pointer to some analogous past experience or generalization from past experiences. Note that our "semantic representation" in Fig. 2b might better be termed a "gist representation." It collapses, for instance, any semantic distinction between an active or passive sentence. Its essential feature is that it reduces the detail of the sentence down to its core meaning.

Again, because the links contain all the structural information, their retrieval will be critical for sentence recall. Note that there are more chunks (in terms of both nodes and links, but links will be our primary interest) in the syntactic encoding (8 links) than in the propositional encoding (3 links). The discrepancy is even greater in the case of the passive sentence *The waiter was paid by Bob*, where the syntactic encoding has 10 links, while the propositional still has only 3. This greater difference in the number of chunks accounts for the apparent superior memory for propositional information because fewer things have to be retrieved to reconstruct the proposition than the syntax. The exact surface structure in Fig. 2a and the exact propositional structure in Fig. 2b depend on representational assumptions that might be questioned but the general principle is that the gist representation will be a smaller representation encoding only significant aspects of the original sentence. Thus, the model is committed to the prediction of poorer memory for surface structure, not because of worse retention of the individual chunks, but because there are more chunks. The more chunks there are, the more likely it is that something will be lost with delay. Ability to recognize the exact sentence depends on all of the elements being present in the surface representation. While the model predicts better memory for the meaning, it is not inconsistent with the observation that surface memory can be improved by manipulations that focus attention on surface details (e.g., Kennan, MacWhinney, & Mayhew, 1977; Murphy & Shapiro, 1994). ACT-R predicts that memory for any chunk, syntactic or semantic, will be enhanced by greater processing. However, the theory does predict inferior surface memory in the absence of special processing.

Figure 3 is an attempt to illustrate the larger structure that is created when story sentences get attached to referents, in this case propositions from a restaurant script. The big boxes, labeled "Story" and "Restaurant," represent the organizing units that are pointed to by the context slots of the individual chunks encoding the links that make up the two sets of propositions. The smaller boxes reflect the individual propositions that are pointed to by the parent slots. The elements within the proposition boxes are pointed to by the child slots. The arrows reflect referent slots pointing from the chunks to the referent proposition. This representation illustrates that
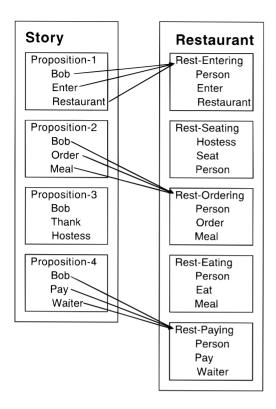
Story

Proposition-1
Bob
Enter
Restaurant

Proposition-2
Bob
Order
Meal

Proposition-3
Bob
Thank
Hostess

Proposition-4
Bob
Pay
Waiter

Restaurant

Rest-Entering
Person
Enter
Restaurant

Rest-Seating
Hostess
Seat
Person

Rest-Ordering
Person
Order
Meal

Rest-Eating
Person
Eat
Meal

Rest-Paying
Person
Pay
Waiter

FIG. 3. A representation of the chunks in a story and their connections to the propositions in a referent.

the participant might not be able to find referents for all the propositions in the story and that there might not be story propositions corresponding to all the propositions in the referent. While the referent propositions in this example come from a classic Schank and Abelson (1977) script, there is nothing in the model that requires this. The referents could come from another story, for instance. The sources of the referent just need to be some well-encoded structure in declarative memory that contains propositions that can be put in correspondence with the propositions in the story. Our concept of a referent is similar to Sanford and Garrod's (1998) scenario and our use of the referents is similar to their scenario mapping except that they do not build up a separate propositional representation.

The representation in Fig. 3 illustrates some of the potential for inferences based on these referent links to prior knowledge. Suppose a participant can retrieve just one chunk from a

story proposition (say the one for *order* in Proposition-2) but this has a referent link. Then the participant can use this referent link to retrieve the corresponding proposition. Furthermore, the participant can use the arguments in the referent proposition to infer the arguments in the story proposition (for instance, that a meal was ordered). Being more adventuresome, participants might also guess that other propositions in the script occurred in the story even if these propositions are not pointed to.

Sentence Processing

We now turn to describing productions that perform three tasks during sentence processing: deriving a parse of the sentence, building a propositional representation, and trying to identify a referent for the proposition. This model makes almost no effort at elaboration, i.e., embellishment of the ideas in the sentence. The reason for this is that the model is constrained to fit the data from experiments where participants are reading stories at the rate of at least a couple of words per second. This implies no more than a few hundred milliseconds to process each word and therefore constrains what can be accomplished in that time. The one bit of embellishment that the model will do is try to find a referent for the sentence. Of course, when participants are given more time to study they often engage in extensive inference and elaboration. Indeed, we have argued elsewhere (Anderson & Reder, 1979; Reder, 1979) that such elaboration can have significant consequences for their memory of the sentences. However, it turns out that we do not need to make such assumptions in order to account for a number of classic results about inference in sentence memory. Rather, they can be explained simply by the use of referents.

The parsing model we use is essentially a scaled-down version of the ACT-R model developed by Lewis (1999) for simulating comprehension effects. It assumes that, with each word processed, the participant retrieves the syntactic category of the word and uses that knowledge to integrate the word into a syntactic parse of the sentence. Lewis's work is more concerned with sentence complexity and garden-path effects

than are we, and he models these effects by retrieval of declarative fragments of the parse tree. We assume the participant is only parsing simple sentences without significant ambiguities or syntactic complexities. Our model builds up a propositional representation as it builds up the parse tree. When the propositional representation is complete, it will attempt to retrieve the referent. Elsewhere (Budiu & Anderson, 2000) we have argued that in at least some situations participants are also retrieving a referent for the sentence before they finish reading it. As a simplification, we postpone retrieval of a referent until the end of the sentence, but it is not essential to the model.

Figure 4 shows how the propositional and semantic representation is built when the ACT-R model processes the active sentence *Bob paid the-waiter*. The noun phrases are hyphenated to represent the assumption that the determiner–noun combination is processed as one encoding. This is roughly consistent with eye movement data (Just & Carpenter, 1987) and serves to eliminate any differences between processing of phrases like *the-waiter* and *Bob*.

For each word, there is a cycle of three productions which fire: *Read-word*, taking 100 ms to encode the current word; *Retrieve-Type*, taking about 50 ms to retrieve the syntactic category of the word; and a variable third production that actually uses this information to appropriately augment the syntactic and semantic structures. To illustrate, at the beginning of the sentence, after reading the word *Bob* and retrieving the fact that *Bob* is a noun, the model builds up the parts of the syntactic tree and of the semantic representation corresponding to *Bob*. For the syntactic tree, the model creates new nodes (*NP1* and *Sentence1*) to denote that it is dealing with a new sentence and a new noun phrase and also new links to relate these nodes (namely, a link which encodes that *Bob* is the head of the new noun phrase *NP1* and a link which records that *NP1* is the first argument of the sentence *Sentence1*). For the semantic representation, the model builds a new node (*Proposition-4*) corresponding to the new proposition, and then it creates a link between *Proposition-4*
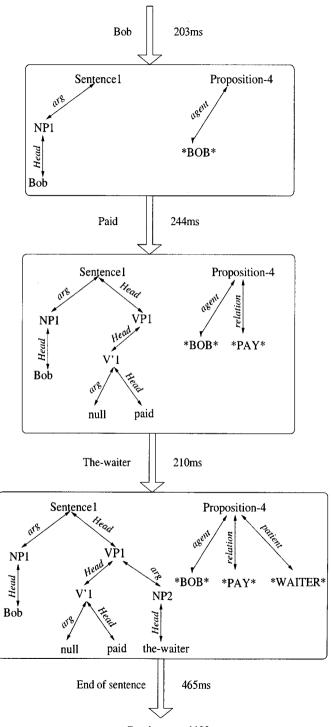
and the meaning of *Bob* (denoted *BOB* in the figure). The model is biased to believe that initial nouns are agents, so this link is labeled *agent*. The context slot of this link is filled with the value *experiment*, and the referent link is left unset to reflect the fact that we postpone the retrieval of a referent until the end of the sentence. The process repeats for each new word, with the category of the word and the state of the trees influencing which productions fire. When the end of sentence is reached, the model looks for a long-term memory referent which has a semantic structure similar to the semantic structure it has just built. The relatively long latency (465 ms) at the end of the sentence reflects the time for separate productions to set up the retrieval, retrieve the referent, and modify the semantic chunks with the referent.

Figure 5 shows how the model comprehends the passive sentence *The-waiter was paid by Bob*. The process is very similar to the one for the active sentence: At first, the model considers the initial noun *The-waiter* as an agent. Only after it recognizes that the auxiliary plus the verb make the sentence a passive does it update the representation to reflect that the concept *WAITER* is a patient. To perform the update, the model takes a little more time because it needs to retrieve the link between *Proposition1* and *WAITER* in order to be able to change the old *agent* label to a *patient* one. As before, the processing of the sentence ends both with the retrieval of a referent and with the updating of the links in the semantic representation so that they point to the retrieved referent.

The traces in Figs. 4 and 5 display the time taken by the productions. We now present the equations that determined these timings.

ACT-R's Subsymbolic Assumptions

To this point we have largely described ACT-R as a symbolic theory in which discrete productions are fired and discrete chunks are retrieved. However, underlying ACT-R is a subsymbolic layer of continuously varying quantities that determine which productions and chunks are selected, if any, and the latency for each chunk's retrieval. Processing at the subsymbolic level is controlled by quantities called

FIG. 4. Time frames in the parsing of the active sentence, "Bob paid the waiter."
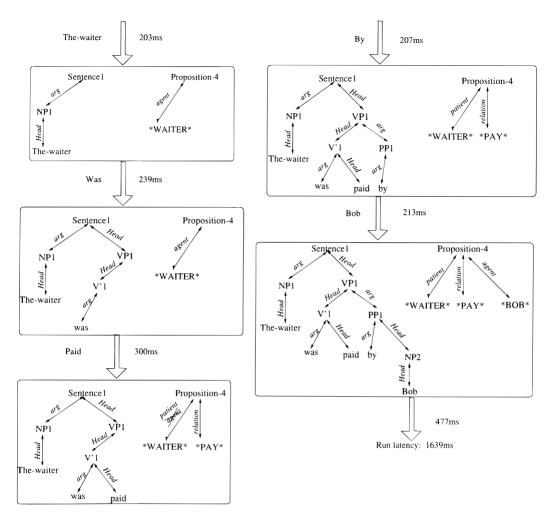
FIG. 5. Time frames in the parsing of the passive sentence, "The waiter was paid by Bob."

activations in the case of declarative memory and utilities in the case of procedural memory. Also, while the computation at the symbolic level is serial, the computation at the subsymbolic level is parallel. Underlying the firing of a single production is a large amount of parallel activation computation and parallel utility computation.

The activation of a chunk is determined by its base level and its associations to elements in the current context. The following equation describes the level of activation, A_i, of a chunk i in terms of its base-level activation, B_i, that reflects its past history of encodings (as defined below) as well as the strengths of association, S_{ji}, to el-

ements j in the goal that send it additional activation:

$$A_i = B_i + \sum_j W_j S_{ji}. \qquad \text{Activation Eq. (1)}$$

The base-level activation varies with the frequency and recency of use according to the following equation:

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right), \qquad \text{Base-Level Learning Eq. (2)}$$

where t_j is the time since the jth use of the chunk and d is a parameter controlling activation decay. As developed in Anderson (1982) and ex-

tensively tested in Anderson, Fincham, and Douglass (1999), this equation both predicts the power law of learning (Newell & Rosenbloom, 1981) and the power law of forgetting (Wickelgren, 1972). For current purposes, the summation in this equation implies that the more a chunk is used, the stronger will be its encoding. The decay function t_j^{-d} implies that the base-level activation will decay with time. Elsewhere (e.g., Anderson & Lebiere, 1998; Anderson & Reder, 1999) we have elaborated a theory of strength of associative activation [the $\sum W_j S_{ji}$ in Activation Eq. (1)], relating it to things like the fan effect; however, for current purposes it is enough to assume that this produces a boost for elements associated to the goal. The base-level learning equation above is at the heart of the applications reported in this article that are concerned with the retention of a sentence over various delays. We model data assuming there is one decay constant d for both syntactic and propositional information about the sentence. Furthermore, taking the strong commitment from other ACT-R models (Anderson & Lebiere, 1998) we have fixed this decay constant at .5. This is one instantiation of our claim that all levels of information about the sentence have the same memory properties.

The activations are noisy quantities and fluctuate around their expected values. A chunk can be retrieved if its activation value is above a threshold τ. The probability of retrieving a chunk with expected activation A is given by the following equation:

$$\text{Probability} = \frac{1}{1 + e^{-(A-\tau)/s}}, \quad \text{Retrieval Probability Eq. (3)}$$

where s reflects the noise in the activation values and is related to the variance, σ, of the noise by the equation $s = \sqrt{3}\sigma/\pi$. The activation, A, of a chunk is also related to the time to retrieve it by the following equation:

$$\text{Time} = Fe^{-A}, \quad \text{Retrieval Time Eq. (4)}$$

where F is the latency scale factor.

The preceding equations describe the subsymbolic part of ACT-R's declarative memory. The procedural memory also has subsymbolic aspects. When there are a set of productions that can apply, ACT-R chooses among them ac-

cording to how well they have performed in the past. The measure of production performance is called utility. There is one such quantity associated with each production and it is calculated as $PG - C$, where P is the probability with which the production has led to a successful completion in past attempts, C is the average amount of time that it took to reach completion, and G is the value of successfully achieving the goal. The parameters P and C are based on past experience[2] with the production while G is a parameter to be estimated. ACT-R selects the production with the highest utility value, but because of noise in these utilities, there is only a probability that any production i will be selected and this is given by the following equation:

$$\text{Probability of choosing } i = \frac{e^{E_i/t}}{\sum_j e^{E_j/t}}, \quad \text{Conflict Resolution Eq. (5)}$$

where the summation in the denominator is over the productions, j, that currently match the goal. This is a softmax rule which tends to select the best production. The parameter t reflects the noise in the estimation of production utility and is related to the variance, σ, of this noise by the equation $t = \sqrt{6}\sigma/\pi$. The units of utility are seconds and throughout this article we use a constant estimate for t of .05 s. One theme in a number of the models that we describe is that there are multiple strategies for answering questions about sentences and that participants choose among these strategies according to their experienced utilities.

Summary

We have now described the basics of the ACT-R theory and the general representation and processing assumptions. We have also described a model for sentence processing within the theory. The important assumptions for purposes of testing the theory are (a) a minimal processing of the sentence which derives a parse tree, a propositional representation, and

---

[2] In the simulations P is set to the actual probability of success in the simulation and C to the actual processing time it took the simulation.

a referent if one can be found and (b) the same retention function for all information. We have yet to describe how the model deals with the memory tests, as this depends on the specifics of the particular experiment's testing procedure. However, data from the experiments will be modeled assuming either a direct effort to retrieve information from the sentence encoding or an effort to use the referent, if there is one, to infer an answer for the memory task.

THE EXPERIMENT MODELS

Table 1 lists the experiments that are modeled in this article and the parameter estimates for these experiments. We start with a model for an experiment described in Anderson (1974) that is concerned with the processing of surface and propositional information. Next, we discuss an experiment by Anderson (1972) that addresses the issue of whether a single proposition is really fragmented into a number of separate chunks as assumed by the ACT-R model. This is the only model that looks at sentence recall measures rather than sentence recognition measures. In our model for the data from this experiment we make extensive use of situational referents. We also make extensive use of situational referents to model plausibility and recognition judgements in Reder (1982). That experiment was primarily concerned with latency measures. We adapt that model to account for a similar experiment by Zimny (1987), which is concerned with probability of recognizing sentences. The Reder model is also adapted to account for data from Schustack and Anderson (1979) showing that sometimes situational referents can result in increased ability to recognize studied sentences. This model is in turn adapted to account for results from Bower,

TABLE 1

The Experimental Models and Parameter Estimates

	Anderson (1974)	Anderson (1972)	Reder (1982)	Zimny (1987)	Schustack & Anderson (1979)	Bower, Black, & Turner (1979)
Latency Scale(F)	0.30 s	As Anderson (1974)	As Anderson (1974)	As Anderson (1974)	As Anderson (1974)	As Anderson (1974)
Time to Read a Word	0.10 s	As Anderson (1974)	As Anderson (1974)	As Anderson (1974)	As Anderson (1974)	As Anderson (1974)
Intercept	0.65 s	Not used	0.85 s	As Reder (1982)	As Reder (1982)	As Reder (1982)
Utility Noise(t)	0.05 s	Not used	As Anderson (1974)	As Anderson (1974)	Not used	Not used
Activation Noise(s)	Not used	0.2	As Anderson (1972)	As Anderson (1972)	As Anderson (1972)	As Anderson (1972)
Ret Threshold(τ)	Not used	0.3	As Anderson (1972)	As Anderson (1972)	−0.05	As Schustack & Anderson (1979)
Slip Probability	Not used	Not used	0.12	0.24	0.125	As Schustack & Anderson (1979)
Goal Value(G)	Not used	Not used	34	10.5	Not used	Not used
Guess Latency	Not used	Not used	0.80 s	As Reder (1982)	As Reder (1982)	As Reder (1982)
Model Unique		p(ref) = .20, .39 $A = 0.25$	p(Plausible) = .90		p(guess) = .06	Plaus rated 3.5 Seen rated 6.0
Number of experiment-specific parameters	4	5	5	2	3	3
Latency R^2	.991	—	.954	—	—	—
Accuracy R^2	—	.992	.859	.923	.999	.995

Black, and Turner (1979) showing that situational referents can sometimes result in poorer discrimination of target sentences. At the end of this article we return to the issue of the stability of the parameter estimates. All of these models are available by following the "Published Models" link from the ACT-R home page (http://act.psy.cmu.edu). The interested reader may inspect the details of these models, observe them run, and check their behavior with other parameter settings.

Anderson (1974): Surface versus Propositional Representations

Anderson (1974) reported an experiment in which participants studied sentences either in the active voice or passive voice and then had to judge whether active or passive test probes were implied by these sentences. The foils switched the roles of the agent and object. Thus, the original study sentence might be either *The-sailor shot the-painter* or *The-painter was shot by the-sailor* and the participants would later be asked to judge whether a test probe followed from the studied sentence. For either of the sentences the true sentence would be either that sentence or the other form. For either of the sentences foil sentences could be either active or passive as in *The-painter shot the-sailor* or *The-sailor was shot by-the-painter*.

Thus, the trials could be classified by the voice of the study sentence (active or passive), the voice of the probe sentence (active or passive), and whether the probe sentence was a target (true) or a foil. Participants were tested either immediately after reading the study sentence or at a 2-min delay. Figure 6 displays the results from these two conditions. The positive judgments in the immediate condition show a strong interaction between the voice of the studied sentence and the probe sentence, with participants much faster for targets for which the voices match. The data at a delay are quite different and show a large effect of the voice of the test sentence with participants taking longer for passives.

At the time this experiment was published, these data were taken as evidence for more rapid forgetting of the surface form of the sentence than of the propositional form. The analysis was basically as follows: Immediately, participants had access to a surface trace and made their judgements on the basis of that, producing a rapid response when there was an exact match of form. This surface trace decayed with delay and the participant was left with the propositional trace that did not encode the voice of the studied sentence. There was a large effect of the voice of the probe sentence at delay because participants had to comprehend the sentence to match propositional traces and passives take longer to comprehend (compare Figs. 4 and 5). The ACT-R model fit to the data in Fig. 6 largely reproduces the account in Anderson (1974) but it does not assume a differential forgetting of the two traces. Still it does a good job of fitting the effect of delay because of the differential complexity of surface and proposition traces (see Fig. 2).

Figure 7 is a schematic representation of the model we implemented, which is essentially the model described in the original Anderson (1974) article. Figure 7 also gives the range of times for each step which vary with delay and voice of the sentences. The actual ACT-R model can be accessed at the "Published Models" link at the ACT-R website. Here we just review its basic logic. The model chooses between a verbatim and propositional strategy. If it chooses the verbatim strategy it never parses the probe sentence but rather immediately retrieves a surface trace from memory that contains the first noun phrase of the probe sentence. Then it checks to see whether the retrieved sentence and the probe sentence match on first noun phrase, verb auxiliary, and verb. As in Anderson (1974), it is assumed that the participant never reads the second noun phrase, as all probes in the experiment can be judged without the second noun. In fact, the model in Fig. 7 only checks for verb auxiliary and does not read the main verb if there is an auxiliary. The model starts out with a response index set to yes and switches it should the subjects mismatch or the verb auxiliaries mismatch. When judging a passive transformation of an active studied sentence or vice versa, both subject and verb auxiliary will mismatch and the response index will be switched twice
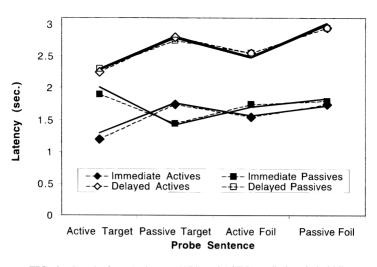
FIG. 6. Results from Anderson (1974) and ACT-R predictions in bold lines.

from yes to no and back to yes. Such sentences take longer to judge, not because of this response switching per se, but because of the more complex processes of retrieving the target sentence. The noun used to retrieve the sentence in step 2 will be the first noun in the probe but the second noun in the retrieved sentence. When the participant has to retrieve the subject of the memorized sentence in step 3 this will be different than the noun retrieved in step 2 and so there is not a benefit of a recent retrieval.[3]

If the participants adopt the propositional strategy they must first comprehend the probe sentence and this comprehension will show a large effect of whether the sentence is active or passive. Having done this, the probe proposition can be more economically matched to the memory representation. In all, four chunks must be retrieved from the propositional representation to complete the matching—one to first retrieve the proposition and three to match the agent, verb, and object (these are the chunks encoding links in Fig. 2). In contrast, seven to nine chunks

need to be retrieved from the verbatim representation—two to four to retrieve the sentence (depending on whether the studied sentence was active or passive) and five to match the subject and verb auxiliary. This reflects the differential complexity of the surface versus propositional representations in Fig. 2. For every chunk in the propositional representation there are two or more chunks that need to be retrieved in the verbatim representation. Moreover, there are fewer cues for retrieving the chunk in the case of the verbatim representation. In checking that the elements of the retrieved proposition match the probe proposition, each chunk can be cued with both the retrieved proposition and the concept (e.g., *Proposition-4* and *Waiter* in Fig. 2b). In contrast, there is only one cue available for each retrieval in checking the verbatim representation because of the extra intervening layer of syntactic phrase structure (NP1, VP1, V′1, and NP2 in Fig. 2a). In summary, there are fewer retrievals in the case of the propositional representation and more sources of activation [j's in the Activation Eq. (1)] to guide these retrievals. Therefore, participants are faster at retrieving the propositional structure.

Table 2 summarizes the comparison of the verbatim and propositional strategies when run through the simulation described above. The propositional strategy requires an initial parsing

---

[3] For example, when using the verbatim strategy, if the probe sentence is "The-sailor shot the-painter," the model looks for any surface representation involving the-sailor (step 2 in Fig. 7). If the studied sentence "The-painter was shot by the-sailor" is retrieved, the-painter will have to be retrieved from this sentence to compare to the-sailor (step 3 in Fig. 7).
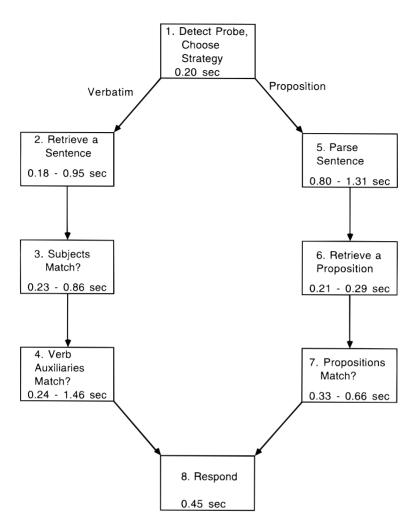
FIG. 7.   The model derived from Anderson (1974) which describes the processing of the sentences.

but places less demand on memory. The initial parsing takes .80 s for actives and 1.31 s for passives for an average of 1.05 s. This parsing time does not vary with delay but the matching time does because it involves retrieving more or less active studied information from memory. In the immediate condition, the matching takes an average of 0.67 s. In the delay condition the matching takes an average of 0.94 s. Thus, the effect of delay for the propositional strategy is to increase the retrieval time by 0.27 s. In addition to the parsing and retrieval times there is an "intercept time," which is the time to initially detect the probe and generate a response and is

estimated as 0.65 s. These intercept times also apply to the verbatim strategy. The verbatim strategy avoids the 1.05-s parsing cost but has a greater matching cost. The matching costs are 1.01 s in the immediate condition and 2.25 s in the delayed condition.

Putting the component times together (intercept, matching, parsing), the model predicts 1.66 s for the verbatim strategy versus 2.37 s for the propositional strategy in the immediate condition and 2.90 s versus 2.64 s in the delayed condition. These times influence choice between the two strategies through the Conflict Resolution Eq. (5) given above. The different

TABLE 2

Analysis of Strategy Selection in Anderson (1974)

	Immediate (seconds)	Delayed (seconds)
Verbatim Strategy		
Matching Time	1.01	2.25
Intercept	0.65	0.65
Total	1.66	2.90
Propositional Strategy		
Parsing	1.05	1.05
Matching Time	0.67	0.94
Intercept	0.65	0.65
Total	2.37	2.64
Difference	−0.71	0.26
Probability of Verbatim	100%	0%

costs in time result in completely different tendencies to select the verbatim strategy—100% in the immediate condition and 0% in the delayed condition. The reader can confirm these percentages by substituting these times (negatively weighted) into the Conflict Resolution Eq. (5) and using the value of $t = .05$ s, which is the noise estimate throughout this article.

In addition to the t parameter, the other parameters estimated for this experiment were as follows: intercept time = 0.65 s, F parameter in the latency time equation = 0.30 s, and time to read a word = 0.10 s.

Thus, in total there are four parameters and, except for the intercept, they are held constant throughout the article. The intercept and word-reading times are reasonable in absolute terms. The F parameter and the expected-gain noise t are both in the ballpark of other estimates in ACT-R modeling (e.g., Anderson & Lebiere, 1998). The overall correlation between theory and data is .996, which compares to the correlation of .976 reported by Anderson (1974) for a model with more parameters.

The good fit of this model derives in large part from the good fit of the model in Anderson (1974), since Fig. 7 is adapted from that article. The substantial parameter reduction reflects the fact that ACT-R was able to unify many things which the other model had to estimate separately such as probabilities of verbatim strategy in various conditions and changes in processing

time with delay. The slightly better fit of the model reflects the fact that this unification captured some subtle trends in the data that were ignored in the original model. The two key elements to the unification that ACT-R provides are the theory of activation decay built into the Base-Level Learning Eq. (2) and the theory of strategy selection built into Conflict Resolution Eq. (5), which determined which branch was followed in Fig. 7.

The basic insight is that the difference between results from using the verbatim and propositional representations is not a consequence of inherent differences in their retention properties. The reason why differences are observed between verbatim and gist information is because the verbatim representation encodes each word in the hierarchical parse structure of the sentence while the propositional representation encodes the essence of the sentence (at least for purposes of this experiment) in a more compact (fewer chunks) form. This compactness means fewer and more efficient retrievals. When we look at the experiment of Zimny (1987), which used accuracy measures with longer delays, we also see that the more compact representation means that fewer things can be lost to forgetting.

Anderson (1972): All-or-None versus Fragmentary Recall

Representational complexity in the previous experiment was measured in terms of the number of chunks it took to encode the propositional representation and the syntactic representation. These representations, with separate chunks for each term, might strike the reader as quite fragmented. For instance, Kintsch (1974) or Anderson (1983) would treat the proposition in Fig. 2b as one unit rather than three separate chunks. Such a fragmented representation implies that we should observe fragmentary sentence recall such that some but not all of the concepts from the proposition might be recalled. There is clearly fragmentary recall of propositional information as was documented in Anderson (1972). There has been some controversy over the magnitude of this partial recall, with R. C. Anderson (1974) dismissing it as insignificant

while others developing special theories to account for it (Jones, 1978). Figure 8 plots the data from Anderson (1972) and illustrates the half-empty, half-full nature of this debate. The figure plots number of concepts recalled from sentences consisting of four (Experiment 1) or five (Experiments 2 and 3) concepts. In the case of four concepts, the sentences were of the form "In the park the hippie touched the debutante." And in the case of five concepts Anderson (1972) used sentences like "In the park the hippie touched the debutante at night."[4]

If a sentence has n concepts and one concept is used to cue recall of the sentence, there are 2^{n-1} possible patterns of recall including all remaining items recalled, no items recalled, and various possibilities of partial recall. The data in Fig. 8 are plotted in terms of the proportion of trials on which various patterns occurred with zero to four concepts recalled. Except in the case of zero items recalled or total recall, there are multiple possible patterns of partial recall. Figure 8a plots the proportion of each possible pattern for a given number of words recalled. Figure 8b plots the total proportion of all patterns for a given number of words recalled. In all of these experiments about 60% of trials resulted in total failure of recall. The real interest lies in the distribution of the remaining data in terms of the probability of a particular pattern of items being recalled as a function of the number of items in the pattern. With the exception of recalling nothing, the event of recalling all elements is much more frequent than any other specific recall pattern (see Fig. 8a); however, there are many possible patterns of partial recall and the total frequency of all of these patterns of partial recall is about double the frequency of perfect recall (see Fig. 8b). The probabilities of partial recall were 24, 26, and 29% in the three experiments while total recall was 12, 10, and 18%. Thus, partial recall is clearly a prominent aspect of recall despite a disproportionate tendency to recall everything.

Figure 8 also displays the predicted recall patterns by ACT-R according to the Retrieval Probability Eq. (3). Because the surface structure was unlikely to be available at the delays used in these experiments (about 10 min), the model we produced only used the propositional representations like those in Fig. 2b. The model depends both on the propositional encoding and on the referent pointed to by the propositional chunks but first we discuss what can be achieved by just the propositional encoding. The propositional representation by itself produces a certain all-or-none character in the recall. The probe consists of a single word and, to begin recall, the participant must retrieve the chunk that contains the probe concept. From this chunk, the participant can retrieve the proposition, which is necessary for the recall of the remaining terms. Thus, conditional on retrieval of the chunk encoding the probe, the probability of the various recall patterns satisfies the binomial formula $p^m \times (1 - p)^n$, where p is the probability of recalling a chunk encoding that a term occurred in the proposition, m is the number of other terms recalled, and n is the number not recalled.[5] However, before any term can be retrieved from the proposition, it is necessary to retrieve the chunk connecting the probe term to the proposition. The probability of retrieving this probe chunk is p. Thus, this model predicts that the probability of retrieving m elements and failing on n is:

$$p \times p^m(1 - p)^n = p^{m+1}(1 - p)^n \text{ if } m > 0$$
$$(1 - p) + p(1 - p)^n \text{ if } m = 0,$$

where the first p in the first line reflects the retrieval of the probe chunk giving the proposition and the first $1 - p$ in the second line reflects the failure to get to the proposition. Interestingly, Ross and Bower (1981) found that a mathematical model such as the one given above does a good job in predicting recall of unrelated word sets. However, such a model cannot predict the pattern of recall from sentences. It can predict the high frequency of zero elements recalled but not the high frequency of all elements recalled.

---

[4] In some experiments Anderson (1972) used other five-concept sentences but these were the ones we used in all of the simulations.

[5] Throughout this discussion we derive the predictions for specific patterns of recall (i.e., Fig. 8b) from which the prediction for total frequency (i.e. Fig. 8b) of all patterns can be derived.
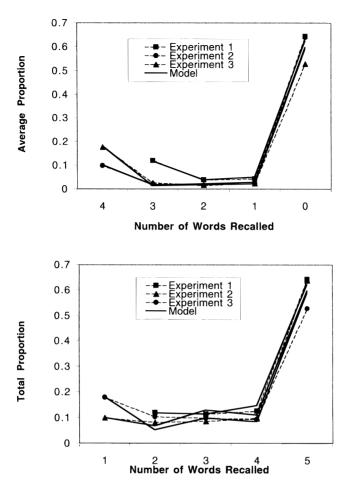
FIG. 8. Proportion of recall of various sentence patterns from Anderson (1972) and ACT-R predictions. (a) The mean proportion of each pattern with the specified number of words recalled. (b) The total proportion of all patterns with the specified number of words recalled.

This model predicts recall patterns that correlate $-.135$ with the data in Fig. 8a (when we exclude the data points for zero items recalled) in contrast to the .995 correlation exemplified by the ACT-R model that we used.

The successful ACT-R model involves an important embellishment. It assumes that at study there is a certain probability that participants are able to retrieve a referent for the target sentence. So, given "The hippie touched the debutante in the park," the participant might retrieve an episode from the movie *Hair* as the referent. If ACT-R can retrieve a chunk that links a probe word to a studied proposition and the chunk contains a pointer to a referent proposition, it can use this proposition to infer what the other terms were (see Fig. 3). Thus, the probability of recalling m and not recalling n in the new model is:

$$p(R + (1 - R)p^m) \qquad \text{if } m = \text{max}$$
$$(1 - R)p^{m+1}(1 - p)^n \qquad \text{if } 0 > m > \text{max}$$
$$(1 - p) + (1 - R)p(1 - p)^n \qquad \text{if } m = 0,$$

where p is the probability of retrieving a chunk encoding that a term is in the studied proposition and R is the probability of finding a referent at study. This implies better recall for the sentence if participants are encouraged to find referents for the sentence. Experiment 3 contained a test of this proposal: Participants were asked to imagine a referent for the sentence and recall

was higher in that experiment. As Fig. 8 shows the major impact of this manipulation is on the frequency with which participants can retrieve all the elements (10% for Experiment 2 vs 18% for Experiment 3).

Three parameters were estimated to fit the model. There was a probability, R, of finding a referent estimated at .20 for the nonimagery Experiments 1 and 2 and at .39 for the imagery Experiment 3. There is p, the probability of retrieving a studied chunk, which was estimated at .44. However, this probability cannot be directly set in ACT-R but results from the setting of three other parameters: the activation of the chunk (A), the threshold (τ), and the activation noise (s) according to Retrieval Probability Eq. (3). Based on prior models (e.g., Lebiere, 1998) we set $s = 0.20$. We chose τ to be 0.3, consistent with the model for the next experiment (by Reder, 1982) that we model. To get a retrieval probability of .44 we estimated A to be .25, just under the threshold.

In addition to providing an excellent fit, this model provides an interesting perspective on sentence memory and all-or-none recall. In this model, perfect recall depends on finding a referent for the sentence in past experience, not on any inherent "Gestalt" properties of a proposition. One consequence of using a referent is that participants may not always recall the same words but rather similar-meaning words. For instance, while "park" may be in the sentence it might really be a "forest" in the referent and so "forest" will be recalled. R. C. Anderson (1974) reports about 20% of all words recalled are not the actual words studied but rather are semantically related to the studied words. Graesser (1978) similarly reports that intrusions (which are a minority of the errors, the majority being omissions) tend to be semantically related.

Reder (1982): Retrieval versus Inference

There are two ways that one can decide that a sentence about a story is true if one has established a referent for the whole story. One is to try to directly retrieve it (its surface encoding or its propositional encoding). The other is to infer the sentence from other sentences that can be re-

called. Thus, even if we cannot directly recall that Bob ate the meal, if he went to a restaurant, ordered a meal, and paid the bill we might be willing to infer that the meal was consumed. Reder (1982) has referred to such a judgement as a "plausibility judgement" and noted that in most real-life situations people are asked to judge what they believe to be true and not to judge what was literally stated. Other researchers (e.g., Graesser & Zwaan, 1995; Kintsch, 1998) have taken such inferences as indicating the creation of a situation model, which involves embellishing the stated material with a mental representation of the situation implied by the material. A significant issue in the literature on text memory is how many of these inferences are made during normal reading of the text and how many are made only when tested. Because of the architectural commitments of ACT-R, we are committed to the position that few inferences can be made at study if study occurs at normal reading or listening rates. In our model, those few inferences generated during reading involved adding a pointer from the chunks encoding the proposition to a past referent. This referent link enables inferences at the time of test.

We first test the ACT-R model of such inferences with Reder's (1982) experiments. These experiments looked at the transition from retrieval-based judgments to plausibility-based judgments over time. In her task, participants read stories and then had to judge either whether sentences were explicitly presented as part of the story (in the recognition condition) or whether they were plausible (in the plausibility condition). Reder's stories consisted of complex, free-form sentences. To simplify the syntactic processing, we presented ACT-R with stories consisting of subject-verb-object sentences like "Bob entered the-restaurant," "Bob ordered the-meal," "The-waiter delivered the-meal," and "Bob ate the-meal." Then ACT-R was tested either with sentences it had studied, like "Bob entered the-restaurant," or sentences which were consistent with the script, like "Bob left the-restaurant," or in the plausibility condition with sentences that did not fit the script, like "Bob delivered the-meal."

Participants were tested either immediately after reading the story (which Reder interpreted as a 120-s delay), after 20 min, or after 2 days. Figure 9 displays the latencies for the old (studied) sentences (which were targets in both the recognition and the plausibility condition), for plausible new sentences (which were foils in the recognition condition and targets in the plausibility condition), and for implausible sentences (which were foils in the implausible condition).[6] With longer delays between reading the story and test, participants showed large increases in latencies in the recognition condition but a net decrease in latencies in the plausibility condition. Figure 10 displays the error data, which show a large increase in error rates for recognition judgments and relatively constant error rates for plausibility judgments.

The ACT-R model for this experiment is a simplified version of the model offered in Reder (1982). Reder's model assumed that participants could judge sentences by either a retrieval strategy or an inference strategy. The retrieval strategy in ACT-R was implemented by the same recognition model (see Fig. 7) that we used for modeling Anderson (1974). The inference strategy involved retrieving the referent of the story (in the preceding example this would be a proposition in the restaurant script) and seeing if the test proposition was stored in the same script. In the plausibility condition the model either (1) tried retrieval first and only switched to inference if it could not retrieve the sentence; or (2) tried the inference strategy first, in which case it just omitted retrieval. The inference strategy is faster because of the stronger encoding of the referent propositions but is somewhat less accurate because some studied sentences might not be judged as plausible (because they are not stored as part of the script for that participant) but could be retrieved. Reder (1982) also assumed that participants mixed strategies in the recognition condition; however, for simplicity the ACT-R model always tried retrieval in this condition and never plausibility.

In modeling the effect of delay we assumed that the immediate condition represented a 120-s

delay, the 20-min delay condition 1200 s, and the 2-day delay 5000 s. The 2-day delay value is taken from other research (e.g., Anderson, Fincham, & Douglass, 1999; McBride & Dosher, 1997) showing that decay dramatically slows after the experimental session is over and can be modeled by a slowing of the clock. The 5000-s estimate is based on Anderson, Fincham, and Douglass, who showed that each day after the experimental session is approximately equivalent to half an hour in the experiment. This may reflect the decrease in interference when the participant leaves the context of the experiment.

ACT-R allows us to model how participants will shift between strategies in the plausibility condition. Table 3 presents an analysis of the relative utilities of the two strategies at various delays. As can be seen, at all delays the inference strategy has a latency advantage. This is because the participants avoid searching for the sentences which will be futile for the three-fourths of the probes that do not involve studied sentences. This advantage slightly increases with delay. The retrieval strategy has a slight accuracy advantage for judging plausibility on those trials involving a studied sentence because sometimes participants did not judge nonpresented plausible sentences as plausible. We estimated that only 90% of these sentences would be judged plausible by the plausibility strategy.[7]

---

[7] In an immediate test Reder found that participants are 10% more likely to judge a sentence as plausible if it has been presented.

TABLE 3

Analysis of Strategy Selection in the Plausibility Condition Reder (1982)

	120 s	20 min	2 days
Retrieval Strategy			
Accuracy (P)	.861	.861	.855
Mean Time (C)	2.89	3.02	3.07
Utility (PG-C)	26.38	26.25	26.01
Inference Strategy			
Accuracy (P)	.842	.842	.842
Mean Time (C)	2.31	2.36	2.44
Utility (PG-C)	26.32	26.27	26.19
Difference in Utility	0.06	−0.03	−0.18
Probability of Retrieval	.78	.37	.03

Note. G = 34.

---

[6] The data in Figs. 9 and 10 are the average of Reder's two experiments.
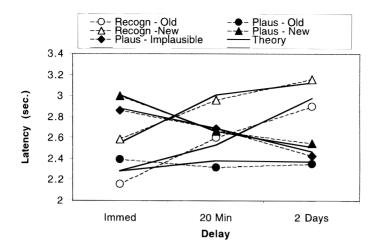
FIG. 9. Latency data from Reder (1982) and ACT-R prediction. Data are plotted for the two types of judgments (Recog vs Plaus) and type of sentence (Old, New, and Implausible).
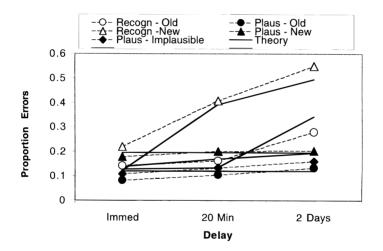

FIG. 10. Error rates from Reder (1982) and ACT-R predictions. Data are plotted for the two types of judgments (Recog vs Plaus) and type of sentence (Old, New, and Implausible).

On the other hand, every retrieved sentence is judged plausible. The accuracy advantage for retrieval is small because there is only a 10% advantage for only one-quarter of the probes that had been studied, and this only occurs if the studied sentence can be retrieved. This advantage reduces with time because a smaller proportion of the studied sentences can be retrieved. Thus, the retrieval strategy has an advantage in terms of probability (P) of a correct answer, while the inference strategy has an advantage in terms of the time (C) to produce an answer. As described with respect to Conflict Resolution Eq. (5), these factors are combined into a net utility that is calculated as $PG - C$. The value estimated for G is 34.[8] Table 3 also shows the differences which lead to the differential choice of strategies according to the Conflict Resolution Equation (5) with the t parameter estimated at .05 as in the model for Anderson (1974). These probabilities are given in the final line of Table 3.

The attempt to judge a sentence can end in one of three ways—ACT-R is unable to retrieve any proposition (studied or script), ACT-R can retrieve a proposition that mismatches the probe sentence, or ACT-R can retrieve a proposition that matches. If it matches, the ACT-R model responds yes; if it mismatches it responds no; if no proposition is retrieved the model guesses between yes and no with equal likelihood. We estimated that participants took .8 s to make that guess but we did not model these guessing processes. We also estimated a .85-s intercept time, which is .2 s longer than in the model for Anderson (1974). This extra time probably reflects the extra time to comprehend the more complex sentences that Reder used. We also fit Reder's error data and to do this we had to estimate a probability of making a slip and giving the unintended response which we estimated to be .12. We achieved a correlation of .977 for latency and .927 for error rates with 5 parameters estimated (see Table 1). These are comparable to the fits reported in Reder (1982), who estimated 20 parameters but also fit other aspects of the data we did not. The two parsimonies achieved by the ACT-R model are that it does not need to estimate separate latency and accuracy parameters for the different delays and it does not have to estimate separate probabilities of strategy selection for the different delays.

The basic insight of this simulation is that we can achieve the inferential capacities associated with situation models by simply storing a pointer to a existing knowledge structure. The previous simulation of Anderson (1972) had shown that this can also serve as the basis for the all-or-none character of recall. The subsequent simulations will show how this mechanism can produce some of the other effects associated with inferential memory. This situational or script information is better retained than the studied propositional information because it has received more practice in the past and not because of different retentive properties. We claim

that equivalent practice would convey the same retentiveness on the studied propositions.

The standard assumption in the literature has been that participants will use the most specific representations if available and only use the more inferential if the others are not available. However, the ACT-R model, like Reder (1982), makes choice among representations strategic. Participants will tend to use whichever representation has the highest net utility. Reder (1987) showed that participants' choice between the retrieval and inference strategies will change depending on which strategy has been locally successful.

Zimny (1987): Surface versus Propositional versus Situational Information

Zimny (1987; reported in Kintsch et al., 1990, who also report a CI model for the experiment) conducted an experiment that had considerable similarity to that of Reder (1982; also Reder, 1976, 1979) but which focused on accuracy of judgments rather than latency. Zimny looked at sentence memory just after reading a story, 40 min after studying the story, 2 days after, or 4 days after. Participants were presented with verbatim sentences, paraphrases (which were identical propositionally to the studied sentences), inferences, or novel unrelated sentences. Unlike the judgments in Reder (1982), Zimny's participants were asked to discriminate verbatim sentences from all other sentences including paraphrases. Figure 11 shows the proportion accepted from the four categories of probe sentences as a function of delay. Participants more rapidly lose ability to discriminate verbatim sentences from paraphrases than they lose the ability to discriminate between studied propositions and inferences. We decided to adapt the two-strategy model that we used for Reder (1982) to make the verbatim judgments in this experiment. We assumed that participants were selecting among the following strategies.

1. Retrieval strategy: Try to retrieve a verbatim trace (e.g., Fig. 2a) to match the sentence. Only if this fails go on to retrieve a propositional trace (e.g., Fig. 2b). If no

---

[8] *G* was not estimated in the model (Table 2) for Anderson (1974) because accuracy remained at ceiling over the short period of that experiment and so did not differ between the two strategies.
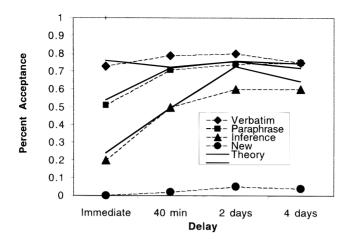
FIG. 11. Results from Zimny (1987) and ACT-R predictions. Data are plotted for the two types of judgments (Recog vs Plaus) and type of sentence (Old, New, and Implausible).

such trace can be retrieved assume the sentence was not studied. This strategy will reject inferences and unrelated sentences since there are no traces of these sentences. It will reject paraphrases if either a mismatching verbatim trace can be retrieved or the propositional trace cannot be retrieved. It will reject verbatim sentences only if neither the verbatim nor the propositional trace can be retrieved.

2. Inference strategy: Simply determine if the sentence is part of the script. This strategy will accept all sentences except novel unrelated sentences.

We estimate that the shortest delay was 60 s. At this delay, the retrieval strategy will enjoy greater success in discriminating verbatim sentences (which is the participants' task) but will also take longer to execute since the chunks formed to encode the study sentence are weaker than the referent chunks. As time passes, however, the accuracy advantage of the retrieval strategy disappears as memories decay and their latency cost increases—just as the retrieval strategy lost relative to the inference strategy in the simulation of Reder (1982). Table 4 presents an analysis of the relative utility of these strategies comparable to Table 3 for Reder (1982). The value of G estimated in this experiment was 10.5. The fact that it is lower

than the G from Reder (1982), which was 34, is interpreted as Zimny's participants placing less emphasis on verbatim accuracy than did Reder's participants on the accuracy of their plausibility judgments.[9] These net utilities can be converted into probability of choice through the Conflict Resolution Equation using the same value of the noise parameter t of .05 that was used in the earlier models. We also estimated a probability .24 of slipping and producing the wrong response. The overall correlation with the data is .956.

As with the Reder model, this model illustrates how participants' choice among strategies is determined by the relative availability of the memory structures. The verbatim structure is the most fragile because it is the most complex and the situation referent is most permanent because it has been well practiced before the experiment. There are no inherent differences in the traces set down in the experiment. It is interesting to note in Fig. 11 that, according to the theory, even acceptance of inferences should start dropping after 4 days. This trend is only slightly apparent in the data but eventually this would happen as participants come to completely forget the stories that they have studied and so forget the connections of the story to the

[9] The value of G is really being constrained to produce a 50% strategy mix at the 40-min delay.

TABLE 4

Analysis of Strategy Selection in Zimny (1987)

	Immediate (60 s)	40 min	2 days	4 days
Verbatim Strategy				
Accuracy (P)	.90	.79	.60	.51
Time (C)	2.71	3.04	3.21	3.21
Utility (PG-C)	4.75	3.78	2.56	2.11
Inference Strategy				
Accuracy (P)	.67	.67	.67	.67
Time (C)	2.27	2.42	2.51	2.60
Utility (PG-C)	3.90	3.78	3.70	3.57
Difference in Utility	0.85	0.00	-1.14	-1.46
Probability of Verbatim	1.00	0.49	0.00	0.00

Note. $G = 10.5$.

referent. Also note that initially the model accepts few inferences and a reduced number of paraphrases. This is because initially the model is predominantly using the verbatim strategy which rejects paraphrases and inferences. This initial blocking of intrusions by the verbatim trace is similar to the proposal of Brainerd, Reyna, and Kneer (1995), who find that a verbatim trace can block false alarms. They also find that this effect decreases with delay.

After reading an earlier draft of this article, Charles Brainerd asked us to consider whether this model predicts the pattern of dependencies reported in an extensive series of sentence memory studies of children and adults (Reyna & Kiernan, 1994, 1995; Kiernan, 1993; Lim, 1993). Those experiments asked participants to try to discriminate among verbatim sentences, paraphrases, and inferences just as in the Zimny experiment. Of interest was how performance varied between immediate recall and delayed recall (often a week later). On immediate memory tests acceptance rates for verbatim sentences were stochastically independent of acceptance rates for paraphrases and inferences but the acceptance rates for paraphrases and inferences were positively correlated. On the delayed test, acceptance rates for all three types of sentences were stochastically dependent.

We examined the issue of stochastic independence in the Zimny simulation and how the

predictions of the ACT-R model would depend on the strategy. In the case of the retrieval strategy, the model produces a dependence between the acceptance of verbatim sentences and paraphrases because both will be accepted if there is a propositional trace and no verbatim trace to reject the paraphrase. This means that in the absence of a verbatim trace either both will be accepted or neither will. However, in the immediate condition of the Zimny experiment, since the propositional trace is almost always present, this source of covariation is removed. In the immediate condition, verbatim sentences are rejected only if the participant slips and slips are random events, uncorrelated with anything else.

The inference strategy produces a dependence between the recall of all three types of sentences because they depend on the finding a successful referent. We assumed in our model of the Zimny data that participants always succeeded in finding a referent at study but to the extent that they did not, there would be stochastic dependence. Since participants only adopt an inference strategy at delay this predicts the observed stochastic dependencies at delay. In summary, the ACT-R model seems generally consistent with the reported patterns of stochastic dependencies. It produces dependencies between all types of sentences except for verbatim sentences in the immediate condition whose acceptance rates are at a maximum.

Schustack and Anderson (1979): Sentences with Referents versus Sentences without Referents

As seen in the previous models, ACT-R can produce inferential recall simply by adding a pointer from chunks encoding the studied proposition to an existing proposition in a referent context such as a script. There is no attempt to copy over the structures from the referent to add explicit inferences to the sentence or story representation. As we saw in the model for Anderson (1972), this can improve memory because one can use the referent proposition to recall the sentence. However, the referent pointer also creates the potential for just guessing any proposition in the referent even if it is not pointed to by a chunk from the memory experi-

ment. Anderson (1972) did not use sentences with known referents and thus guessing could not be assessed. We now consider two studies that explicitly manipulated the availability of known referents.

The experimental literature is not consistent on whether memory is enhanced for referent-consistent material. The best way to assess this issue is with a recognition memory paradigm in which participants are tested with referent-consistent sentences that came from the story and referent-consistent sentences that did not. Improved memory would be reflected in greater discriminability, poorer memory in worse discriminability, and a "guessing bias" in the form of a greater tendency to accept referent-consistent sentences, whether they occurred or not. We describe below an experiment by Bower, Black, and Turner (1979) that can be interpreted as showing poorer discriminability and bias. However, first we describe an experiment by Schustack and Anderson (1979) that can be interpreted as showing increased discriminability as well as increased bias.

Schustack and Anderson (in an elaboration of Sulin & Dooling, 1974) had participants study stories about fictional figures that had parallels to well-known public figures. Thus, they might be told that Yoshida Ichiro was a Japanese politician of the 20th century who was "responsible for intensifying his country's involvement in a foreign conflict" and other such facts consistent with the American president Lyndon Johnson.[10] In the experimental condition participants were told about the parallel and were reminded at test. They were asked to identify sentences which they had studied. They were tested with sentences that they had studied and that were true of the parallel as well as sentences that they had not studied and were true of the parallel. Participants achieved 87.9% hits on the targets while showing only 17.9% false alarms on related targets. In one control condition they were not informed about a parallel at study or test and achieved 67.3% hits and 13.6% false

alarms. Perhaps a better control was one in which they were given the name of a nonanalogous public figure at study and test—here they achieved 71.6% hits and 12.6% false alarms. In terms of d' and bias measures, participants who studied and judged the sentences with a referent had d' values of 2.09 in the experimental condition versus 1.55 and 1.72 in the two control conditions. In terms of bias, the value of β was .77 for the experimental condition versus 1.67 and 1.63 in the control conditions (where values less than 1 indicate a tendency to say "yes" while values greater than 1 indicate a tendency to say "no"). Figure 12 graphically represents these data, averaging together the two control conditions (which is referred to as no-referent). Thus, participants were better when they had an appropriate referent. Another experiment also established that they had to have the referent given both at study and at test to enjoy this benefit.

The ACT-R model we have presented provides a basis for enhanced memory when there is a referent because it stores a pointer to the referent proposition. Just as in the model for recall in Anderson (1972), participants can use this referent proposition to reconstruct the sentence when they cannot directly recall it. This referent-based recall can be further enhanced if we assume that participants have some tendency to accept any proposition in the referent structure, not just the one pointed to in the referent slot. The former process is responsible for the better memory while the latter process is responsible for the bias.

In adapting the ACT-R model of the Reder task for this experiment, we estimated three parameters. One was the retrieval threshold τ [see Retrieval Probability Eq. (3)], which was set to −0.05. The second parameter was the slip parameter, which was .125. The third was the probability of accepting the probe if it was part of the referent's history but not connected to a studied proposition. This was .06 and reflects the bias to accept related sentences. The d' values are 2.10 for the experimental conditions and 1.71 for the controls and the β values are .82 for the experimental condition and 1.65 for the controls. Under any parameter setting the model would predict greater bias and discriminability in the

---

[10] Note that these analogies are not scripts in the Schank and Abelson sense but reflect the more general sense of referents in our model.
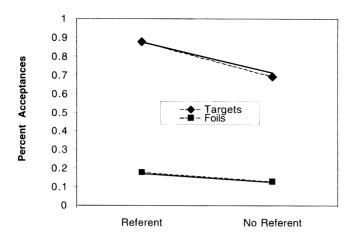
FIG. 12.   Percentage acceptance of targets and foils from Schustack and Anderson (1979) and ACT-R predictions.

referent condition. Given that ACT-R predicts the qualitative result, its good quantitative fit is not surprising, as there are three parameters and four data points. Thus, the most important result is the qualitative conclusion that ACT-R predicts a discriminability advantage for the referent condition in this paradigm. We use the parameter estimates from this experiment to predict the next.

Bower, Black, and Turner (1979): Single versus Multiple Uses of Scripts

Although Schustack and Anderson (1979) presented a situation in which providing a referent improved recognition accuracy, an experiment described by Bower, Black, and Turner (1979) reversed this result. In their experiment, participants studied one, two, or three stories involving the same script such as visiting a health professional. Their participants were asked to give recognition ratings of sentences on a scale from 1 to 7 (1 = *high confidence rejection*, 4 = *guessing*, 7 = *high confidence acceptance*). Figure 13 displays the recognition rates for targets, script-related foils, and script-unrelated foils. The recognition ratings for studied sentences and unrelated foils did not vary much as a function of the number of stories studied. On the other hand, the ratings for script-related foils increased from 3.91 to 4.62 to 4.81 for one, two, and three stories, respectively. It is worth noting about the design that the probability that these foils appeared in another story varied with num-

ber of stories—0% for one story, 50% for two stories, and 100% for three stories.

We attempted to fit these data with the same model and parameters that were used for Schustack and Anderson (1979). This required finding a way for ACT-R to give confidence measures. While we could have developed a more elaborate theory of confidence judgments and have done so elsewhere (Anderson, Bothell, Lebiere, & Matessa, 1998b), it would be a digression to do so here. Therefore, we simply assumed that participants assigned a mean rating of 1.0 to unrecognized script-unrelated sentences, 3.5 to unrecognized script-related sentences, and 6.0 to script-related sentences that they thought they recognized. Otherwise the model and parameters were the same as for Schustack and Anderson. As Fig. 13 illustrates, the model did a good job of reproducing these data (the correlation is $r = .998$). The model produced an increasing effect of number of stories on related foil acceptance because a proposition studied in one story can be accepted as foil in another story. As an example of how this can happen, suppose the participant has studied one restaurant story that includes

"Dan ordered the-meal" and another restaurant story that includes "Bob ate the-meal." In the structure of the Bower et al. materials, the "ordered the-meal' proposition would not be studied with Bob and "ate the-meal" proposition would not be studied with Dan. Then the participant was tested with "Bob ordered the-meal."
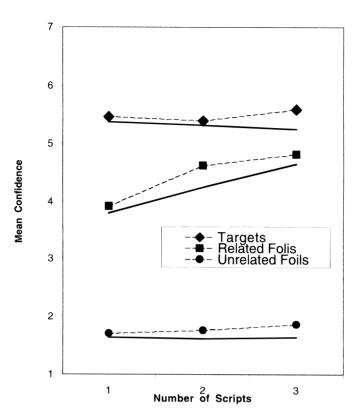
FIG. 13. Mean ratings for targets and foils from Bower, Black, and Turner (1979) and ACT-R predictions.

The participant can find a referent pointer from the-meal to the "person orders the-meal" in the restaurant script because of the story studied about Dan. Retrieving a referent proposition serves as the basis for accepting the probe proposition just as it had in the previous models. The conclusion from this model and the one for Schustack and Anderson is that use of a script sentence in one story makes it available both for correct recognition in that story and for false recognition in other script-related stories. It is worth understanding why Bower et al. found poorer discriminability while Schustack and Anderson found increased. Bower et al. used foils from other stories which produced increased false alarms. On the other hand, they did not have a condition like Schustack and Anderson where there was no recognizable referent. It is in this condition that targets are more poorly recognized. In summary, if a referent is used for a single story it conveys a benefit on that story

relative to conditions in which the story has no referent or the referent is also used for other stories.

CONCLUSIONS

It is not a trivial matter that one can implement models of sentence memory in a cognitive architecture. This is because the architecture comes with certain commitments that are not present when building a model from scratch. ACT-R has commitments about the nature of the retention function which are at odds with commonly held beliefs about the differential forgetting of different types of sentence information. It also has a commitment to serial processing at the symbolic level which might seem at odds with evidence about inferential processing. Thus, success in this modeling enterprise constitutes a significant test of the architecture. Also, since this architecture models cognition in multiple domains (Anderson & Lebiere, 1998) our

success provides support for the view that there is nothing special about sentence processing or sentence memory. Finally, the architecture can bring new integration to a domain like sentence memory by explaining the selection among the various strategies that a participant might bring to bear in recalling a sentence. Basically, participants tend to choose the strategy that delivers the best combination of high accuracy and short processing times and the best strategy can change with delay (basically, the point made in Reder, 1988).

The significance of modeling these six experiments somewhat depends on the consistency of parameter estimates. The decay parameter d was kept at .5 throughout all simulations as it is in all ACT-R models (Anderson & Lebiere, 1998) and as it has been estimated in a extensive empirical investigation (Anderson, Fincham, & Douglass, 1999). The rest of the parameters are displayed in Table 1. With two exceptions the common parameters are remarkably consistent. Both exceptions are associated with the Zimny model that dealt with verbatim memory judgments at very long delays. The G parameter, measuring the value of accuracy was lower by a factor of 3 and the slip probability was higher by a factor of 2. Our model for this task was built on the assumption that the latencies for the memory judgments could be predicted from the model for the Reder task. However, since no latency data are available it was not possible to check these assumptions.[11] A qualification on the generality of the conclusions here is that our model only has been developed to apply to simple and unambiguous sentences. It is an open question how well it will generalize to more complex sentence forms.

Our model has numerous similarities to the fuzzy trace model of Reyna and Brainerd (1995). Like that theory we assume these two traces—a verbatim trace and a propositional trace—and that participants vary in their preference for using the two traces. However, unlike Reyna and Brainerd, the ACT-R model does not assume the differential decay although the verbatim trace is harder to reinstate at a delay because it is more complex. The ACT-R model also offers a systematic basis for deciding which strategy participants will prefer.

An important consequence of the model's parameter commitments was minimal inferential processing. Like other theorists (Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992), we acknowledge that, given enough time, people can elaborate what they are studying with a great many inferences. Indeed, we (Anderson & Reder, 1979; Reder, 1979) have argued that in many conditions where participants are trying to remember material they elaborate richly on the material with great consequence for their memory. However, what is striking to us is that such elaborations are not necessary to account for much of the data. By simply establishing a pointer to a referent, the participant can both enhance memory for the target material and prime retrieval of related material. It is not necessary to make explicit inferences by mapping over the information to the current context. Not only would the generation of such inferences be time consuming but, unless we wanted to attribute special mnemonic properties to these inferences, they would be unlikely to be successfully retrieved at delay. The way to get such strong inferential effects in memory at delay is to count on well-established referents already in long-term memory.

Graesser, Singer, and Trabasso (1994) lay out a set of different types of inferences that might be made during comprehension and they classify different comprehension theories according to which of those inferences a given theory claims that participants make. It is worth reviewing how our own model stands with respect to this set of inferences. The ACT-R model builds chunks that represent the role of the arguments in the sentence. This might require resolving the referent of a noun or pronoun or deciding the role of an argument—which Graesser et al. call local coherence inferences. However,

---

[11]Actually, we have since learned from Zimny (personal communication) that her study involved a word-by-word presentation procedure with 300 ms/word and participants took less than a second after presentation of the sentence to make their judgments. This yields total times comparable to those produced in Table 4 by the Reder model, but the different procedure suggests our extrapolation of the Reder model to her task will be only approximate.

our model will not build inferences if they re-flect new propositions that require new chunks. The only inferential elaboration postulated by our model is the tagging of the chunks repre-senting the proposition with a pointer to a refer-ent. This might also be viewed as in the service of building local coherence. Except as implicit in the referent link, our model does not make goal inferences, causal inferences, inferences of implicit arguments, or any of the other infer-ences that Graesser et al. list.

Verification latency has been used to deter-mine what inferences a participant has made. If a participant recognizes an inference as fast as a stated proposition, the assumption is often made that the inference must have been made while the sentences are studied. While disagreeing on just what inferences are made, Graesser, Singer, and Trabasso (1994) and McKoon and Ratcliff (1992) agree that such latency measures are not strong evidence that the inference has been drawn during initial reading. This is a point that was made earlier (Reder, 1979). This is because postcomprehension processes cannot be ruled out. The ACT-R model presented here illustrates this point. Even though the inference is not gen-erated at study participants can sometimes ver-ify an inference faster than a stated sentence be-cause the referent is more strongly encoded than the sentence and so its components can be more rapidly retrieved.

Much of the research on different inference types has used a word priming methodology (e.g., Long, Golding, & Graesser, 1992; Magliano, Baggett, Johnson, & Graesser, 1993). If it can be shown that words appearing in cer-tain inferences can be recognized more rapidly, it is assumed that these inferences were made during comprehension. Research has docu-mented that words from certain kinds of infer-ences are likely to be primed, particularly if the participants are of high knowledge (e.g., Long et al., 1992; Long & Golding, 1993). We think these results can be understood within the cur-rent theory in terms of the probability that the participants have referent experiences for the stories studied and the probability that these ref-erents have the inferences represented as part of them. If the referent experience can be found

and the relevant inference is strongly associated to the referent, spread of activation will cause these terms to be primed as a consequence of the comprehension process. For instance, a favorite story of Graesser and his colleagues involves a story about a dragon kidnapping the daughter of a Czar. Presumably participants will vary in the amount of prior experience they have had with dragon stories and what facts are represented in their dragon stories. Participants who know a lot about dragon stories are more likely to have a strongly encoded referent in memory that en-ables spread of activation to highly associated concepts. Thus, in our view, this research need not indicate that the inferences are explicitly drawn; only that they are available from the ref-erents. This view is consistent with the recent research on memory-based text processing (e.g., Cook, Halleran, & O'Brien, 1998; Gerrig & McKoon, 1998) that shows that, rather than making explicit inferences, participants just prime relevant background information.

Two of the experiments we modeled (Ander-son, 1972, 1974) involved sentences that were presented out of a prose context while the other experiments involved sentences that were pre-sented in the context of coherent stories. The difference in our treatment of these two classes of experiments was the availability of a referent. We assume that the effect of a coherent story is to help establish a referent for the sentence. Such a referent enables the inferential process-ing that tends be more substantial for sentences presented in a coherent context.

It is worth comparing the ACT-R model with Kintsch's construction-integration (CI) model, which similarly integrates sentence processing with a general theory of cognition. Kintsch em-phasizes the notion of different types of repre-sentations and, unlike ACT-R, does attrib-ute different mnemonic properties to them. Nonetheless, he represents the text and the situ-ation model in terms of propositions and our propositional representation can be basically seen as an incorporation of his representational theory into ACT-R's general chunk-based, de-clarative structure. Kintsch emphasizes the idea that a separate situation model is created for the current text in contrast to our simpler addition

of pointers to an existing referent. The representations postulated by Kintsch are usually created through a hand simulation of a set of rules and so there is not a strong commitment to the processing time for individual steps of comprehension. In contrast, it is ACT-R's commitment to processing time that forces us to our minimalist position. The CI model assumes a spreading activation process at study that operates over a network of propositions to converge on asymptotic values that play an important role in determining the long-term memory familiarity of the propositions, which, in turn, influences recognition judgment. In contrast, activation in ACT-R [Activation Eq. (1)] operates at test to directly determine recognition judgments. Sentence recognition itself is modeled in Kintsch's theory as a familiarity judgment in which the probe evokes some global familiarity response as a function of the strengths of associations to elements in the probe. This is explicitly an importation of the Shiffrin and Gillund (1984) SAM memory model. Our model is quite sensitive to strengths of association but attempts to explicitly retrieve the elements of the original proposition rather than make a global judgment.

In general terms, it can be said that the two models use similar concepts in different ways. ACT-R paints a picture of remembering a sentence that is much more discrete (i.e., discrete steps due to sequential production firing) and Spartan than the one painted by CI. Nonetheless, at least in the case of the Zimny data, the two theories result in roughly equivalent predictions. The Zimny data set is well chosen for the purposes of establishing that ACT-R can offer a competitive account in the domain of language processing where the CI theory has had its most extensive application. However, it is not well chosen to provide a discriminative test of the two theories. The account of the results depends on the existence of three types of representation—an assumption common to both models and basically forced by the data. From the ACT-R perspective, the most critical predictions concern the details of the time course of processing and the CI theory has not been developed for such predictions. In con-

trast, the CI model has been elaborated to account for priming and inference effects that we have not addressed. It would be a good idea to develop both models toward tasks that address issues in common. Until this is done we cannot make strong claims about the real differences between the two theories or their relative merits. However, given that we have advanced the ACT-R theory here, we should say what attracts us to its account: It is committed to the moment-by-moment steps of processing such that it does all tasks from input of the words at study to the production of memory responses at test.

In conclusion, this research has three major implications for sentence memory research: (1) It is not necessary to assume different retention functions for different types of information, (2) it is possible to produce rich inferential effects without extensive elaborations or parallel threads of processing, and (3) the choice among different ways of answering a memory probe is strategic in response to the relative utilities of these strategies.

REFERENCES

Anderson, J. R. (1972). *A stochastic model of sentence memory*. Doctoral dissertation, Stanford University.

Anderson, J. R. (1974). Verbatim and propositional representation of sentences in immediate and long-term memory. *Journal of Verbal Learning and Verbal Behavior*, **13,** 149–162.

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. (1982). Acquisition of cognitive skill. *Psychological Review*, **89,** 369–403.

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

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

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). New York: Worth.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, **38,** 341–380.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1120–1136.

Anderson, J. R., & Lebiere C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

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

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, **128,** 186–197.

Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 385–404). Hillsdale, NJ: Erlbaum.

Anderson, R. C. (1974). Substance recall of sentences. *Quarterly Journal of Experimental Psychology*, **26,** 530–541.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, **11,** 177–220.

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language*, **34,** 157–185.

Budiu, R. (in preparation) *The role of background knowledge in sentence and discourse processing*. Doctoral dissertation, Carnegie Mellon University.

Budiu, R., & Anderson, J. R. (2000). Integration of background knowledge in language processing: A unified theory of metaphor understanding, Moses illusions, and text memory. In *Proceeding of the Third International Conference on Cognitive Modeling*, (pp. 50–57). Groningen, The Netherlands: Universal Press.

Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available during reading? A memory-based view of text processing. *Discourse Processes*, **26,** 109–130.

Fletcher, C. R. (1994). Levels of representation in memory for discourse. In M. A. Gernsbacher (Ed.) *Handbook of psycholinguistics* (pp. 589–608). San Diego, CA: Academic Press.

Gerrig, R. J., & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, **26,** 67–86.

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

Graesser, A. C. (1978). Tests of a holistic chunking model of sentence memory through analyses of noun intrusions. *Memory & Cognition*, **6,** 527–536.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, **101,** 375–395.

Graesser, A. C., & Zwaan, R. A. (1995). Inference generation and the construction of situation models. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Strategies and processing revisited* (pp. 117–139). Hillsdale, NJ: Erlbaum.

Jones, G. V. (1978). Tests of a structural theory of the memory trace. *British Journal of Psychology*, **69,** 351–367.

Kennan, J. M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior*, **16,** 549–560.

Kiernan, B. J. (1993). *Verbatim memory and gist extraction in elementary-school children with impaired language skills*. Unpublished doctoral dissertation, University of Arizona.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, **95,** 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge Univ. Press.

Kintsch, W., Welsch, D. M., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, **29,** 133–159.

Lebiere, C. (1998). Cognitive arithmetic. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought* (pp. 297–342) Mahwah, NJ: Erlbaum.

Lewis, R. L. (1999, March). *Attachment without competition: A race-based model of ambiguity resolution in a limited working memory*. Paper presented at the CUNY Sentence Processing Conference, New York.

Lim, P. L. (1993). *Meaning versus verbatim memory in language processing: Deriving inferential, morphological, and metaphorical gist*. Unpublished doctoral dissertation, University of Arizona.

Long, D. L., & Golding, J. M. (1993). Superordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes*, **16,** 55–73.

Long, D. L., Golding, J. M., & Graesser, A. C. (1992). The generation of goal related inferences during narrative comprehension. *Journal of Memory and Language*, **5,** 634–647.

Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes*, **16,** 35–53.

McBride, D. M., & Dosher, B. A. (1997). A comparison of forgetting in an implicit and explicit memory task. *Journal of Experimental Psychology: General*, **126,** 371–392.

McKoon, G., & Ratcliff, R. (1992). Interference during reading. *Psychological Review*, **99,** 440–466.

McKoon, G., & Ratcliff, R. (1995). The minimalist hypothesis: Directions for research. In I. C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 97–116). Hillsdale, NJ: Erlbaum.

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition*, **22,** 85–94.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–56). Hillsdale, NJ: Erlbaum.

Reder, L. M. (1976). *The role of elaborations in the processing of prose*. Unpublished doctoral dissertation, University of Michigan.

Reder, L. M. (1979). The role of elaborations in memory for prose. *Cognitive Psychology*, **11**, 221–234.

Reder, L. M. (1982). Plausibility judgments vs. fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, **89,** 250–280.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, **19,** 90–138.

Reder, L. M. (1988). Strategic control of retrieval strategies. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. **22,** pp. 227–259). New York: Academic Press.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, **7,** 1–75.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, **30,** 178–191.

Reyna, V. F., & Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor and Symbolic Activity*, **10,** 309–331.

Ross, B. H., & Bower, G. H.(1981). Comparisons of models of associative recall. *Memory and Cognition* **9,** 1–16.

Salvucci, D. D., & Anderson, J. R. (1998). Tracing eye movement protocols with cognitive process models. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 923–928). Hillsdale, NJ: Erlbaum.

Salvucci, D. D., & Anderson, J. R. (2001). Integrating analogical mapping and general problem solving: The path-mapping theory. *Cognitive Science*, **25,** 67–110.

Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, **26,** 159–190.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.

Schustack, M. W., & Anderson, J. R. (1979). Effects of analogy to prior knowledge on memory for new information. *Journal of Verbal Learning and Verbal Behavior*, **18,** 565–584.

Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, **103,** 255–262.

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology*, **9,** 418–455.

Zimny, S. T. (1987). *Recognition memory for sentences from a discourse*. Unpublished doctoral dissertation, University of Colorado, Boulder, CO.