JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 22, 0–00 (1983) DUPLICATE

A Spreading Activation Theory of Memory

JOHN R. ANDERSON

Department of Psychology, Carnegie–Mellon University, Pittsburgh, Pennsylvania 15213

The ACT theory of factual memory is presented. According to this theory, information is encoded in an all-or-none manner into cognitive units and the strength of these units increases with practice and decays with delay. The essential process to memory performance is the retrieval operation. It is proposed that the cognitive units form an interconnected network and that retrieval is performed by spreading activation throughout the network. Level of activation in the network determines rate and probability of recall. With these assumptions in place, the ACT theory is shown to predict interference results in memory, judgements of associative relatedness, impact of extensive practice on memory, the differences between recognition and recall, effects of elaborative processing, and effects of reconstructive recall.

A simple observation about human experience is that we encounter various facts and retain them for varying periods of time. There has probably been more research in experimental psychology studying the many aspects of retention than any other topic. This research has indicated that there are many variables relevant to understanding this retention phenomenon besides the obvious ones of amount of initial study and passage of time. The purpose of this paper is to discuss some of the more important variables in terms of the ACT theory (Anderson, 1976, 1983a). This is a theory which represents knowledge in a network and which has its memory processes defined on that network. It will be shown that many memory phenomena can be understood in terms of the network structures that encode the to-be-recalled facts and the network structures which surround these fact encodings. It will also be shown that the memory process of spreading activation plays a key role in explaining these phenomena.

Preparation of this paper and the research described herein was supported by NSF grant BNS78-17463. I am grateful to Matt Lewis, Lynne Reder, and Miriam Schustack for their comments on the manuscript and to the members of the ACT research group (Gary Bradshaw, Bill Jones, Matt Lewis, Peter Pirolli, and Jeff Schrager) for their discussions relevant to the paper.

This paper considers a wide variety of memory phenomena, but only a fraction of the phenomena that have been documented in experimental research on memory. The criterion in choosing phenomena has been to find ones that nicely illustrate the explanatory power of the ACT spreading activation mechanism operating on a memory network. (This is not to say that I have ignored data that contradict ACT; rather, I have ignored data that seemed irrelevant). While the paper will detail other aspects of the ACT memory theory, this will only be to set the stage for discussion of spreading activation.

The first part of this paper will set forth the principles of the ACT theory of fact memory (Anderson, 1983a). This theory has undergone some significant modifications since it was set forth by Anderson (1976). Using the three-stage organization proposed by Melton (1963), the ACT theory can be divided into principles concerning initial encoding, principles concerning storage, and principles concerning retrieval. The presentation of the ACT theory will be divided according to this categorization. With this theory in place I will apply it to explain a number of memory phenomena.

The Cognitive Unit

Before we can specify the processes of memory it is necessary to say something about the units that these memory processes operate on. In the ACT theory as developed by Anderson (1976), these units were individual associative links where a set of links made up a proposition. There has been a considerable literature now addressed to the issue of whether whole propositions might better be considered the units of memory (e.g., Anderson & Bower, 1973, 1980; Goetz, Anderson, & Schallert, 1981; Graesser, 1978; Jones, 1978; Kintsch, 1974). In the interest of getting on with the major points of this paper, I will not review the considerations but simply state that in the current ACT theory the units of memory are larger structures like the proposition (but see Anderson, 1980, for a discussion).

We use the term *cognitive units* to refer to the units of memory in the current ACT theory. A cognitive unit consists of a *unit node* plus a set of *elements*. For instance, a proposition is a cognitive unit where the elements are the relation and arguments of a proposition and the unit node is the proposition itself. There are clear similarities between my use of a "unit node" and Estes (1972) use of control nodes, Mandler's (1967), Miller's (1956), Simon's (1974), and Wickelgren's (1979) use of chunks, to name just a few of the predecessors to this idea. Cognitive units can be organized hierarchically as when one proposition occurs as a subproposition of another. While propositions are cognitive units, I do not mean to imply that they are the only type of cognitive unit. I have argued elsewhere (Anderson, 1983b) that images and temporal strings (e.g., of words) can also be cognitive units. The points of this paper will not depend critically on whether we assume that this material is encoded as propositions, strings, images, or whatever. For present purposes the encoding can be considered to be the "generic" cognitive unit. Whatever the character of the cognitive unit, it can be represented in network form with the unit node connected to the elements.

An essential feature of a cognitive unit is that it is limited in the number of elements that it can contain. Currently, I have been working with a limitation set at five elements. This means that it is reasonable to consider a paired associate or simple sentence to be encoded by a cognitive unit but that it is not reasonable to consider a paragraph or 30-word list as encoded by a single cognitive unit.

Cognitive units gather their functional significance because they are the units of encoding and retrieval. When part of a cognitive unit is formed in long-term memory, all of it is encoded. Similarly, when part of a cognitive unit is retrieved from long-term memory, all of it is. For instance, when a proposition consisting of subject, verb, and object is formed, all three elements will be encoded in the unit, not just one or two.

Encoding

When a cognitive unit is created, either to record some external event or the result of some internal computation, a transient copy of it is placed in working memory. The basic encoding assumption of the ACT theory is that there is a probability that a transient working memory structure will be turned into a permanent long-term memory *trace*. This encoding assumption is spectacularly simple. The probability is constant over many manipulations. For instance, it does not vary with intention or motivation to learn, consistent with the ample research indicating that intention and motivation are irrelevant if processing is kept constant (e.g., Nelson, 1976; Postman, 1964).

Also, the probability of forming a long-term memory trace does not vary with the duration of residence in working memory. This is consistent with research (e.g., Nelson, 1977; Woodward, Bjork & Jongeward, 1973; Horowitz & Newman, 1969) that fails to find much effect of study time when the information is not being actively processed during study. However, probability of recall is found to increase with repetition even if that repetition is back to back

(Horowitz & Newman, 1969). One would suppose that a second presentation of an item has some chance of creating a new working memory copy. It is also worth noting here the results of Loftus (1972) with respect to picture memory: duration of a fixation on a picture part has no effect on its probability of recall but number of fixations on that part does.

The ACT theory is also quite straightforward about the impact of additional learning opportunities once a trace has been established: all traces have a strength associated with them. The first successful trial establishes the trace with a strength of one unit. Each subsequent trial increases the strength by one unit. Strength of a trace will be important to determining its probability and speed of retrieval. Thus, ACT clearly makes the prediction that overlearning will increase the probability of retention and speed of retrieval—predictions which are equally clearly confirmed.

Retention

According to the ACT theory, traces once formed are not lost but the strength of a trace can decay. Based on data summarized by Wickelgren (1976) and data of our own we assume that trace strength S is a power function of time with the form

$$S = t^{-b} \qquad (1)$$

where t is measured from the point at which the trace was created in working memory and the exponent b has a value on the interval 0 to 1. The function has a strange value at $t = 0$, namely infinity. However, the strength value is only relevant to performance when the trace is out of working memory and must be retrieved (i.e., at times after $t = 0$). I regard this decay function as reflecting a fundamental fact about the system incapable of any further theoretical unpacking short of getting into the physiological character of the system (e.g., see Eccles' (1972) discussion of neural effects of use and disuse). Such a power function is to be contrasted with a exponential function (e.g., $S = a^t$) where $a < 1$. Such exponential functions would produce much more rapid forgetting than is empirically observed.

One of the interesting issues is what the retention function is like for a trace which has had multiple strengthenings. I will simply assume that its total strength is the sum of the strengths remaining from the individual strengthenings

$$S = \sum_i t_i^{-b} \qquad (2)$$

where t_i is the time since the ith strengthening. Evidence for this assumption will be given later when I discuss effects of extensive practice.

Retrieval

To explain the retrieval process in ACT it is necessary to explain more fully the concept of working memory. Working memory contains the information currently available to the system for processing and so combines encoding of information about the current environment, inferences, current goal information, and traces from long-term memory. Since working memory contains traces from long-term memory and since new traces in working memory may be permanently encoded in long-term memory, working memory and long-term memory overlap in terms of their contents. Working memory elements are active to varying degrees. The continuous nature of activation (in contrast to the 1976 ACT) means that membership in working memory is a matter of degree. Less active working memory elements are processed less rapidly, for instance, in a recognition task.

A spreading activation process determines the level of activity in long-term memory. At any point in time certain working memory elements are sources of activation—either because they are encodings of perceptual events or because they are internal concepts currently being processed. Activation can spread from these elements to associated elements in the network of elements and units. As soon as source drops from attention, its activation begins to decay, as does the activation of the network supported by spread from it. Some of the best evidence for this activation analysis of memory is the accumulating evidence for an automatic process that makes information available on the basis of an associative relatedness (e.g., Fischler, 1977; Meyer & Schvaneveldt, 1976; Neely, 1977; Warren, 1972, 1977). Much of this research uses a priming methodology and has been concerned with semantic memory although similar effects have been shown in episodic memory (McKoon & Ratcliff, 1979). One does not think of the typical priming paradigms, such as naming, lexical decision, or Stroop tasks, as memory tasks of the same character as paired-associate recall or sentence recognition. Nonetheless, the claim made here is that the same spreading activation mechanism is involved in memory retrieval. The evidence for the claim is the coherence with which a wide variety of memory phenomena can be interpreted in terms of this mechanism. A basic purpose of this paper is to document this theoretical coherence.

It is assumed that if a trace has level of activation A converging on it, the time to retrieve that trace will behave as an exponential with rate parameter A. In the typical memory experiment, retrieving a trace amounts to processing to the point where some simple recall or recognition task can be performed. This means that retrieval time in an experiment should be of the form

$$RT = I + 1/A \qquad (3)$$

where I is the intercept. It is also assumed that there is a cutoff time K such that if the trace is not processed by then, there will be a retrieval failure. Since rate of retrieval is an exponential function of activation level, probability of successful recall should obey a function of the form

$$PR = 1 - e^{-KA}. \qquad (4)$$

The cutoff means that the previous equation (3) for mean reaction time needs to be amended because the long times are edited out by the cutoff time. Mean time for successful retrievals will be

$$RT = I + 1/A - Ke^{-KA}/(1 - e^{-KA}) \qquad (5)$$

where the third factor is the correction. The form of the correction derives from the underlying assumption of exponential processing times.

The probabilistic retrieval process implies that we should see some mixture of recall of an item and failure to recall it if repeated memory tasks are administered. This is observed (Estes, 1960; Goss, 1965; Jones, 1962). It is also observed that an item successfully recalled on one trial has a greater probability of recall on a second trial. The above analysis would appear to imply independence of recall, but there are a number of explanations for this observed nonindependence. First, the above is only an analysis of retrieval and ignores the all-or-none encoding phenomena which will produce nonindependence among successive recalls. Second, nonindependence would be produced by the successful trial providing a strengthening experience and so increasing the level of activation for the second test. Third, nonindependence could be produced by item selection effects if there were considerable variation among items in the level of activation they could achieve (Underwood & Keppel, 1962).

Computation of asymptotic levels of activation. A schematic retrieval situation is illustrated in Figure 1. I assume that the subject is focusing on an encoding of some event or stimulus (e.g., a sentence) and this is the source of activation. In Figure 1, I have a pair of units focused in working memory and their elements. These elements are also part of units in long-term memory, as is indicated by the associative links going to long-term memory units. The elements of long-term memory are also interassociated with other units, some of which are illustrated in Figure 1. Activation

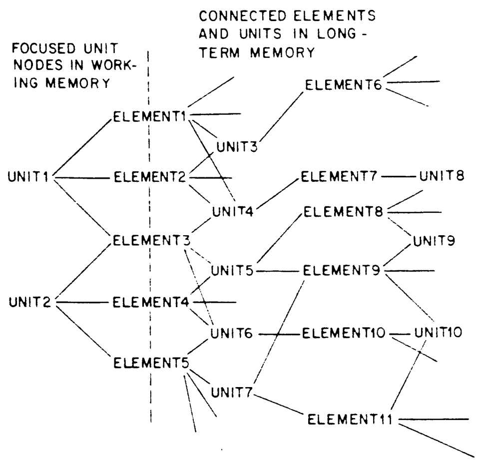


FIG. 1. A schematic representation of the structure of memory relevant to retrieval dynamics. Several units and their elements are active in working memory and activation spreads from these elements through the long-term memory network.

can spread from the elements of focused units throughout the long-term memory structure. These elements of focused working-memory units are referred to as *sources of activation*. The amount of activation they emit is a function of their strength.

A pattern of activation in long-term memory is set up in response to the input of activation from the focused units. The elements and units are the nodes in this activation pattern. Basically, if a node n_y receives activation a_{1y} to a_{iy} from nodes n_1 to n_i, its level of activation is $\sum_i a_{iy}$. The activation that node n_x sends to nodes n_1 to n_j is determined by the strength s_1 to s_j of each node and the activation level of node n_x. If its level of activation is a_x, the amount of activation it sends to node n_k is $la_x s_k/\sum_j s_j$ where l is the loss in activation and $s_k/\sum_j s_j$ is the *relative strength* of node n_k from n_x. Let $f_{xj} = ls_k/\sum_j s_j$ for all nodes j connected to n_x and 0 otherwise. This means that the level of activation of node y is

$$a_y = \sum_x f_{xy} a_x + c_y \qquad (6)$$

where c_y is 0 unless y is a focused element, in which case c_y is the amount of activation coming from this source. Thus, if we have a network of m nodes, we have m simultaneous equations with variables a_1–a_m which we can solve to find the pattern of activation set up by a particular set of focused elements in working memory. While it undoubtedly takes some time for the network structure to reach this asymptotic pattern of activation, I will assume that it is relatively small compared to other processing times and can be ignored. (See Anderson, 1983a, for a discussion of the circumstances under which a stable pattern of activation will be achieved and the time to achieve this pattern.) This is in keeping with a suggestion by Wickelgren (1976) for rapid spread of activation. Recently, Ratcliff & McKoon (1981) have found evidence to support this assumption.

Because of the loss factor l in the above equations, there is a bound on the total amount of activation that will be "pumped" into the network from a source. This is part of what guarantees that the network will move to a stable asymptotic pattern. If the source nodes provide A units of activation, the total activation of all nodes in the asymptotic pattern will be $A/(1 - l)$. This asymptotic amount of activation is distributed among the nodes in the network with nodes getting more activation to the degree they are closely and strongly connected to sources of activation.

Note that this scheme also allows activation to reverberate back. That is, if node 1 connects to node 2, activation from node 1 will spread to node 2 and activation from node 2 will spread back to node 1. Contrary to many people's intuitions, these reverberatory possibilities do not change expectations about a stable asymptotic pattern of activation.

As a small example, consider the network in Figure 2 where we denote the strengths of the long-term memory nodes by s_i and their activation levels by a_1. Let us consider the activation level of element 1 which we denote by a_1. It is part of the focused unit 1 and so will be a source of activation proportional to its strength. Since its strength is 4, it will receive four units of ac-

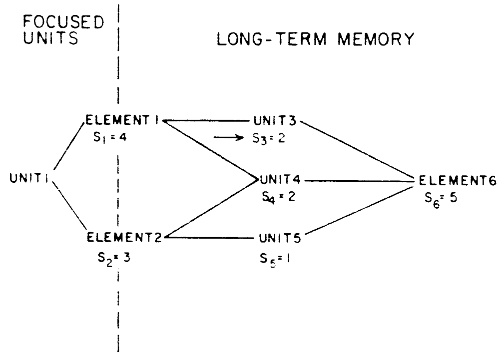


FIG. 2. A simple schematic network structure for illustrating the computation of asymptotic levels of activation.

tivation from itself (the c_y in Eq. 6). Element 1 is also connected to unit 3 and unit 4 and so will receive activation from them. The relative strength of element 1 to all the elements out of unit 3 is $^4/_9$. Similarly, its relative strength from unit 4 is $^1/_3$. I assume that the activation of unit 3 less a loss factor l, is sent out to all the attached nodes according to their relative strengths. The level of activation of unit 3 is a_3. Therefore, element 1 receives $^4/_9 la_3$ activation from unit 3. Similarly, it receives $^1/_3 la_4$ activation from unit 4. The activation of element a_1 is the sum of its self-activation and the activation from units 3 and 4, that is, $4 + {}^4/_9 la_3 + {}^1/_3 la_4$. Thus, we can express the activation patterns in this network by the simultaneous equations

$$
\begin{aligned}
a_1 &= 4 + .44la_3 + .33la_4 \\
a_2 &= 3 + .25la_4 + .38la_5 \\
a_3 &= .5la_1 + .4la_6 \\
a_4 &= .5la_1 + .67la_2 + .4la_6 \\
a_5 &= .33la_2 + .2la_6 \\
a_6 &= .55la_3 + .42la_4 + .62la_5.
\end{aligned}
\tag{7}
$$

Supposing the loss factor l is .8, these equations can be solved as $a_1 = 7.97$, $a_2 = 5.32$, $a_3 = 5.16$, $a_4 = 7.99$, $a_5 = 2.41$, and $a_6 = 6.15$. Given this pattern of activation, unit 4 should clearly be retrieved faster than unit 3 which should be retrieved faster than unit 5.

One of the problems in using this model is that one needs to specify all of the long-term memory network connected at any distance to the working memory elements to derive precise predictions about activation patterns. However, the impact of distant structure is minimal and one's derivations will be quite accurate assuming only the proximal network structure (see Anderson, 1983a).

Another approximating assumption that can be made in computation is to assume that activation will only flow forward from the working memory elements and not reverberate backwards. Under this assumption, derivations become quite direct. Consider the analysis of Figur 2 under these assumptions. In the example, half of the four measures of activation from element 1 go to unit 3 and the other half to unit 4. Similarly, two thirds of the three measures of activation from element 2 will go to unit 4 and one third to unit 5. This means that 2 l measures of activation will go to unit 3, 4 l measures to unit 4, and 1 l measure to unit 5. This clearly preserves the ordering of the three units as obtained from the more exact derivation.

Basically, the view of activation is one in which activation reverberates throughout the network, setting up a stable pattern of activation that reflects how closely connected various nodes are to elements of focused units.

Effects of node strength. Under this view, the critical factor determining retrieval dynamics is the strength of the individual nodes. The strength of a node is a function of its frequency of exposure. The strength of the elements of focused units determines the amount of activation they can emit into the network. Also, more activation is sent down the paths leading to the stronger nodes. Thus, more activation will accumulate in those parts of the network that have stronger units.

Note that frequency of exposure to facts involving a concept will increase the concept's strength and so influence the amount of activation it can emit. On the other hand, learning additional facts about a concept will create competitors to take strength away from existing facts. Thus, we see additional knowledge has both beneficial and harmful effects. Much of this paper is concerned with understanding these mixed effects of more knowledge.

Activation and temporal data. An important distinction between this version of spreading activation and others (Anderson, 1976; Collins & Quillian, 1972; Collins & Loftus, 1975) is that the factor that determines processing times is not time for activation to spread but rather asymptotic level of activation. Asymptotic level of activation affects processing time through ACT's pattern matcher for productions. Although we will not consider in detail the characteristics of ACT's production system, it is the case that any long-term memory fact can affect behavior (as in the generation of a memory report) only by being matched to part of the condition of a production. It is assumed that the production pattern matcher matches more rapidly those facts which are most active.

There are a number of reasons for believing that the important factor is asymptotic level of activation and not time for activation to spread. There are data (Ratcliff & McKoon, 1981; Schustack, 1981; Warren, 1977) which suggest that priming effects due to activation do not have a significant rise time and that the only important factor is overall level of activation. There are data from my laboratory (see Anderson, 1983) that show that fan does not interact with memory complexity and distance but it does interact with complexity of patterns being searched for. This is what one would expect if time to activate a structure was minimal (so distance and complexity were not important) but time to match a pattern was significant.

Summary

So, to recapitulate, experience establishes a network of nodes connected by links of varying strengths. This network consists of cognitive units (e.g., propositions) encoding various facts. At any point in time certain nodes are sources of activation. The levels of activation of the nodes in the network reflect their degree of association to the source nodes. when the source nodes change, spreading activation rapidly adjusts the levels of activation to achieve a new asymptotic pattern. The speed with which information in any part of the network can be processed is a function of its level of activation.

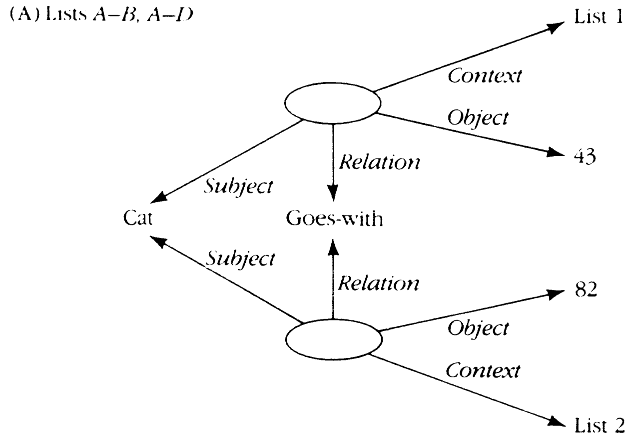
Interference

The analysis of interference effects in memory has been very influential in the development of the ACT theory. Much of the traditional research on interference has been based on a paired-associate paradigm with a recall measure, and it is this paradigm that I would like to discuss first.

The Classic Paired-Associate Paradigm

A frequent contrast in discussion of interference is between the A–B, A–D condition and the A–B, C–D condition. In the first condition, the subject learns two lists composed of paired associates where the two lists involve the same stimuli (A's) but different responses (B's and D's). In the second condition, the lists involve different stimuli (A's, C's) and different responses (B's, D's). The different network representations for the two lists are illustrated in Figure 3.

When a stimulus is presented at the time of test, we assume that it and the list context are elements of a focused unit—for instance, a unit encoding the fact that the stimulus is being presented in the current context. Therefore, these elements serve as sources of activation. The response will be retrieved if (a) a trace connecting stimulus, response, and context has been formed, and (b) it can be retrieved within the cutoff time. Let S^+ be the strength of the trace, S_s^- be the strength of all other traces connected to the stimulus, and S_c^- be the strength of all other traces connected to the context. Let A_s be the activation from the stimulus and A_c be the activation from the context. Then, ignoring interactions due to activa-



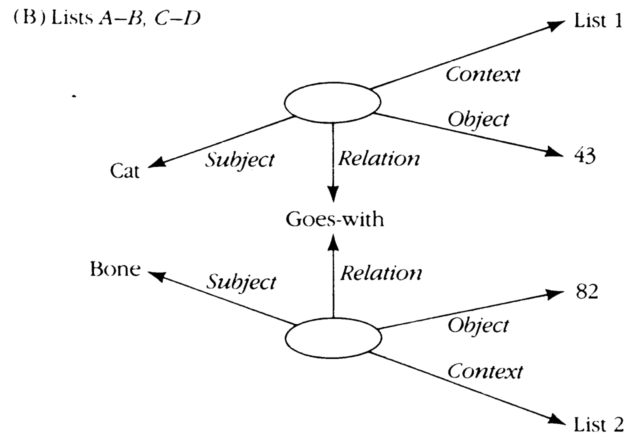


FIG. 3. (a) Network representation for the typical interference condition of a paired-associate paradigm; (b) network representation for the typical control condition of a paired-associate paradigm.

tion flowing backward, the activation converging on the trace from the stimulus and context will be

$$A = A_s(S^+/(S^+ + S_s^-)) + A_c(S^+/(S^+ + S_c^-)). \quad (8)$$

Probability of recall will increase with A as implied by the earlier equation (4). From Equation (8) we see that as the competing strengths (S_s^-,S_c^-) increase we will need more target strength (S^+) to maintain the same level of activation. From this observation derive the basic predictions of negative transfer and retroactive interference. To achieve the same level of performance on the second list in the A–B, A–D paradigm, as in the A–B, C–D, extra study will be required in the interference condition to increase its trace's strength to compensate for the strength of the competing first-list trace connected to the stimulus. This is the prediction of negative transfer. In a retest of list 1 after list 2, the relative strength of the list 1 trace in the interference condition will be reduced by the competing list 2 trace and so lower recall will be observed. This is the prediction of retroactive interference.

The Problem of Proactive Interference

One of the proactive interference paradigms has been difficult to explain in the

ACT framework. This is the two-list paradigm that contrasts A–B, A–D list with the A–B, C–D control. Proactive interference in this paradigm refers to the fact that a *retest* of the second list (at a delay) is worse if it had been in an interfering relation to the first list (ie., retention of A–D is worse than C–D). This is true even though the second lists (A–D or C–D) are brought to the same level of recall before the retention interval. I will assume that if the control and interference conditions had the same level of recall at the beginning of the retention interval, they had the same level of activation. The activation in the control (noninterference) case is

$$A_N = A_S \frac{S_N}{K + S_N} + A_c \frac{S_N}{nS_1 + nS_N}$$
$$= A_S \frac{S_N}{K + S_N} + A_c' \frac{S_c}{S_1 + S_N} \qquad (9)$$

where A_S is the activation from the stimulus, A_c the activation from the context, S_N the strength of the second test trace, S_1 the strength of the first list trace, K the strength of prior associations to the stimulus, and n the number of items in a list. A_c' is A_c/n and reflects the contextual contributions to each list element. The activation in the interference case is

$$A_I = A_S \frac{S_I}{K + S_1 + S_I} + A_c' \frac{S_I}{S_1 + S_I} \qquad (10)$$

where the only differences are that the strength of the second list trace, S_I, must be larger than S_N to compensate for the interference of the first list trace from the stimulus. Note that it is assumed in both cases that there is interference of the first list from the context. At a delay the activation formulas become

$$A_N' = \frac{A_S d_2 S_N}{K + d_2 S_N} + \frac{A_c' d_2 S_N}{d_1 S_1 + d_2 S_N} \qquad (11)$$

$$A_I' = \frac{A_S d_2 S_I}{K + d_1 S_1 + d_2 S_I} + \frac{A_c' d_2 S_I}{d_1 S_1 + d_2 S_I} \qquad (12)$$

where it is assumed that the activations (A_S,A_c) do not change with time and the decay in prior strength (K) is also negligible. The decay (d_1) in the first list strength is smaller than the decay (d_2) in the second list strength, that is, $d_1 > d_2$. These assumptions about decay of strength are implied by the power law retention functions. Under a wide range of reasonable parameter assumptions, although not under all possible parameter assumptions, it can be shown that $A_N = A_I$ implies $A_N' < A_I'$. To take an example, suppose $A_s = 2$, $A_c' = 1$, $d_1 = .75$, $d_2 = .5$, $K = 1$, $S_1 = 1$, and $S_N = 1$. Then $A_N = 1$; to get $A_I = 1$ it is necessary to set $S_I = 2$. Under these parameter assumptions, $A_N' = .733$ and $A_I' = .773$. That is, there is more activation at a delay in the interference condition.

Thus, it can be shown that ACT predicts just the opposite of proactive interference. Interestingly, despite the traditional belief, proactive inhibition has seldom been shown in a contrast of an A–B, A–D paradigm with an A–B, C–D paradigm. Most experiments that have brought the control and interference conditions to the same learning criterion have contrasted a one-list control with a two-list interference (e.g., Koppenaal, 1963; Ceraso & Henderson, 1965; Houston, 1969; Postman, Stark, & Fraser, 1968). In fact, in our own laboratory we have confirmed ACT's prediction of proactive facilitation in the A–B, A–D versus A–B, C–D contrast and have shown proactive inhibition in the A–B, A–D versus C–D contrast. For instance, in one experiment each list in each condition involved 20 word–number pairs (a methodology similar to Anderson, 1981). We looked at retention of the last list a week later after bringing it to 90% initial learning in all conditions. The retention of A–D in the A–B, A–D condition was 70%, the retention of C–D in the A–B, C–D condition was 63%, and retention of C–D in the single-list paradigm was 81%. It should be emphasized that these results were obtained in an unpaced MMFR test. The ACT predictions about the ordering of these paradigms only apply in this situation. List discrimination problems might well produce poorer performance in a A–B, A–D paradigm than a A–B, C–D paradigm when a paced, nonMMFR test is used.

Our history in this effort was that we started from the traditional wisdom that proactive interference could be obtained in the contrast between A–B, A–D and A–B, C–D paradigms. Starting with HAM (Anderson & Bower, 1973) and then ACT (Anderson, 1976), it was observed that these theories could not predict proactive interference when the two lists were trained to the same criterion. Our first response was to try to explain proactive interference away as an experimental artifact (e.g., Anderson & Bower, 1973), and we engaged in a series of experiments to confirm these attempted explanations. The outcome of these experiments was the discovery of failure to confirm an interfering relationship. Subsequently, a search of the literature indicated that most demonstrations of proactive interference involved contrasting a single list with a two-list paradigm. Indeed, Postman, Stark, and Burns (1974) and Postman and Gray (1977) report failure to find proactive interference comparing the two two-list paradigms in an unpaced MMFR test. The demonstrations of poorer second-list retention in the A–B, C–D paradigm have involved paced nonMMFR tests (Postman & Gray, 1977; Underwood & Ekstrand, 1967).

Cumulative Proactive Interference

To restate the conclusion of the previous section, proactive interference (PI) seems to be obtained only comparing two lists with single lists. In this case, the contrast seems a special case of the cumulative PI design (e.g., Greenberg & Underwood, 1950; Keppel, Postman, & Zavortink, 1964; Postman & Keppel, 1977; Underwood, 1957) where it is found that retention deteriorates for successive unrelated lists. The typical design involves having subjects learn one list one day, having them return the next day to recall the previous day's list and then learn a second list, repeating this pattern over days. The ACT analysis, with certain assumptions about context, does predict cumulative PI. The following are the equations for activation for the first and second lists immediately and at delay.

$$A_1 = A_s \frac{S_1}{K + S_1} + A_c' \frac{S_1}{S_1} \tag{13}$$

$$A_2 = A_s \frac{S_2}{K + S_2} + A_c' \frac{S_2}{d_1 S_1 + S_2} \tag{14}$$

$$A_1' = A_s \frac{d_1 S_1}{K + d_1 S_1} + A_c' \frac{d_1 S_1}{d_1 S_1} \tag{15}$$

$$A_2' = A_s \frac{d_1 S_2}{K + d_1 S_2} + A_c' \frac{d_1 S_2}{d_2 S_1 + d_1 S_2} \tag{16}$$

where S_1 is the strength of the first list trace, S_2 the strength of the second list trace, d_1 the decay after 24 hours, and d_2 the decay after 48 hours. Assuming a power law for decay, we have $d_{\mathrm{n}} = t^{-\mathrm{d}}$ where t will be the number of hours. Then we can derive the expression for the *n*th list in a cumulative PI design.

$$A_n = A_s \frac{S_n}{K + S_n} + A_c' \frac{S_n}{\sum_{i=1}^{n} S_i((n-1)24)^{-\mathrm{d}}} \tag{17}$$

$$A_n' = A_s \frac{d_1 S_n}{K + d_1 S_n} + A_c' \frac{d_1 S_n}{\sum_{i=1}^{n} S_i((n-i+1)24)^{-\mathrm{d}}}. \tag{18}$$

The essential observation is that the number of interfering contextual associations increases with number of lists. Therefore, greater trace strength will be required to achieve the same level of activation, that is, $S_n > S_{n-1}$. It can also be shown under a wide range of plausible parameter values that if $A_n = A_{n-1}$ then $A_n' < A_{n-1}'$. This is because with each successive trial (1) a larger portion of the activation comes from the stimulus which decays rapidly, and (2) the decay of the interfering associations from the context slows down.

As an illustration, we considered the case where $A_s = 5$, $A'_c = 1$, $K = 10$, and the exponent for decay $d = .5$. Setting $S_1 = 2$, we got $A_1 = 1.8$. To get $A_2 = 1.8$, we had to set $S_2 = 2.3$ and similarly $S_3 = 2.6$, $S_4 = 2.9$, and $S_5 = 3.1$. Then we calculated the activation at delay: $A'_1 = 1.11$, $A'_2 = .84$, $A'_3 = .73$, $A'_4 = .69$, $A'_5 = .65$. We set K in equation (11) at 1 and got the following probabilities of recall for successive lists at a 24-hour delay: .67, .57, .52, .50, and .48. Thus, the explanation of cumulative proactive interference may lie in increasing contextual interference.

This analysis implies that there should not be cumulative proactive interference to the extent that one could create a novel context for each study. We performed a modest context manipulation experiment to put this prediction to test. Subjects learned three lists of 20 paired associates, either from a computer in a windowless cubicle or from a human experimenter in a windowed seminar room. As in the classic paradigm for cumulative PI, list n was learned on day n to a 95% recall criterion and tested for retention on day $n + 1$. We used all $2^3 = 8$ sequences of contexts for the three lists, each sequence with 10 subjects. Thus, we counterbalanced for any effect of a particular context on learning.

Averaged over sequences, performance was 87.6% on list 1, 86.1% on list 2, and 82.3% on list 3. Thus, there is only a modest buildup of PI which tempers one's expectations about effects of context change. Nonetheless, retention was 88.4% for list 2 when context was changed from list 1 versus 83.9% when it was the same. Similarly, it was 84.5% for list 3 when context was changed from list 1 and 2 versus 78.3% when it was unchanged. The combined differences reach the conventional level for statistical significance, $t_{108} = 2.31$; $p < .05$.

There is very little PI over the three tests in a changed context—87.6% for list 1, 88.4% for list 2, and 84.5% for list 3. On the other hand, the PI is fairly clear when context does not change—87.6%, 83.9%, and 78.3%. This analysis of cumulative PI implies that it results from contextual interference and not stimulus-specific interference. This is consistent with conclusions of Wickens, Moody, and Dow (1981) that PI is due to interference on response set or list differentiation.

In summary, this theory is consistent with the major trends of traditional research on interference. Before going on to other topics, I would like to consider two other subissues about interference because they serve to lay the groundwork for subsequent analyses. These are interference effects on reaction-time measures and effects of network integration on interference.

Reaction Time and Interference

In the ACT framework, reaction time should be a purer measure of interference than percent recall. This is because percent recall reflects both the formation of links, which is not subject to interference, and retrieval, which is. In contrast, reaction time reflects only the retrieval of those links formed. It is also the case that after near-asymptotic levels are achieved in percent recall, reaction time will continue to reflect interference effects. Figure 4 displays data from a paired-associate paradigm (Anderson, 1981). Subjects were given eight trials of study test on list 1, then eight trials on list 2, then four trials of retest of list 1, and then four more trials of retest on list 2. For the interference subjects, the two lists were in an A–B, A–D relation; for the control subjects, they were in a A–B, C–D relation. We have plotted reaction time against accuracy. Note that there are separate functions for interference and control, both of which can be approximated by straight line relationships. However, the interference function is above the control function, indicating that when the two conditions have been equated for percent recall, the

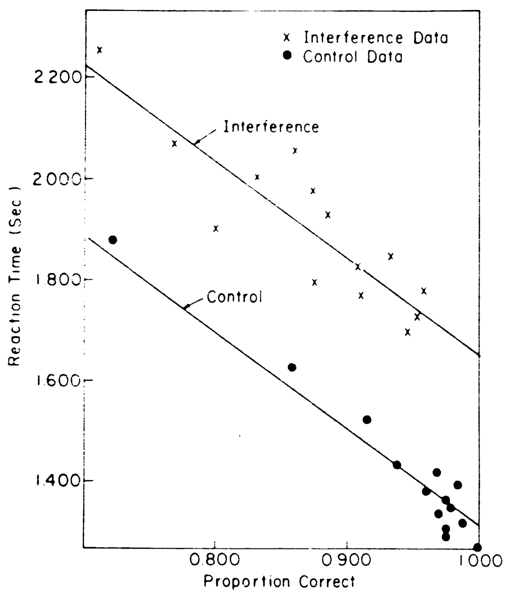


FIG. 4. Plot of reaction time as a function of percent recall for interference condition and control condition from Anderson (1981).

interference condition is at a disadvantage with respect to reaction time. This is consistent with the ACT claim that reaction time is a more sensitive indicant of interference.

Many of the reaction-time experiments we have done have involved subjects recognizing whether they saw particular sentences. In one experiment (Anderson, 1976) we had subjects recognize whether they had studied location–subject–verb sentences such as *In the winery the fireman slept*. We manipulated independently whether subjects studied one or two facts about subject, verb, and object. Table 1 reproduces the results from that experiment classified according to the number of facts studied about each concept. Data are also presented for foil sentences which were created from the same words as the target sentences but were in novel combinations. As can be seen, reaction time increases as number of facts increase on any of the three dimensions. This increase in reaction time has been referred to as the *fan effect* because more facts increase the fan of propositional associations leading from a concept. Increasing the number of facts on any dimension will reduce the amount of activation converging on the memory trace.

It is interesting that the foil items also show this fan effect. The general proposal that we (Anderson, 1976, 1983a; King & Anderson, 1976) have advanced for recognition of foils is that subjects wait a period of time and reject the sentence if they fail to retrieve a matching trace. The amount of time a subject waits is adjusted to reflect the fan of the probe. This is essential for a waiting strategy to work since the subject must allow enough time to elapse for the positive probes to be recognized. Note that subjects take somewhat longer to reject foils. They would have to wait longer than normal target time to make sure the probe was not a target.

Effects of Network Integration

One of the results that has been shown in the interference literature (e.g., Postman & Gray, 1977; Postman & Parker, 1970) is that interference is diminished in an A–B, A–D paradigm if the subject makes an effort to maintain B responses during A–D learning. The consequence of maintaining the two responses should be to integrate them together. Figure 5 illustrates in very simplified network terms the single list condition (a), the two-list nonintegrated condition (b), and the two-list integrated condition (c). Assuming that one unit of activation is flowing into the A stimulus and that all nodes are of equal strength, it is possible to calculate the levels of network activation given the earlier model. Assuming a loss of .8 in spread in equations like those in (8) on page ■, 2.22 units will accumulate at D in the single-list condition, 1.11 in the nonintegrated condition, and 1.43 in the integrated condition. Thus, ACT does predict a benefit of such integrating structure. As we will see in the discussion of elaborative processing, this is just the tip of the iceberg in terms of the memory phenomena that can be accounted for in terms of such network integration.

TABLE 1
LOCATION–SUBJECT–VERB EXPERIMENT (*In the church the sailor sang*)—MEAN REACTION TIMES AND ERROR RATES (IN PARENTHESES)

	$L = 1$				$L = 2$		
Targets	$V = 1$	$V = 2$	Mean		$V = 1$	$V = 2$	
$S = 1$	1220 (.034)	1183 (.017)	1202 (.026)	$S = 1$	1297 (.046)	1402 (.086)	1350 (.066)
$S = 2$	1232 (.069)	1421 (.080)	1327 (.075)	$S = 2$	1358 (.046)	1500 (.082)	1429 (.064)
Mean	1226 (.052)	1302 (.049)	1264 (.051)	Mean	1328 (.046)	1451 (.084)	1390 (.065)
Foils	$V = 1$	$V = 2$	Mean		$V = 1$	$V = 2$	Mean
$S = 1$	1323 (.029)	1387 (.035)	1355 (.032)	$S = 1$	1404 (.034)	1469 (.023)	1437 (.029)
$S = 2$	1320 (.028)	1371 (.028)	1346 (.028)	$S = 2$	1488 (.023)	1511 (.063)	1500 (.043)
Mean	1322 (.029)	1379 (.031)	1351 (.030)	Mean	1446 (.029)	1490 (.043)	1468 (.036)

JUDGMENTS OF ASSOCIATIVE RELATEDNESS

The discussion to date has assumed that activation makes information available for inspection by other procedures. Thus, activation determines the amount of cognitive resources available to process information. In addition to this role, level of activation can also provide information in and of itself. The more active a particular part of the network is, the more it must be related to the current context. Level of activation does not tell us how the network is related, but it does provide information about degree of relation. There is evidence that information about level of activation is available relatively immediately and is used by subjects to facilitate certain types of memory judgments.

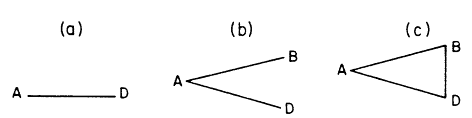


FIG. 5. (a) Schematic representation of a paired associate without interference; (b) representation of two interfering and unintegrated paired associates; (c) representation of two interfering but integrated paired associates.

Rejection of Foils

The experiment by Glucksberg and McCloskey (1981) nicely illustrates the use of information about associative relatedness. They had subjects study affirmative facts such as *John has a pencil*, negative facts such as *Bill doesn't have a shovel* and "don't know" facts such as *It is unknown whether Fred has a chair*. Then subjects had to judge of various statements whether they were true or false or whether their truth was unknown. An interesting comparison involves the speed with which subjects can judge that they do not know something when they have explicitly learned that they do not know (e.g., Fred has a chair), and their speed in cases where their lack of knowledge is only implicit (e.g., Bill has a pencil). Subjects are faster in the implicit case. In this case, subjects can monitor for an intersection of activation between subject and predicate at a trace. If no such locus of high activation appears, subjects can use this low level of activation as evidence that subject and predicate are unrelated and so conclude that the sentence is not known. In the explicit case, however, the "don't know" fact creates a locus of high activation which the subject must explicitly reject. Apparently it is easier to use the low level of activation as a basis for a response than it is to explicitly retrieve the "don't know" tag.

The experiment by Shoben, Wescourt, and Smith (1978) is another example that illustrates the importance of level of activation to rejection of foils in a sentence recognition task. One type of foil involved sentences that were true but not studied (e.g., Tigers have stripes), and the other type involved sentences that were false and not studied (e.g., Tigers have fingers). Subjects were slower and made many more errors at rejecting the true foils, suggesting that subjects were sometimes responding just to the high activation created by intersection of subject and predicate.

Thematic Judgments

The experiments of Reder and Anderson (1980) and Smith, Adams, and Schorr (1978) are examples of situations where subjects can use this strategy of activation monitoring to make positive judgments. These researchers had subjects study a set of facts about a person that all fell under some theme such as going to the circus. So a subject might study

Marty laughed at the clown.
Marty ate cotton candy.
Marty cheered the trapeze artist.

Subjects were transferred to a fact recognition situation where they had to recognize these facts about the individual and reject facts from different themes (e.g., Marty waited for the train). The number of such facts studied about a person was varied. Based on the research on the fan effect one might expect recognition time to increase with the number of facts studied about the individual. However, the material in the typical fan experiment is not thematically integrated as is this material. In these experiments, recognition time did not depend on how many facts a subject studied about the individual.

Reder and Anderson postulated on the basis of data from Reder (1979) that subjects were actually judging whether a probe fact came from the theme or not and that subjects were not carefully inspecting to see if the fact about the individual was studied. To test this idea, we examined what happened when the foils were other predicates consistent with the theme. So if the subject has studied

Marty laughed at the clowns.

a foil might be

Marty liked the animal trainer.

In this situation subjects took much longer to make their verifications and the fan effect reemerged.

We proposed that subjects set up a representation such as that in Figure 6. We assumed that thematically related predicates are already associated to a theme-node like circus. We assume that subjects create a subnode to represent same-theme facts about a person. This subnode is associated to the theme and to the individual theme predicates. We have represented two such theme nodes in the figure and two subnodes. Reder and Anderson manipulated whether subjects studied one or two themes in that study and found evidence that subjects organized knowledge by subnode according to theme. Specifically, we found that subjects were slowed in their judgments of facts in one theme by the addition of another theme but not affected by the number of facts associated with that other theme. Recently, Reder and Ross (in press) extended this finding to the contrast between two and three themes. This is what would be predicted on the basis of Figure 6.

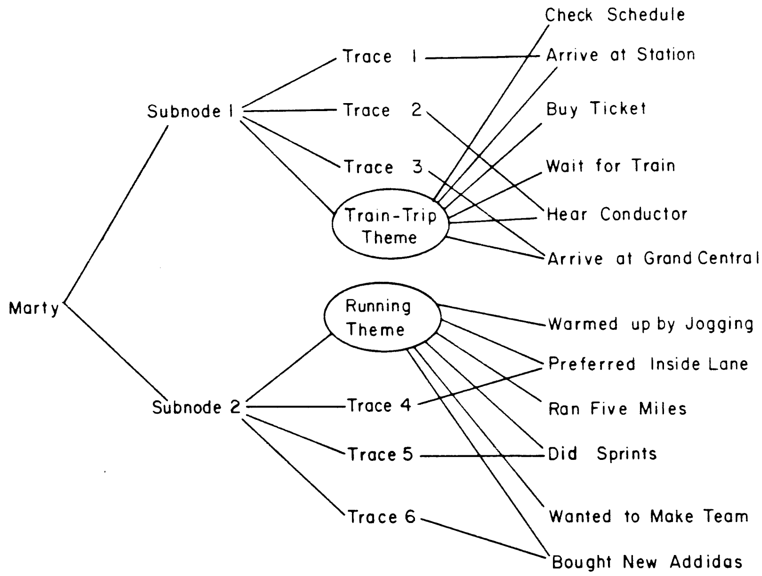


FIG. 6. Network representation for the Reder and Anderson experiments. Facts about Marty are organized into two subnodes according to theme.

Additional themes create a fan out of the person node but the number of facts associated to the second subnode should have little or no impact on the amount of activation intersecting between person and predicate.

In Reder and Anderson we proposed that in the presence of unrelated foils subjects simply retrieved the subnode while in the presence of related foils they had to retrieve the target trace. We were somewhat vague about how a subject might retrieve the subnode and evaluate it, but in the context of the current framework there is an easy explanation. Subjects could respond to the level of activation of the subnode. A high level of activation intersecting from the person and the predicate would be evidence that the subnode was thematically relevant.

Calculation of activation patterns. I have calculated what the activation patterns would be for the Anderson and Reder experiments, assuming a representation like that in Figure 6. In doing so, I set the amount of activation from the predicate (which for simplicity is represented by a single node) to be ten and from the person to be one since I expected more activation from the multiple familiar concepts of the predicate. The conditions of the Anderson and Reder experiment can be classified according to the number of things learned about the tested dimension (1 or 3) and the number of things about the untested dimension (0, 1, or 3). I calculated the pattern of activation in each of these situations. For each of these conditions, Table 2 reports the level of activation of the trace for targets and also the level of activation of the subnode in the presence of a target, a related foil, and an unrelated foil.

Note that the activation of the target trace decreases with the number of facts in the same theme as the probe, decreases when there is a second theme, but shows very little variation of one versus three facts in the nontested theme. This is precisely the fan effect reported by Reder and Anderson when related foils were used. Thus it appears, as hypothesized for the related-foil condition, that activation of the trace is controlling judgment time.

It was hypothesized that when foils were unrelated, level of activation of the subnode and not of the trace would control

TABLE 2
LEVEL OF ACTIVATION OF VARIOUS NODES IN FIGURE 6 UNDER VARIOUS CONDITIONS

		Number of facts about target theme	
1. Target trace in the presence of the target		1	3
Number of facts about other theme	0	7.78	6.83
	1	7.26	6.57
	3	7.24	6.56
2. Subnode in the presence of the target		1	3
Number of facts about other theme	0	7.06	7.49
	1	5.59	6.29
	3	5.50	6.22
3. Subnode in the presence of a related foil		1	3
Number of facts about other theme	0	5.30	5.94
	1	4.07	4.88
	3	4.00	4.81
4. Subnode in the presence of an unrelated foil		1	3
Number of facts about other theme	0	1.40	1.42
	1	.70	.75
	3	.66	.71

judgment time. The subnode actually shows a reverse fan effect in Table 2 when a target is presented as a probe—greater activation when there are three facts. This is because there are more paths converging on the subnode in this case. Although these additional paths may not be direct routes from sources of activation to the subnode, they are nonetheless indirect routes. Thus, presented with *Marty arrived at the station*, activation can spread from *arrive at station* to the train theme to *hear conductor* to subnode 1. Also note that the subnode has a high level of activation in the presence of a related foil. Although this level is not as high as for a target, it is sufficiently high to cause trouble for a scheme of responding to level of activation as a basis for discriminating between targets and related foils.

Reder and Anderson (1980) and Smith et al. (1978) both report no fan effects in the presence of unrelated foils. If subjects were responding purely on the basis of level of activation, there should be a reverse fan effect as Table 2 indicates. Reder and Ross (in press) speculated that the subjects adopt a mixed strategy in this case, sometimes responding to subnode activations and sometimes responding to the trace. The direct and reverse fan effects would then tend to cancel themselves out. Consistent with this hypothesis, Reder and Ross showed that when subjects are explicitly instructed to respond on a thematic basis (i.e., either accept studied sentences or unstudied but related sentences) they do show a reverse fan effect. Reder and Ross also found subjects slower to accept unstudied related sentences than studied sentences in these thematic judgment conditions. This is to be expected from Table 2 because in the presence of related foils (which are the same as Reder and Ross's related, nonstudied targets) there is a lower level of activation of the subnode than in the presence of targets.

Refocusing on Subnodes

The activation patterns were calculated in the previous section under the assumption that the subject spreads activation from the person node like Marty in Figure 6. In these calculations the activation from *Marty* is broken up twice before getting to the target predicate. It is divided once between the two subnodes and once among the facts attached to the subnode. This implies that the activation level of the traces should be no different if six facts were attached to one subnode than if the six facts were divided between two subnodes. In both cases, one sixth of the activation reaches the subnode. In fact, however, there is evidence (Reder & Anderson, 1980; McCloskey & Bigler, 1980) that subjects are faster in the two subnode condition.

These and other results (Anderson, 1976; Anderson & Paulson, 1978) lead to the second aspect of the subnode model, the refocusing process. Even in cases where subjects must retrieve the specific fact, subjects can use level of activation of the subnode to refocus their activation. They select the most active subnode and make that the source of activation rather than the original node. This is a two stage process; first the subnode is selected and then activation spreading from the subnode enables identification of the target fact. Since activation spreads rapidly, the time for the subnode selection should be relatively brief.

This subnode-plus-refocusing model explains the low estimate of strength of prior associations that we have obtained in some previous experiments (e.g., Lewis & Anderson, 1976; Anderson, 1981). As suggested in Anderson (1976), subjects may create an experimental subnode and use contextual associations to focus on it. This would protect them from the interference of prior associations. This model also offers a reason why we may be faster at retrieving information about familiar concepts. Presumably such concepts have a well-developed and perhaps hierarchical subnode structure which can be used to focus the retrieval process on a relatively small subset of the facts known about that concept.

Summary

This section has reviewed how subjects can use judgments of associative relatedness to avoid direct retrieval. When such judgments are implemented in the subnode-plus-refocusing model they can eliminate interfering fan effects and even produce positive fan effects. This subnode focusing strategy will only work when the subnode can reliably be activated above the level of other subnodes. This is presumably why it is not to the advantage of the system to create a subnode for each fact. There must be enough facts converging on the subnode with enough associative interconnections to guarantee the subnode a high level of activation in the presence of a related fact.

Practice

People get better at remembering facts by practicing them and it should come as no surprise that ACT predicts this basic fact of memory. However, the serious issue is whether the ACT theory can predict the shape of the improvement function and how this varies with factors such as fan.

Accumulation of Strength with Practice

ACT makes some fairly interesting predictions about the cumulative effects of extensive but widely distributed practice. By widely distributed I am referring to practice at intervals on the order of 24 hours. The reason for looking at such wide spacings is to avoid complications due to diminished effects of massed presentations. ACT does not really have an analysis of spacing effects except to note that it can implement the encoding variability explanation (see Anderson, 1976, for how). However, an assumption of the following analysis of practice is that each unit of practice is as effective as the next. By looking at widely spaced units this assumption becomes plausible. In the analyses to follow I will designate the cumulative impact of multiple massed practices per day as one unit of strength. Given this scaling of strength, I will be concerned with the accumulation of strength over the multiple days which are spaced.

With spaced repetitions the strength of a trace will just be a sum of the individual strengthenings. However, because of the delay between repetitions, it will not be simply a linear function of number of repetitions. We can assume that the early strengthenings have considerably decayed by the time of the nth strengthening. Assuming n spaced repetitions each t time units apart, the total strength of a trace after the nth strengthening, and just before the $n + 1$st, will be (by Eq. 2)

$$S = \sum_{i=1}^{n} s(it)^{-b} \qquad (19)$$

assuming the power law function for decay where s is the strength of each repetition and $b(<1)$ is the exponent of the power function. It can be shown (Anderson, 1982) that this sum is closely approximated as

$$S = d(n)^{+c} - a \qquad (20)$$

where $c = 1 - b$, $d = St^{-b}/(1 - b)$, and $a = (1 + b)/2(1 - b)\ st^{-b}$. Thus, strength approximately increases as a power function of time. As we will see, this power function prediction corresponds to a good deal of data. However, before presenting this data, it is necessary to consider the impact of this strength accumulation on spreading activation.

Effects of Extensive Practice

A set of experiments was conducted to test the prediction given earlier about a power law increase in the strength of a trace with practice. In these experiments, subjects were given extensive practice (i.e., hundreds of trials over many days). In one experiment subjects studied subject–verb–object sentences of the form *The lawyer hated the doctor*. After studying these sentences they were transferred to a sentence recognition paradigm in which they had to discriminate these sentences from foil sentences made of the same words as the target sentence but in new combinations. There were 25 days of tests. Each day subjects were tested on each sentence 12 times (in one group) or 24 times in the other group. There was no difference between these two groups (which is consistent with earlier remarks about massing of practice), so these two groups will be treated as one in the analysis.

There were two types of sentences—*no fan* sentences made from words that appeared in only one sentence and *fan* sentences made from words that appeared in two sentences. Figure 7 shows the change in reaction time with practice (number of days). The functions that are fit to the data in Figure 7 are of the form $T = K + AP^{-b}$ where K is an intercept not affected by strengthening, $K + A$ is the time on day 1, P is the amount of practice (measured in days), and the exponent b is the rate of improvement. It turns out that this data can be fit assuming different values of A for the fan and no fan and keeping K and b constant. The equations are

$$T = .36 + .77(P - ½)^{-.36} \quad \text{for no fan} \quad (21)$$
$$T = .36 + 1.15(P - ½)^{-.36} \quad \text{for fan.} \quad (22)$$

The value $P - ½$ appears in these equations as this is the average practice on a day P.

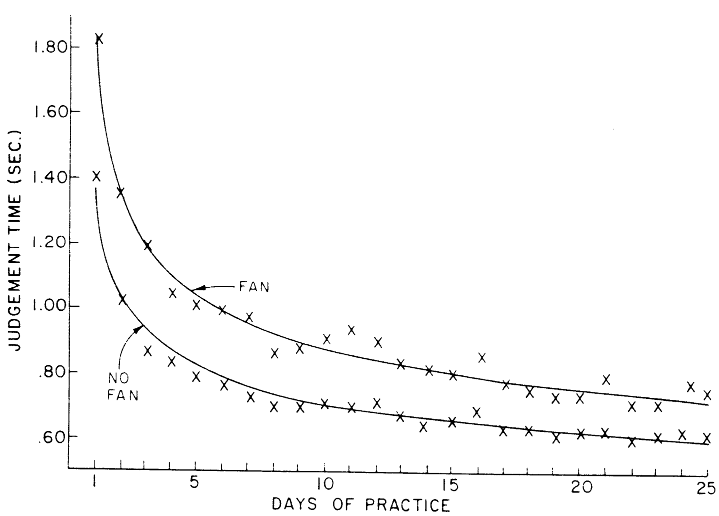


FIG. 7. Recognition times for fan and no fan sentences as a function of practice.

One implication of these equations is that the fan effect diminishes with practice. They also imply that the fan effect never disappears. After P days the fan effect is $.38(P - ½)^{-.36}$ according to these equations. Hayes-Roth (1977) reported data on practice from which she concluded that the fan effect disappeared after 10 days and 100 practice trials. However, this is not what these equations imply and Figure 7 shows that there still is a fan effect after 25 days and 600 trials. Perhaps the Hayes-Roth conclusion was a case of erroneously accepting the null hypothesis.

Earlier, I showed that strength increased as a power function of practice. Now I will show that this implies that reaction time should decrease as a power function. Recall that the amount of activation sent to a trace from a concept is a product of the activation emitted from the concept and the strength of that trace relative to competing traces. The activation emitted by a concept is a function of its strength. Let I be the prior strength of a concept. Then the strength of the concept after P days of practice will be

$$I' + AP^{+c} \tag{23}$$

where $I' = I - a$, $A = d$, and $P = n$ from Equation (20) earlier. In this we are assuming that the prior strength of I maintains a stable value over the experiment. Again drawing on an assumption about the ineffectiveness of mass practice, I will assume the strength of the concept to be equal for fan and no-fan concepts.

The relative strength of one of n experimental facts attached to a concept will be $1/n$ where I am making the assumption that subjects can completely filter out by a subnode structure any interference from preexperimental associations. This implies that the activation converging on the trace will be

$$\frac{3}{n}(I' + AP^{+c}) \tag{24}$$

where the 3 reflects the fact that activation is converging from 3 concepts (subject, verb, object).

According to the earlier retrieval assumptions, recognition time will be a function of the inverse of this quantity or

$$\frac{n}{3(I' + AP^{+c})} = \frac{nA'P^{-c}}{[I'/AP^{+c} + 1]} \tag{25}$$

where $A' = A/3$. To the extent that I, prior strength of the concept, is small relative to the impact of the massive experimental practice this function becomes

$$nA'P^{-c} \tag{26}$$

and total reaction time is predicted to be of the form

$$K_1 + A''P^{-c} \tag{27}$$

where K_1 is the intercept not affected by strength of traces of concepts and $A'' = nA' + K_2$. The quantity K_2 reflects that part of the improvement that is due to general practice and not retrieval of the fact, and, I assume, is improving at the same rate as memory retrieval. In general, even Equation (25) above, with a significant I factor, will yield a good fit to such a power function. Thus, we would expect the data in Figure 7 to be fit well by a power function. Note also that according to the ACT analysis, these functions should only vary in the parameter A'' and not in intercept K_1 or exponent c. The A'' will increase with n, the number of studied facts. As the reader can confirm from Figure 7, we get a good fit to the data based on these assumptions.

Interaction Between Practice and Prior Familiarity

One basic consequence of this increase in concept strength is that subjects can remember more facts about frequent concepts and retrieve facts of similar relative strength more rapidly. Anderson (1976) reported that subjects can retrieve facts about more familiar people (e.g., Ted Kennedy is a senator) more rapidly than facts about less familiar people (Birch Bayh is a senator—experiment done when B. B. was still a senator). Anderson (1981) noted that there are serious issues about whether pairs of facts like these are equated in terms of other properties. In that research report, I had subjects learn new facts about familiar or unfamiliar people and tried to control such things as degree of learning for these new facts. Still we found subjects at an advantage both in learning and retrieving new facts about the familiar person.

We recently performed an experiment in which we compared time to verify sentences studied in the experiment such as *Ted Kennedy is in New York* with other sentences studied such as *Bill Jones is in New Troy*. We found that subjects were initially more rapid at verifying the experimental facts about the familiar concepts, consistent with Anderson (1981). However, we also looked at the effects of fan and practice on these verification times. Figure 8 shows what happened to the effects of fan and familiarity over nine days of practice. As can be seen, the effects of fan largely maintained themselves over the period while the effects of familiarity diminished. This is what would be predicted on the basis of Equation (24). As practice P increases, the effect of prior familiarity I diminishes dramatically.

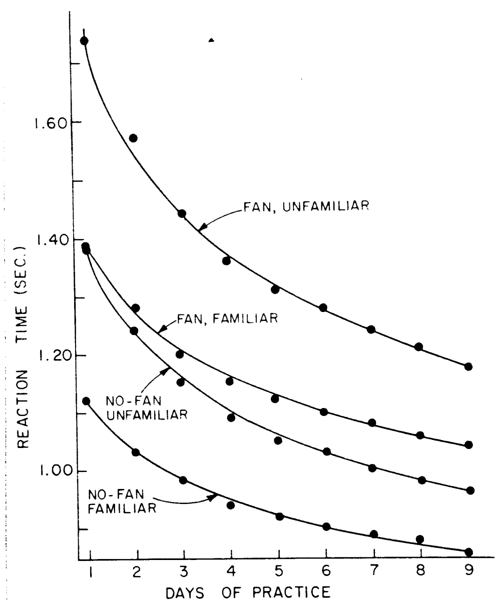


FIG. 8. Recognition of fan and no fan sentences about familiar and unfamiliar concepts as a function of practice.

The functions fit to the data are of the form $a + b/(I + P^{c})$ where a is the asymptote, b is the retrieval time parameter, I is prior strength and strength accumulated in original learning, P is the independent variable (number of days), and c is the exponent controlling growth of strength. The quantity P^c reflects the strength after P days. One value of a was estimated for all four conditions; this is .36 seconds. Separate values of b were estimated for the no fan (1.42 sec) and fan conditions (1.94 sec). Separate values of K were estimated for the familiar material (.88) and the unfamiliar material (.39). Finally a single parameter for c was estimated for all four conditions; this was .31.

On day 10, subjects were asked to learn some new facts of different form (e.g., Bill Jones hated the doctor) about the *old people* studied in the experiment plus some *new people* not yet studied. Some of the new people were familiar famous names and others were unfamiliar. After learning these new facts, the subjects went through one session of verification for these. There was no difference in the time they took to recognize the new facts about old familiar or new familiar people. In both cases, they took .96 seconds. They were still longer to recognize new facts about the old unfamiliar people (1.00 sec). Thus, the practice had not completely eliminated the differences between familiar and unfamiliar. However, they took longest to recognize facts about the new unfamiliar people (1.06 sec). So, the practice had increased the capacity of the unfamiliar nodes.

Summary

This section has reviewed the consequence of increased node strength with practice. Reaction times decrease and fan effects decrease for stronger nodes which can emit greater activation. However, the data reported here show that the basic pattern of results do not change.

Recognition versus Recall

The basic difference between the recall and the recognition paradigm is quite straightforward under the ACT analysis. In the recognition paradigm, parts of a trace are presented and the subject is asked whether he recognizes their combination. In the recall paradigm the subject is also asked to retrieve other components of the trace. In ACT, activation converges on the trace from all presented components. If there is sufficient activation the trace becomes available. In a recognition task the subject simply says that he recognizes the test probe. In a recall task the subject retrieves part of the trace according to task specifications. It is typical to present more of the trace in a recognition paradigm. For instance, in paired-associate learning both stimulus and response are typically presented in a recognition test but only the stimulus is presented in a recall test. However, it is possible to do a recognition test by simply presenting a stimulus. One would expect a high conditional probability between success at recognizing the stimulus and success at recalling the response, and indeed there is (Martin, 1967). Under the ACT analysis, recognition is typically better than recall because more of the trace is typically presented, not because of any inherent superiority of recall over recognition.

Paired-Associate Recognition

One of the interesting analyses of recognition versus recall is that of Wolford (1971). He showed that, correcting for guessing, recognition of a paired associate could be predicted by the probabilities of forward and backward recall. His model was basically that a paired associate could be recognized if the subject could retrieve the response from the stimulus or the stimulus from the response. Let P_f and P_b be these two probabilities of recall. On the assumption that the two directions of recall are independent, Wolford derived, for corrected recognition P_r, the equation

$$P_r = P_f + (1 - P_f)P_b. \tag{28}$$

Under the ACT theory, the subject is not viewed as performing two independent retrievals in recognition but rather converging activation from the two sources. This is an important way in which the current ACT differs from the earlier ACT (Anderson, 1976) and its predecessor, HAM (Anderson & Bower, 1973). Nonetheless, ACT predicts the relationship documented by Wolford. We may assume that an amount of activation A_S comes from the stimulus and an amount A_R from the response. This means that we have the following equations for forward recall, backward recall, and recognition

$$P_f = 1 - e^{-CA_s} \tag{29}$$
$$P_b = 1 - e^{-CA_R} \tag{30}$$
$$P_r = 1 - e^{-C(A_s+A_R)} \tag{31a}$$
$$= 1 - e(^{-CA_s})\,(e^{-CA_r}) \tag{31b}$$
$$= 1 - (1 - P_f)\,(1 - P_b) \tag{31c}$$
$$= P_f + (1 - P_f)P_b. \tag{31d}$$

The above analysis assumes that the probability of forming the trace is one and all failures of recall derive from failures of retrieval. If there is some probability of failing to encode the trace, the ACT analysis would predict that probabilities of forward and backward recall would overpredict probability of recognition. This is because forward and backward recall would no longer be independent. While Wolford found no evidence at all for nonindependence, other researchers (e.g., Wollen, Allison, & Lowry, 1969) have found evidence for a weak nonindependence.

Word Recognition

Another major domain where recall and recognition have been contrasted is memory for single words. A subject is presented with a list of such words and then must either recall (typically free recall) or recognize them. Recognition performance can be much higher than recall performance in such experiments. According to the framework set forth by Anderson (1972) and Anderson and Bower (1974), it is assumed that the subject forms traces linking the words to the various contextual elements. Although the contextual elements are undoubtedly more complex, the contextual factor is often represented as an association between the word and a single context element. This simplified situation is illustrated in Figure 9 where each line corresponds to a trace. Under this model, recognition involves retrieving the context from the word and verifying that it is indeed a list context. Direct recall involves retrieving list words from the list context. However, because of the high fan out of the context, the subject will have limited success at this. Thus, an auxiliary process involves using various strategies to generate words which can then be recognized. Because of this feature, this has been called the generate–recognize model of recall. For a review of relevant positive evidence see Anderson and Bower (1973) or Kintsch (1970). The major challenge to this analysis has come from various experiments showing contextual effects. These results will be reviewed in the next subsections.

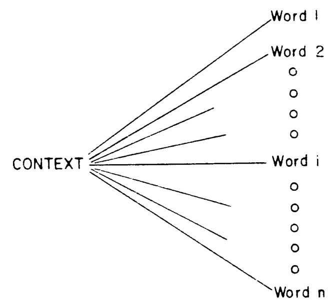


FIG. 9. Network representation of the word-context associations for a single list.

The basic assumptions of the generate–recognize model are consistent with the current framework. The current framework makes it clear that recognition is better than recall because of the difference in fan out of the context node versus word nodes. If the same word appeared in multiple contexts we might expect this situation to be reversed and, indeed, Anderson and Bower (1972, 1974) present evidence that recognition performance degrades as a word appears in multiple contexts.

Effects of Encoding Context

A large number of studies have been performed that display an effect of encoding context on both recognition and recall (e.g., Flexser & Tulving, 1978; Tulving & Thomson, 1971; Watkins & Tulving, 1975). These experiments are said to illustrate the *encoding specificity principle* that memory for an item is specific to the context in which it was studied. The experiment by Tulving and Thomson (1971) is a useful one to consider. They had subjects study items (e.g., black) either in isolation, in the presence of a strongly associated encoding cue (e.g., white), or in the presence of a weakly asso-iated encoding cue (e.g., train). The strong and weak cues were selected from association norms. Orthogonal with this variable, subjects were tested for recognition of the word in one of these same three contexts. Recognition was best when study context matched test context.

We have explained these results in terms of selection of word senses (Reder, Anderson, & Bjork, 1974; Anderson & Bower; 1974) or in terms of elaborative encodings (Anderson, 1976). These explanations still hold in the current ACT framework. Figure 10 illustrates the network structure assumed in this explanation for the case of study with a weak encoding cue. The basic idea is that a to-be-recalled word like *black* has multiple senses. In this case we have *black* illustrated with two senses, black1 and black2. Attached to *black1* is a weak associate *train* and attached to *black2* is a strong associate *white*. The oval nodes in Figure 10 are the traces encoding these as-

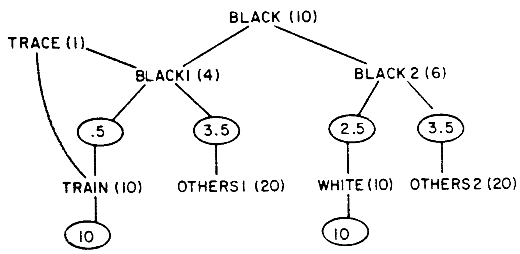


FIG. 10. A representation of the relevant network structure in the encoding specificity experiments. The numbers associated with the nodes are the strengths associated with the spreading activation analysis on p. ■■■.

sociations. The oval nodes in Figure 10 leading to others1 and others2 represent other unidentified associations. Similarly, the nodes at the bottom attached to *train* and *white* represent other unidentified associations. For simplicity, I have only represented the multiple senses attached to *black*.

At first blush, people often have the intuition that there is only one sense for a word like *black*. However, there are a number of distinct if similar senses. In the presence of *white*, one is likely to come up with senses of *black* that refer to a prototypical color or race of people. In the presence of *train* one is likely to come up with the sense associated with soot or the glistening black of a polished toy train.

I assume that the encoding context determines the sense of the word chosen and that a trace is formed involving that sense and, perhaps, the encoding context. When the subject is tested, context will again determine the sense chosen and activation will spread from the chosen sense. Probability of recognition will be greater when the same sense is chosen because activation will be spreading from a sense node directly attached to the trace.

It should be noted that the choice of a sense for the word can also be accomplished by means of spreading activation. That is, the sense of *black* chosen when train-black is presented is the one that receives the greatest activation from *train* and *black*. In Figure 10 this will be *black1* which lies at the intersection of train and black. Thus, there are two "waves" of activations. The first determines the sense of the words and the second spreads activation from the word senses to retrieve the trace. It is this same double activation process that is used in selecting a subnode.

Evidence for Multiple Sense Nodes

While one can explain the encoding specificity result assuming multiple senses as in Figure 10, it is not really necessary to assume multiple senses to explain the basic effect that recognition is higher if the study context is presented at test. To see this, note in Figure 10 that the study context, *train*, is associated to the trace. This means that there will be more activation converging at the trace at test if *train* is presented again, independent of any sense selection.[1] However, there are a number of additional results that indicate the need for the multi-sense-node explanation. One is that it has been found (Reder, et al. 1974; Tulving & Thomson, 1973; Watkins & Tulving, 1975) that recognition is worse in a test context that promotes selection of the wrong sense than a neutral test context that just fails to present any encoding word. For instance, after studying *train–black* subjects are worse in recognition of *black* in the context *white–black* than when *black* is presented alone. Thus, it is not just the absence of *train* that hurts recognition. It is the presence of another context that actively selects against the original interpretation.

[1] Past versions of the generate–test model assumed that there was no contribution of the context word to the retrieval of the trace except in terms of sense selection. However, in this model the context word can be just as important a source of activation as the target word. In fact, there are some results that suggest it might be a more important source (Bartling & Thompson, 1977; Rabinowitz, Mandler, & Barsalou, 1977). To the extent that it is the more important source, encoding specificity results should occur even for those words that truly have a single sense (Tulving & Watkins, 1977).

The multiple sense representation is important for understanding the results of Light and Carter-Sobell (1970). They had subjects study a pair like *raspberry–jam* with *jam* as the target and then tested subjects with *raspberry–jam*, or *strawberry–jam* which tapped the same sense of that word, or *log–jam* which tapped a different sense. They found the three conditions were in this order in terms of decreasing level of recall. Figure 1 shows a schematic of the memory representation for their experiment. I constructed sets of equations like those given in (14) to derive the patterns of network activation when various cues were presented to serve as sources of activation. The predictions were derived under the assumption that all links have equal relative strength where multiple links emanate from a node. I also assumed in these calculations that one unit of activation would spread both from the context word and from the selected sense of the target word. Presented with *raspberry* and *jam* were 1.65 units of activation accumulating at the trace. Cued with *strawberry* and *jam*, 1.26 units were accumulated at the trace. Finally, cued with *log* and *jam*, .29 units accumulated at the trace. This corresponds to the ordering found by Light and Cartel-Sobell. The difference between raspberry–jam versus strawberry–jam is a result of the fact that *raspberry* is directly connected to the trace. The difference between strawberry–jam versus log–jam is due to the difference between the two senses selected.[2]

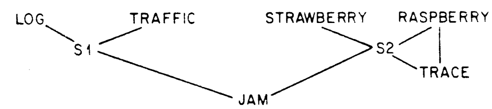


FIG. 11. A representation of the relevant memory structures for the Light and Carter-Sobell experiments.

Recognition Failure

Experiments such as those reported by Tulving and Thomson (1973) and by Watkins and Tulving (1975) are thought to be damaging for the generate–recognize models. In a typical example of these experiments, subjects study a word with a weak associate, are then asked to recognize it in the context of a strong associate, and are then asked to recall it in the context of the old weak associate. In these cases it is often found that recall is superior to recognition and that many words are not recognizable but can be recalled. The phenomenon of recallable words not being recognized is called *recognition failure* and it is sometimes interpreted as disproving generate–recognize models. This is because recognition is one of the subprocesses in the generate–recognize model of recall so recognition of a word is a precondition for its recall. However, the activation patterns set up in ACT for the strong-associate recognition condition are very different than the activation patterns set up in the weak-associate recall condition. Therefore, there is no reason to assume that the recognition should be predictive of recall. We would only expect recognition to be predictive of recall in the same context.

I used Figure 10 to calculate how much activation should converge on the trace in these two conditions. My calculations were performed under the assumption that active nodes (representing presented words) would send out one unit of activation each. Thus, when the target and the strong cue are presented, I assumed nodes *white* and *black2* would be active and send out one unit of activation each. In this case, .24 units of activation converge on the trace. When just *train* (the weak cue) is presented and activated, .88 units of activation are expected to converge. So in point of fact, the weak associate is a better prompt for the memory trace than is the target and strong cue. This is because *train* is directly con-

[2] Underwood & Humphreys (1977) have basically replicated the results of Light and Carter-Sobell, but they argue that the magnitude of the results do not justify the multiple-sense interpretation. It is hard to make clear predictions about the magnitude of the effect.

nected to the trace while *white* and *black2* are not. This analysis is consistent with the research and ideas of Rabinowitz, Mandler, and Barsalou (1977) and of Bartling and Thompson (1977) who showed that a considerable asymmetry existed between the forward recall of the target in response to the weak associate cue versus the opposite backward recall of the weak associate in response to the target cue with backward recall much lower. Rabinowitz et al. also noted a much reduced incidence of recognition failure when recognition was conditionalized on backward rather than forward recall.

One of the phenomena that has captured considerable attention (e.g., Flexser & Tulving, 1978; Wiseman & Tulving, 1975) is the relationship that occurs across experiments between probability of recognition conditional on recall $P(\text{Rn/Rc})$ and unconditional recognition $P(\text{Rn})$.

$$p(\text{Rn/Rc}) = p(\text{Rn}) + c[\text{p}(\text{Rn}) - p(\text{Rn})^2] \qquad (32)$$

where c has been estimated to be .5. This indicates that the probability of recognition conditional on recall is only marginally superior to the unconditional probability of recognition. Under the current framework we would predict that, if the trace has been formed, the probability of retrieving the trace from the cue is independent of the probability of retrieving it from the target. The reason there is not complete independence is that if there was no trace formed, there will be failure of recall in both cases. This is substantially the explanation of the function offered by Flexser and Tulving (1978).

Summary

According to the current ACT analysis, the difference between recognition and recall is one of the number and directness of the sources of activation. Context affects recall and recognition by providing sources of activation. Recognition will only be predictive of recall in the same context. It makes no more sense to talk about recognition conditional on recall when the contexts are different than it does to talk about conditional measure when the targets are changed. In different contexts different activation patterns will be set up just, as different patterns will be set up when the subject is asked to recognize different words. It is an interesting question to what extent differences in activation patterns instantiate what Tulving means by encoding specificity.

Elaborative Processing

One of the most potent manipulations that can be performed in terms of increasing a subject's memory for material is to have the subject elaborate on the to-be-remembered material (see Anderson & Reder, 1979; Anderson, 1980) for reviews. As Anderson and Reder argue, much of the research under the rubric of "depth of processing" (see Craik & Lockhart, 1972; Cermak & Craik, 1979, for a current survey) can be understood in terms of elaborative processing. That is, instructions which are said to promote "deeper processing" of the input can often be viewed as encouraging the subject to engage in more elaborative processing of the input. The phrase *elaborative processing*, though, is not much more technically precise than is the much lamented term *depth of processing* (Nelson, 1977). What I would like to do is spell out some of the ways elaboration can occur within this ACT framework. There are three basic ways that elaborations can improve recall. The first, and weakest, occurs when study elaborations serve to redirect activation away from interfering paths and towards the target path. (This is related to the concept of network integration discussed earlier.) The second occurs when subjects spread activation at test from additional concepts which were not in the probe but were part of the elaboration at study. Basically, this involves elaborating on the probe at test to try to generate additional concepts from which to spread activation. The third method involves using inferential methods to reconstruct from the elaborations that can be retrieved at test what the target trace must have been. I will go through each of these uses of elaboration in turn.

Redirecting Activation

To illustrate the first possibility, let us consider an example of elaboration given in Anderson (1976). One subject who was asked to memorize the paired associate *dog–chair* generated the following elaboration.

> The dog loved his masters. He also loved to sit on the chairs. His masters had a beautiful black velvet chair. One day he climbed on it. He left his white hairs all over the chair. His masters were upset by this. They scolded him.

Figure 12 illustrates this elaboration in approximate network form proposed in Anderson (1976). Note that the impact of this elaboration is to introduce multiple paths between *dog* and *chair*. There are two effects of this structure. First, it redirects activation that would go directly from *dog* to *chair* to other parts of the elaborative structure and this activation will only arrive at *chair* in a less direct and dissipated form. On the other hand, this activation is also being taken away from the prior facts. For example, activation is taken away from the prior associates to spread to *master* and some of the activation arriving at *master* spreads on to *dog*. Thus, the experimental fan out of *dog* somewhat dissipates direct activation of *chair* but somewhat redirects activation toward *chair*. We used the network in Figure 12 to see what the overall effect would be. Again we solved for asymptotic patterns of activation using equations such as (7). We assumed the total strength of the prior nodes attached to *dog* was nine and the strength of all experimental nodes was one. In this case, inputting one unit of activation at *dog*, .19 units arrived at *chair*. In contrast, when there was a single experimental path from *dog* to *chair*, only .12 units arrived at *chair*. Clearly, whether the elaborated structure will be better will depend on the specifics of the elaboration, but this example at least illustrates that it is possible that an elaborated structure can result in greater activation of the target trace.

A question that naturally arises about such elaborations is how the subject discriminates between target traces and elaborations. For instance, how is it that the subject knows it was the *dog–chair* that he studied and not *dog–master?* It is assumed that part of a trace is a tag indicating whether it is an encoding of a study event or part of a subject elaboration. Reder (personal communication), who had subjects explicitly generate elaborations of text, found her subjects very good at discriminating what they explicitly studied from what they generated as elaboration. However, to whatever extent subjects do lose these tags and to whatever extent they are willing to venture guesses in the absence of such tags, we would see inferential and semantic intrusions and false alarms.

In some experiments (e.g., Bransford, Barclay & Franks, 1972; Owens, Bower, & Black, 1979; Sulin & Dooling, 1974; Thorn-

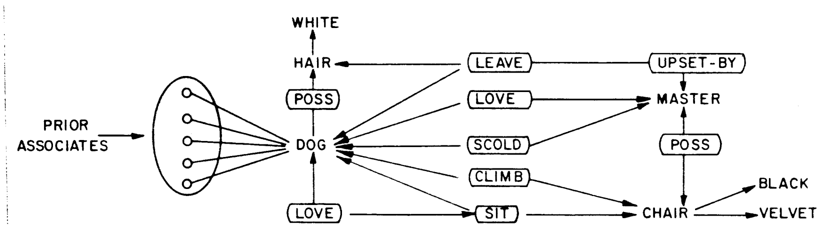


FIG. 12. Network representation of the elaborative structure generated to connect the pair *dog–chair*.

dyke, 1977), false acceptance of inferentially related foils is almost as high as acceptance of presented targets. I doubt that these very high false alarm rates can be attributed to loss of tags. This is because of Reder's evidence that subjects are quite good at distinguishing elaborations from studied material. Also, it seems improbable that subjects would have explicitly generated all these inferences. A better explanation of these intrusions and false alarms involves the notion of reconstructive recall which I will discuss later in this section.

People have commented on the superficial contradiction between the fan analysis (which claims that the greater the number of experimental paths leading from a concept, the poorer the memory) and this elaborative analysis (which says the greater the number of experimental paths the better the memory). To help clarify the situation I (Anderson, 1980) have coined the terms *irrelevant fan* and *relevant fan*. In the typical fan experiment, we are creating irrelevant fan in that the paths lead away from each other. In Figure 12 we have relevant fan in that the various paths leading from *dog* converge back on *chair*.

Elaborative Sources of Activation

There are other more powerful advantages to the structure in Figure 12 than its ability to direct more activation to the target. The subject can use any of the retrieved elaborations as additional sources for activating the target. So if the subject recalled his elaboration *The dog loved his master*, he can use *master* as another point from which to spread activation to the target structure. That is, the subject need not confine himself to spreading activation from the presented word. He can retrieve concepts used in the elaboration, focus on these, and spread activation from them.

Configural cueing. The research on configural cueing of sentence memory can be understood in terms of this elaborative analysis. Anderson and Bower (1972) had subjects study sentences such as

The minister hit the landlord

and then cued subjects for memory of the objects of the sentence with prompts such as

The minister ___ the _______	S cue
The _______ hit the _______	V cue
The minister hit the _______.	SV cue

When the instructions to the subject were just to study the sentences and when the subjects presumably studied the sentences passively, the experiments uncovered the relationship between recall to the three cues

$$P(\mathrm{SV}) \leq P(\mathrm{S}) + (1 - P(\mathrm{S}))P(\mathrm{V}). \quad (33)$$

On the other hand, when subjects were asked to generate meaningful continuations to the sentences, the relationship obtained was

$$P(\mathrm{SV}) > P(\mathrm{S}) + (1 - P(\mathrm{S}))P(\mathrm{V}). \quad (34)$$

Subsequent research by Foss & Harwood (1975), Jones (1980), and Anderson (1976) has confirmed that whether one gets the first or second relationship depends on how much subjects process the meaning of these sentences. The research caused a minor flap about whether or not there are configural cues in memory, but the basic results fall out quite neatly in the current framework and in such a way that the configural issue is blurred.

If subjects generate no elaborations, ACT predicts relationship (33) between probability of recall to the configural SV cue and the single-word S or V cue. Subject and verb contribute independent and additive activations to the trace. As we noted in the analysis of Wolford's recognition paradigm (p. ■■■) the effect of summing activation from cues C^1 and C^2 is to produce a level of activation that gives us the relation between probabilities of recall

$$P(C_1 \;\&\; C_2) = P(C_1) + (1 - P(C_1))P(C_2). \quad (35)$$

One can get the inequality in Equation (33) to the extent where a trace is not formed. That is, let $P^*(C)$ be the probability of reviving the trace from cue C conditional on the trace being formed and let $P(C)$ be the unconditional probability of trace retrieval. Let a be the probability of trace formation. Then

$$P(C_1) = aP^*(C_1) \tag{36}$$
$$P(C_2) = aP^*(C_2) \tag{37}$$
$$P(C_1 \,\&\, C_2) = a[P^*(C_1) + (1 - P^*(C_1))P^*(C_2)] \tag{38}$$
$$< aP^*(C_1) + (1 - aP^*(C_1))aP^*(C_2) \tag{39}$$
$$= P(C_1) + (1 - P(C_1))P(C_2). \tag{40}$$

The advantage of the subject and verb cue under meaningful processing instructions can be explained if these instructions cause the subject to process the sentence elaboratively. A schematic memory structure for such a sentence showing the effect of elaborations is illustrated in Figure 13. There is a trace interconnecting the subject (minister), verb (hit), and object (landlord). The assumption is that under meaningful processing instructions, the subject will retrieve one or more schemata for elaboration. So, he might retrieve *people hit people with hand-held objects*. This would lead to the elaboration that the minister hit the landlord with an object and that he held the object. This elaboration process can continue by retrieving another schema. Suppose the subject can recall a movie where a minister carried a wooden cross when confronting the devil. This would lead to the elaborations that the hitting instrument was a wooden cross and that the landlord was the devil. (Actually, one subject given this sentence did continue it *with a cross*—but I have no idea about the exact elaborative processes that led her to this continuation). Figure 13 indicates some of the impact of such elaborative activity by including *cross* in the trace for the sentence.

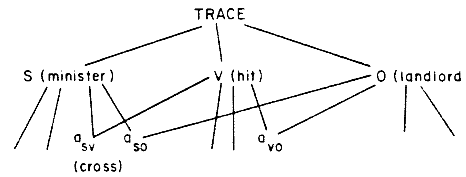


FIG. 13. Network representation of a subject–verb–object trace and the overlapping network of associations among the concepts. One of the intersecting concepts has been included in the trace.

Such elaborative encoding by itself will not produce configural effects. If activation spreads from S, or from V, or from both, the same prediction holds for the structure in Figure 13 about amount of activation of the trace and probability of recall, as when cross is not part of the trace. Basically, S and V make independent contributions to the overall pattern of activation. When presented together their contributions are just summed.

However, suppose that the subject at test also tried to elaborate with associates of the presented items and that he selected some associates of the words, and spread activation from these as well as from the presented word. In the presence of S or V alone he has a poor probability of regenerating the old associate. However, presented together there is a good probability that the old associate will be at the intersection of activation of these two words, have the highest activation, and so be selected. In that case, the subject could use this associate as an additional source of activation to converge on the trace and boost his recall over what we would expect on the basis of the single word cue.

The basic idea here is that, if the subject can recreate at test the elaborative activities he engaged in at study he can boost his recall. If the subject chooses as elaborations the most strongly activated associates of the probe, there is a chance that test associates will overlap with study associates. The probability and degree of overlap will vary with the number of terms in the test probe. This is basically the "semantic triangulation" explanation proposed by Anderson and Bower (1972) or the sense selection explanation proposed by Anderson (1976) where it is assumed that elaboration selects a particular word sense. Anderson and Bower (1973) showed heightened object recall conditional on a subject's ability to regenerate the old continuation and Anderson (1976) showed heightened object recall conditional on ability to recognize the sense of the subject or verb.

Configurally related sentence cues. A similar analysis can be made of the experiment of Anderson and Ortony (1975) who had subjects study sentences such as

A. Nurses are often beautiful.
B. Nurses have to be licensed.
C. Landscapes are often beautiful.

and cued subjects for recall with a term like *actress*. They point out that *nurses* has two interpretations, one as a female person and the other as a profession. Similarly, *beautiful* has two interpretations, one appropriate to women and one appropriate to landscapes. In the case of A, *actress* is appropriate to the selected senses of *nurse* and *beautiful*, but it is not in the case of B or C. They were interested in the relationships among the probability that *actress* evokes recall of sentence A, the probability of it evoking recall of sentence B, and the probability of it evoking recall of sentence C. Referring to these three probabilities as t, s, and p, they observed the following: $t > s + (1\text{-}s)p$. This was interpreted as contrary to associative theory.

Figure 14 shows the network schematics for situations A, B, and C. In each case, we have assumed that concepts at the intersection of subject and predicate are chosen for inclusion in the trace elaboration. As an example, for case A we have chosen *glamour* at the intersection of *nurse* and *beautiful*. The word *glamour* is also closely associated to *actress* and hence *actress* is closely associated to the trace. In contrast, the elaborations at the intersection of the subject and predicate in B and C are not closely associated to *actress* and hence *actress* is not closely associated to the trace. Thus, we would expect out of this associative analysis just what Anderson and Ortony found—much greater recall of the sentence with the *actress* cue in case A.

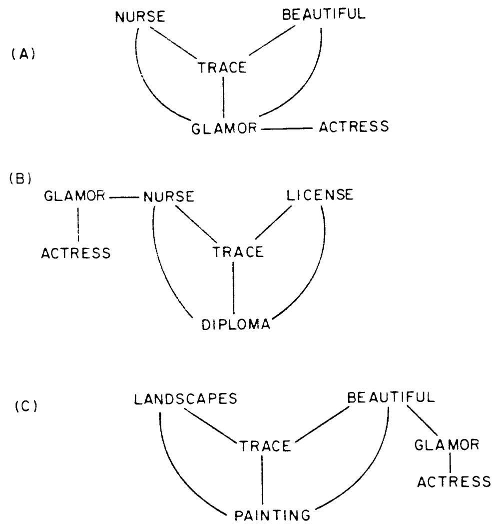


FIG. 14. Network representation for the three conditions of the Anderson and Ortony experiment.

Generation of Elaborations

An important issue about elaborations concerns how they are generated. Anderson (1976) proposed that they could be generated by inferential productions. I still believe this to be plausible in some circumstances, but a more important mechanism involves analogy to prior knowledge structures and this is the mechanism that I would like to develop here. With this mechanism it is possible to explain how memory elaboration and memory reconstruction can be combined to yield enhanced memory performance. The mechanism provides an explanation for Bartlett's (1932) observation that memory often depends critically on a match between the interpretive activities at study and the interpretive activities at test.

The details of this elaborative mechanism are described in Anderson (1983a), but the basic idea is that the subject has stored in long-term memory sets of events which can serve as analogs for elaborating the current event. I call these event sets *schemata*. For instance, the *chair* story in Figure 12 may have been based on one or more event sets in the subject's life. This elaboration notion can be made more powerful if the elaboration the subject generates is an interweaving of more than one event set. To consider a new example, suppose the subject is given as a to-be-remembered sentence

The janitor chased the cat.

and has in memory the schemata

A. The janitor found the rat in the basement.
   The janitor chased the rat.
   He killed the rat.
B. The dog chased the cat.
   He cornered the cat.
   The cat scratched him.

These schemata can be actual events in the subject's life (indeed, they are events in mine). Combined together we can construct the elaboration

The janitor found the cat in the basement.
The janitor chased the cat.
The janitor cornered the cat.
The cat scratched him.
The janitor killed the cat.

This elaboration is generated from the original memory by simple substitution of terms to achieve compatibility with the target trace. In the first case, *cat* is substituted for *rat*, and in the second case, *janitor* for *dog*. The schemata do not have to be encodings of specific events. They can also be more general event characterizations as envisioned in some of the current work on schemata.

Inferential Reconstruction

So far we have a mechanism for generating elaborations at study. This can be useful in numerous ways as already discussed. However, these schemata become much more powerful when we consider their potential use at recall. Suppose the subject was not able to remember the target sentence and was only able to remember two of his elaborations.

The janitor cornered the cat.
The janitor killed the cat.

The subject can apply the same elaborative process to these sentences as he applied to the original studied sentence. Suppose that the subject is able to reevoke the *dog chasing cat* schema from *The janitor cornered the cat* and generate an elaboration based on it. Then he might infer that the study sentence was one of the sentences he generated by elaborating with the schema, that is, either *The janitor chased the cat* or *The cat scratched the janitor*. In this situation the subject would have a fair probability of an inferential intrusion but also a fair probability of a correct recall. If the subject were also able to recall the rat-chasing schema and elaborate with it he would be in a stronger position. He could infer from these elaborations that the studied sentence might have been *The janitor chased the cat* or *The janitor found the cat in the basement*. Intersecting this with the elaborations from the other schema, the subject would have strong circumstantial evidence that what he had really seen was *The janitor chased the cat*.

The above example illustrates the basic processes underlying the phenomenon of *inferential redundancy*. At recall, the subject can use those study elaborations he can recall to infer what schemata he must have been using at study. He can then use these schemata to elaborate on the study elaborations and use the intersection of these test elaborations to infer what the original sentence had been. This is a very important role for prior knowledge.

As we noted above, the subject may have to select from a number of possible inferred candidates for the target trace. This can produce a high rate of false alarm and intrusion. However, the subject need not be at chance in choosing among the alternatives. He can try to recognize the alternative he generates. That is, when he spreads activation from *janitor* it may not have been enough to activate the trace. However, when he elaborates *The janitor chased the* *cat* as a candidate and spreads activation from *janitor*, *chased*, and *cat*, this may raise the activation level to the point where the original study trace is available.

The subject may also use statistical properties of the study material to reject various candidate sentences. Thus, he may know that none of the study sentences involved locations and so on this basis reject a sentence such as *The janitor found the cat in the basement*.

Experimental evidence. A clear prediction is that manipulations which increase the use of elaborations should increase both the level of recall and of intrusion. Experiments consistent with this prediction are reviewed by Anderson and Reder (1979). For instance, the prediction was nicely confirmed in the experiments of Owens et al. (1979) who found that providing subjects with thematic information increased both the level of recall and of intrusion. One of their examples involved having subjects study a story about a coed with or without the knowledge that she had an unwanted pregnancy. Presumably, the effect of this additional information is to increase the likelihood of evoking pregnancy-related schemata. So, when one of the target sentences read

> The doctor said, "My expectations are confirmed."

subjects could embellish this with their schema for a doctor telling a woman that she is pregnant. The effect of this embellishment would be to increase memory for the target sentence but also to increase the likelihood that subjects would intrude other elements from this schema such as

> The doctor told Nancy that she was pregnant.

We can understand the influence of the Owens et al. priming manipulation in terms of the spreading activation theory that we have developed. Recall that schemata are data structures in memory that have to be retrieved themselves before they can be used for embellishments. Thus, the statement

> The doctor said, "My expectations are confirmed."

would spread some activation to the doctor-tells-woman-she's-pregnant schema but not much. On the other hand, with the pregnancy concept already active, considerably more activation would converge on this schema, making its selection for elaboration more likely.

This same consideration can explain the effect of pretest instructions such as those given by Sulin and Dooling (1974). They told subjects only at test that a passage they had read about "Carol Harris" was really about Helen Keller. This produced a large increase in subjects' willingness to accept foils such as

> She was deaf, dumb, and blind.

The impact of identifying Carol Harris with Helen Keller should be to spread activation to the Helen Keller facts and thus strongly influence the elaborations that subjects generate at test. As a consequence, their elaborations will contain the candidate sentence which may then be falsely accepted.

This idea, that contextual information can influence the elaboration process, explains the powerful interaction that we obtained in Schustack and Anderson (1979). There we presented subjects with 42 short historical passages for study and then tested their recognition memory for sentences taken from these passages. We provided subjects with analogies for the referents of these passages either at study or at test. For instance, it might be pointed out to the subjects that the passage about Yoshida Ichiro had strong analogies to what they knew about Lyndon Johnson. We found that subjects' memory for this material was greatly improved but only if the same analogy was presented at study as at test. Thus, it was important that subjects' study elaborations and test elaborations both take advantage of information to be found with the Lyndon Johnson schema.

The notion of fluctuations in schema availability also helps to explain an interesting result reported by Bower, Black, and Turner (1979). They had subjects study 1, 2, or 3 stories that constituted variations on the same general theme. For instance, stories could be about visiting a doctor, visiting a dentist, or visiting a chiropractor. Bower et al. found an increased tendency, the more stories the subject had studied, to intrude theme-related inferences and to false alarm to such inferences. With each story studied, the subject would have another opportunity to rehearse and strengthen the elaborative schemata. He would also have in memory these earlier stories which could themselves serve as elaborative schema. Therefore, these schemata for making theme-related inferences would be more available the more stories that had been studied.

Summary

This section has considered the various ways activation influences the generation of elaborations and their use. It was argued that spread of activation determined which schemata were chosen for elaboration both at study and at test. Elaborations can influence recall by redirecting activation towards the to-be-recalled material, providing additional sources of activation, and providing a means for reconstructing what had been studied. Many of the interesting interactions in memory can be understood as depending on activation selecting the right schemata for elaboration.

Conclusions

We have reviewed some of the memory phenomena that can be understood in terms of the spreading activation mechanism of the current ACT theory (Anderson, 1976). Many of these phenomena had been thought to be difficult for the old ACT (Anderson, 1976) or its predecessor, HAM (Anderson, 1983a). It is worth noting the features of the new ACT that enable it to deal with these formerly troublesome phenomena. Most fundamentally, activation is a continuously varying quantity. This enables it to produce more subtle behavior than was possible with the all-or-none mechanisms of the former ACT or HAM. Because activation can sum and varies with associative distance and strength, level of activation of a node is sensitive to the particular configuration of activation sources. This enables ACT to produce some of the phenomena in the memory literature that depend on the context of study and test.

(Received February 24, 1982)

References

Anderson, J. R. FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation*. New York: Academic Press, 1972. Vol. 5.

Anderson, J. R. *Language, Memory, and Thought*. Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. R. Concepts, propositions, and schemata: What are the cognitive units? *Nebraska Symposium of Motivation*, 1980, **28**, 121–162. (a)

Anderson, J. R. *Cognitive Psychology and its Implications*. San Francisco: Freeman, 1980. (b)

Anderson, J. R. Interference: The relationship between response latency and response accuracy. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, **7**, 326–343.

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

Anderson, J. R. *The Architecture of Cognition*. Cambridge, Mass.: Harvard Univ. Press, 1983. (a)

Anderson, J. R. Representational types: A tricode proposal. *Cognitive Psychology*, submitted. (b)

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

Anderson, J. R., & Bower, G. H. Configural properties in sentence memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 594–605. (b)

Anderson, J. R., & Bower, G. H. *Human Associative Memory*. Washington, D.C.: Winston, 1973.

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

Anderson, J. R., & Bower, G. H. *Human Associative Memory: A Brief Edition*. Hillsdale, N.J.: Erlbaum, 1980.

Anderson, J. R., & Paulson, R. Representation and retention of verbatim information. *Journal of Verbal Learning and Verbal Behavior*, 1977, **16**, 439–451.

Anderson, J. R., & Reder, L. M. An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of*

Processing in Human Memory. Hillsdale, N.J.: Erlbaum, 1979.

ANDERSON, R. C., & ORTONY, A. On putting apples into bottles—A problem of polysemy. *Cognitive Psychology*, 1975, **7**, 167–180.

BARTLETT, F. C. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge Univ. Press, 1932.

BARTLING, C. A., & THOMPSON, C. P. Encoding specificity: Retrieval asymmetry in the recognition failure paradigm. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 690–700.

BRADSHAW, G. L., & ANDERSON, J. R. Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior*, 1982, **21**, 165–174.

BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 1972, **3**, 193–209.

BOWER, G. H., BLACK, J. B., & TURNER, T. J. Scripts in memory for text. *Cognitive Psychology*, 1979, **11**, 177–220.

BROADBENT, D. E. *Perception and Communication*. Elmsford, N.Y.: Pergamon, 1958.

CERASO, J., & HENDERSON, A. Unavailability and associative loss in RI and PI. *Journal of Experimental Psychology*, 1965, **70**, 300–305.

CERMAK, L. S., & CRAIK, F. I. M. *Levels of Processing in Human Memory*. Hillsdale, N.J.: Erlbaum, 1979.

COLLINS, A. M., & LOFTUS, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, **82**, 407–428.

COLLINS, A. M., & QUILLIAN, M. R. Experiments on semantic memory and language comprehension. In L. Gregg (Ed.), *Cognition and Learning*. New York: Wiley, 1972.

CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*. 1972, **11**, 671–684.

CRAIK, F. I. M., & WATKINS, M. J. The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 599–607.

ECCLES, J. C. Possible synaptic mechanisms subserving learning. In A. G. Karyman & J. C. Eccles (Eds.), *Brain and Human Behavior*. New York: Springer-Verlag, 1972.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychological Review*, 1960, **67**, 207–223.

ESTES, W. K. An associative basis for coding and organization. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington: Winston, 1972.

FISCHLER, I. Semantic facilitation without association in a lexical decision task. *Memory and Cognition*, 1977, **5**, 335–339.

FLEXSER, A. J., & TULVING, E. Retrieval independence in recognition and recall. *Psychological Review*, 1978, **85**, 153–172.

FOSS, D. J., & HARWOOD, D. A. Memory for sentences: Implications for human associative memory. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 1–16.

GLUCKSBERG, S., & MCCLOSKEY, M. Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, **7**, 311–325.

GOETZ, E. T., ANDERSON, R. C., & SCHALLERT, D. L. The representation of sentences in memory. *Journal of Verbal Learning and Verbal Behavior*, 1981, **20**, 369–385.

GOSS, A. E. Manifest strengthening of correct responses of paired-associates under postcriterion zero percent occurrence of response members. *Journal of General Psychology*, 1965, **72**, 135–144.

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

GREENBERG, R., & UNDERWOOD, B. J. Retention as a function of stage of practice. *Journal of Experimental Psychology*, 1950, **40**, 452–457.

HOROWITZ, L. M., & NEWMAN, W. An interrupted stimulus can facilitate PA learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 219–224.

HOUSTON, J. P. Proactive inhibition and undetected retention interval rehearsal. *Journal of Experimental Psychology*, 1969, **82**, 511–514.

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

JONES, G. V. Interaction of intrinsic and extrinsic knowledge in sentence recall. *Attention and Performance VIII*. Hillsdale, N.J.: Erlbaum, 1980.

JONES, J. E. All-or-none versus incremental learning. *Psychological Review*, 1962, **69**, 156–160.

KEPPEL, G., POSTMAN, L., & ZAVORTINK, B. Studies of learning to learn: VIII. The influence of massive amounts of training upon the learning and retention of paired-associate tests. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 790–796.

KING, D. R. W., & ANDERSON, J. R. Long term memory search: An intersecting activation process. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 587–605.

KINTSCH, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of Human Memory*, New York: Academic Press, 1970. Pp. 307–373.

KINTSCH, W. *The Representation of Meaning in Memory*. Hillsdale, N.J.: Erlbaum, 1974.

KOPPENAAL, R. J. Time changes in the strengths of A–B, A–C lists: Spontaneous recovery? *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 310–319.

LEWIS, C. H., & ANDERSON, J. R. Interference with real world knowledge. *Cognitive Psychology*, 1976, **8**, 311–335.

LIGHT, L. L., & CARTER-SOBELL, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1–11.

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

MCCLOSKEY, M., & BIGLER, K. Focused memory search in fact retrieval. *Memory & Cognition*, 1980, **8**, 253–264.

MCKOON, G., & RATCLIFF, R. Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1979, **18**, 463–480.

MANDLER, G. Organization and memory. In K. W. Spence & J. A. Spence (Eds.), *The Psychology of Learning and Motivation*. New York: Academic Press, 1967, 328–372. Vol. 1.

MARTIN, E. Relation between stimulus recognition and paired-associate learning. *Journal of Experimental Psychology*, 1967, **74**, 500–505.

MELTON, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 1–21.

MEYER, D., & SCHVANEVELDT, R. Meaning, memory, and mental processes. *Science*, 1976, **192**, 27–33.

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

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

NELSON, T. O. Reinforcement and human memory. In W. K. Estes (Ed.), *Handbook of Learning and Cognitive Processes*. Hillsdale, N.J.: Erlbaum, 1976. Vol. 3.

NELSON, T. O. Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, **16**, 151–172.

OWENS, J., BOWER, G. H., & BLACK, J. B. The "soap opera" effect in story recall. *Memory and Cognition*, 1979, **7**, 185–191.

POSTMAN, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of Human Learning*. New York: Academic Press, 1964.

POSTMAN, L., & GRAY, W. Maintenance of prior associations and proactive inhibition. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 255–263.

POSTMAN, L., & KEPPEL, G. Conditions of cumulative proactive inhibition. *Journal of Experimental Psychology: General*, 1977, **106**, 376–403.

POSTMAN, L., & PARKER, J. F. Maintenance of first-list associations during transfer. *American Journal of Psychology*, 1970, **83**, 171–188.

POSTMAN, L., STARK, K., & BURNS, S. Sources of proactive inhibition on unpaired tests of retention. *American Journal of Psychology*, 1974, **87**, 33–56.

POSTMAN, L., STARK, K., & FRASER, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 672–694.

RABINOWITZ, J. C., MANDLER, G., & BARSALOU, L. W. Recognition failure: Another case of retrieval failure. *Journal of Verbal Learning and Verbal Behavior*, 1977, **16**, 639–663.

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

REDER, L. M., & ANDERSON, J. R. A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology*, 1980, **12**, 447–472.

REDER, L. M., & ROSS, B. H. Integrated knowledge in different tasks: Positive and negative fan effects. *Journal of Experimental Psychology: Human Learning and Memory*, in press.

REDER, L. M., ANDERSON, J. R., & BJORK, R. A. A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, 1974, **102**, 648–656.

SHOBEN, E. J., WESCOURT, K. T., & SMITH, E. E. Sentence verification, sentence recognition, and the semantic-episodic distinction. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, **4**, 304–317.

SHIFFRIN, R. M. Short-Term Store: The basis for a memory system. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive Theory*. Hillsdale, N.J.: Erlbaum, 1975. Vol. 1.

SIMON, H. A. How big is a chunk? *Science*, 1974, **183**, 482–488.

SMITH, E. E., ADAMS, N., & SCHORR, D. Fact retrieval and the paradox of interference. *Cognitive Psychology*, 1978, **10**, 438–464.

SCHUSTACK, M. W., & ANDERSON, J. R. Effects of analogy to prior knowledge on memory for new information. *Journal of Verbal Learning and Verbal Behavior*, 1979, **18**, 565–583.

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

THORNDYKE, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, **9**, 77–110.

TULVING, E., & THOMSON, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

TULVING, E., & THOMSON, P. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352–373.

UNDERWOOD, B. J. Interference and forgetting. *Psychological Review*, 1957, **64**, 49–60.

UNDERWOOD, B. J., & EKSTRAND, B. R. Studies of Distributed Practice: XXIV Differentiation and Proactive Inhibition. *Journal of Experimental Psychology*, 1967, **74**, 574–580.

UNDERWOOD, B. J., & KEPPEL, G. One trial learning? *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 1–13.

WARREN, R. E. Stimulus encoding and memory. *Journal of Experimental Psychology*, 1972, **94**, 90–100.

WARREN, R. E. Time and the spread of activation in memory. *Journal of Experimental Psychology: Learning and Memory*, 1977, **3**, 458–466.

WATKINS, M. J., & TULVING, E. Episodic memory: When recognition fails. *Journal of Experimental Psychology: General*, 1975, **104**, 5–29.

WICKENS, D. D., MOODY, & DOW, R. The nature of timing of the retrieval process and of interference effects. *Journal of Experimental Psychology: General*, 1981, **110**, 1–20.

WICKELGREN, W. A. The long and the short of memory. *Psychological Bulletin*, 1973, **80**, 425–438.

WICKELGREN, W. A. Single-trace fragility theory of memory dynamics. *Memory and Cognition*, 1974, **2**, 775–780.

WICKELGREN, W. A. Memory storage dynamics. In W. K. Estes (Ed.), *Handbook of Learning and Cognitive Processes*, Hillsdale, N.J.: Erlbaum, 1976. Vol. 4 (a)

WICKELGREN, W. A. Network strength theory of storage and retrieval dynamics. *Psychological Review*, 1976, **83**, 466–478. (b)

WICKELGREN, W. A. Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review*, 1979, **86**, 44–60.

WISEMAN, S., & TULVING, E. Encoding specificity: Relation between recall superiority and recognition failure. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 349–361.

WOLFORD, G. Function of distinct associations for paired-associate performance. *Psychological Review*, 1971, **73**, 303–313.

WOLLEN, K., ALLISON, T., & LOWRY, D. Associative symmetry versus independent association. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 283–288.

WOODWARD, A. E., BJORK, R. A., & JONGEWARD, R. H. Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 608–617.