Psychological Review

A Theory for the Storage and Retrieval of Item and Associative Information

Bennet B. Murdock, Jr.
University of Toronto, Toronto, Ontario, Canada

A theory for the storage and retrieval of item and associative information is presented. In the theory, items or events are represented as random vectors. Convolution is used as the storage operation, and correlation is used as the retrieval operation. A distributed-memory system is assumed; all information is stored in a common memory vector. The theory applies to both recognition and recall and covers both accuracy and latency. Noise in the decision stage necessitates a two-criterion decision system, and over time the criteria converge until a decision is reached. Performance is predicted from the moments (expectation and variance) of the similarity distributions, and these can be derived from the theory. Several alternative models with varying degrees of distributed memory are considered, and expressions for signal-to-noise ratio and relative efficiency are derived.

The nature of associations is a classic problem in the area of human learning and memory. According to the traditional view, ideas (events, items) exist separately in memory but are somehow connected or related (see, e.g., Anderson & Bower, 1973). The labeled paths of current network models are an elaboration of these simple connections, but the essential idea is the same. The two items entering into an association are stored separately, and the association is the link between them.

A basic problem with such models is the search problem. How is the stored information located? Search of an associative network is easy to imagine but hard to describe precisely. Alternatives to connectionistic models are distributed-memory models (see, e.g., Hinton & Anderson, 1981). They have suggested a radically different possibility. Associative memory could occur without any localized storage of individual items at all. This more wholistic view is reminiscent of Gestalt notions (Asch, 1969) and has also been suggested by recent work on Markov models (Greeno, James, DaPolito, & Polson, 1978). Primarily, however, it is most closely allied with holographic models of memory (Cavanagh, 1976; Pribram, Nuwer, & Baron, 1974; van Heerden, 1963).

Distributed-memory models have powerful mechanisms but have not yet been applied to the extensive data already available in the literature. What is needed is a distributed-memory model that generates testable predictions about experimental results. That is my intention here. It is limited in that it applies only to rather simple phenomena of

recognition and recall. If it proves fruitful, then it certainly can be extended.

Following the content-addressable distributed associative memory (CADAM) model of Liepa (Note 1), I suggested a particular version of an associative mechanism based on convolution and correlation (Murdock, 1979). Eich (1982) has applied these concepts to a paired-associate situation, and her work illustrates what can be done with this approach. In this paper I present a fairly detailed model of associative memory based on convolution and correlation. Eich (1982) has done this using computer simulation to generate the predictions. I would like to buttress her work by using a more analytic approach.

Any general theory of memory must specify at least four things: how information is represented, the type of information that is stored and retrieved, the nature of the storage and retrieval operations, and the format of the store. I suggest that information is represented by random vectors, item information and associative information are stored and retrieved, the storage and retrieval operations are convolution and correlation, and memory storage is distributed or composite, not discrete or compartmentalized. Evidence for the distinction between item (occurrence) and associative (relational) information may be found in Murdock (1974). A complete theory must encompass serial-order information as well, but that is much more complex.

Theory

Items or events can be represented as vectors of attributes (Bower, 1967; Horowitz & Manelis, 1972; Kintsch, 1970; Underwood, 1969; Wickens, 1972). Item information and associative information are both stored in a common memory vector **M**. For item information, each item vector is simply

added to the memory vector. For associative information, two separate item vectors are first convolved; then the convolution is added to the memory vector. Specifically,

$$\mathbf{M}_j = \alpha\mathbf{M}_{j-1} + \gamma_1\mathbf{f}_j + \gamma_2\mathbf{g}_j + \gamma_3(\mathbf{f}_j*\mathbf{g}_j), \quad (1)$$

where **f** and **g** are the two members of a pair (names and faces, graphemes and phonemes, or the A–B pairs in a list of paired associates) and **f**∗**g** is the convolution of the two separate items. The Greek letters denote parameters: α is the forgetting parameter, γ_1 and γ_2 are the weighting parameters for item information (A and B items, respectively), and γ_3 is the weighting parameter for associative information. These parameters can vary over the range 0–1.

As each new pair of items is presented, the items are associated, or convolved, and both item and associative information are added to the memory vector as described by Equation 1. To illustrate, suppose three successive pairs A–B, C–D, and E–F are presented. Then the contents of memory after the jth pair has been presented are shown in Table 1. The symbolism here uses A, B, and C to obviate the need for subscripts in the **f** and **g** notation. Also, the weighting parameters (γ_1, γ_2, and γ_3) have been set to 1 to simplify the table. The table may suggest that the memory vector gets bigger as list presentation proceeds, but that is not the case. There are simply more entries summed.

A different representation is shown in Figure 1. The items are considered to be three-element vectors, with the elements represented by dashes. As will be explained, their convolution is a five-element vector with nine components. The middle element is actually the sum of three components, but it is represented by three dashes rather than one to show these components. The two flanking elements are each the sum of two components, again represented by dashes, whereas

Table 1

Composition of the Memory Vector $\mathbf{M}_j$ After Presentation of the jth Pair

j	Pair	$\mathbf{M}_j$
1	A–B	A + B + A∗B
2	C–D	C + D + C∗D + α(A + B + A∗B)
3	E–F	E + F + E∗F + α[C + D + C∗D + α(A + B + A∗B)]

the outside elements each have only a single component. This representation highlights the fact that all inside elements of a convolution are sums, a fact that must be kept in mind when deriving expressions for means and variances. It also shows that item information is based on simple summation, whereas for associative information the two items are first convolved and then added to the memory vector.

At the end of list presentation, the memory vector **M** consists of a single five-item vector whose contents are the sum of the six items (A, B, ..., F) and three pairwise convolutions (A*B, C*D, and E*F). So, it is partly a sum of items and partly a sum of sums. There is nothing in memory corresponding to any single item. There is only a single composite trace (the memory vector) where everything has been added together.

A good metaphor would be to imagine what the waves on the surface of a pond might look like after various objects had been dropped in the pond. Say the objects were an automobile tire, a beer bottle, the kitchen sink, and a telephone. The wave action on the surface of the pond shows traces of all but is specific to none. However, by the
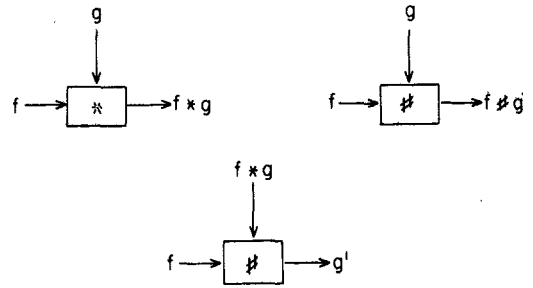

Figure 2. Black-box account of convolution (*) and correlation (#). The inputs are the vectors **f** and **g**, and the output is their convolution (**f*g**) or their correlation (**f#g**).

proper comparison process, one could determine that the telephone had been thrown in but a pair of skis had not. Further, one could "retrieve" beer bottle rather than kitchen table, telephone, or skis given tire as the cue for recall. However, what is retrieved is not the object itself but information sufficient to support veridical recall or recognition.[1]

The storage operation is convolution, but the retrieval operation is correlation. Convolution and correlation are treated more formally in the next section. For now, think of them simply as operations that can be performed on vectors. A black-box approach is shown in Figure 2. In the upper left (convolution), if the inputs to the black box are the vectors **f** and **g**, the output is the convolution **f*g**. In the upper right (correlation), with the same inputs the output is their correlation, **f#g**. As shown at the bottom, if one input to the correlator is the vector **f** and the other input is the vector **f*g**, the output will be **g′**, where **g′** is a vector similar to **g**. This is the retrieval mechanism for cued recall.

Think of **g′** as the retrieved information and **g** as the target item. Whether **g** will be recalled depends on the similarity of **g′** to **g**. Similarity will be measured by the dot product (**g′·g**) of the two vectors. Over some number of trials, there will be a distribution of similarity values, and the mean and variance of this similarity distribution can be computed as a function of the parameters of the model (*N*, the number of elements in

Figure 1. Memory representation for a three-pair list. (The items A, B, C, D, E, and F are represented by rows of dashes. The pairwise convolutions—A*B, C*D, and E*F—are represented by dashes in a diamond shape.)

---

[1] I would like to thank Jim Hinrichs for a suggestion that considerably improved this metaphor.
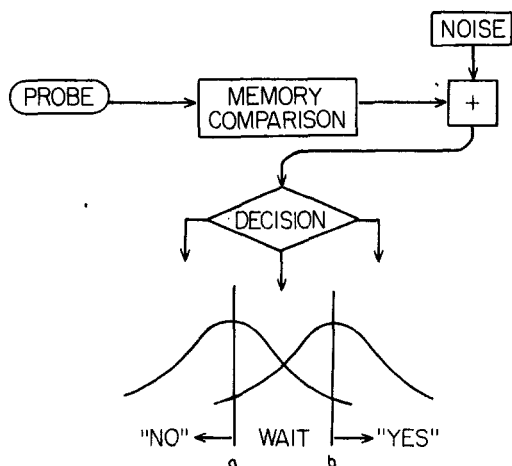
Figure 3. Decision system for recognition. (The input to the decision system is the output from the memory comparison plus uncorrelated [independent] noise. If the sum of these is above an upper criterion [b], the response is "yes." If the sum of these is below a lower criterion [a], the response is "no." If the sum is between Criteria a and b, the system waits because the noise is randomly varying over time.)

each item vector; α, the forgetting parameter; and γ_1–γ_3, the weighting constants for item and associative information).

If the retrieved information were exactly g (i.e., if $g' = g$), then the similarity of the retrieved information to the target information would be P, the power of the vector ($P = f \cdot f$). Recall probability depends on two factors. The first factor is how close the dot product ($g' \cdot g$) is to P. The second factor is whether any other item (e.g., h) is closer to P than g' is. If in fact g' were more similar to h than to g, then an intrusion might occur. However, there are probably error-detection capabilities. If the item information about h were too different from the item information about the probe f, then perhaps h could be rejected in favor of g.

Item recognition is much simpler. The probe item is compared to the memory vector by taking the dot product of the two, and the result is fed into the decision system. The time taken by the decision system will be one of the factors determining reaction time. The main feature of the decision system is that it is designed to cope with a noisy signal. Not only is there noise arising from the distributed memory but there is also extraneous random noise added to the output from the memory system. This is illustrated in Figure 3.

In a recognition task there are three options: "no," "yes," and wait. The function of the third option (wait) is to trade time for accuracy. Noise eventually may put a wait-listed outcome in one of the two action regions. That is, the result of the comparison process stays constant but the noise continually varies, so on a moment-by-moment basis, the input to the decision system also varies. Initial observations near the lower criterion are more likely to become "no" (i.e., fluctuate below Criterion a rather than above Criterion b), whereas initial observations near the upper criterion are more likely to become "yes" (i.e., fluctuate above Criterion b rather than below Criterion a). As soon as either one happens, a response is assumed to occur. Depending on the distance between the two criteria and the variance of the random noise, the wait may be short or long.

Under some cases, observations may continue to accrue in the middle region without any decision being reached. Some stop rule is necessary. If a "don't know" response is allowed, that may end it. It is assumed that over time the criteria come together. After each uncertain observation, the distance between the two criteria is reduced by a constant fraction. Thus, eventually, some response must occur. In the limit, the distance between Criterion a and Criterion b is zero. Eventually, some response must be made.

For associative recognition (Is A–B a correct pairing? Is A–D a correct pairing?), the retrieval process goes as follows: The two items (A and B, or A and D) are convolved, and the convolution is compared to the memory vector in the usual way (dot product). Then this dot product is fed into the decision system, which returns a yes or no response. Principles determining latency are as described above.

I have now presented a fairly complete overview of the theory. To be useful, however, derivations based on detailed assumptions are necessary. These are given in the next section. The reader who is more interested in the general characteristics of this approach should skip the derivations and proceed directly to the section on application.

Derivations

Items may be represented as random vectors. That is, as a vector whose elements are random variables. (For a discussion of random variables, see, e.g., Feller, 1968.) There are N elements in the vector, N being one of the parameters of the model.

To say an item is a random vector is to say the item has a random value on each of its attributes. These values are random in the sense that the suit (clubs, diamonds, hearts, and spades) and rank (two, three, . . . , queen, king, ace) of a card one draws from a deck of cards is random. Suit and rank are unpredictable card by card and uncorrelated. However, the jack of diamonds, say, always has the same suit and the same rank whenever it is drawn. A given item (at least in the same context) always has the same value on each dimension, but the values are random (independent or uncorrelated) both within and across items.

As in Anderson (1973), each element in the vector is assumed to be a random sample from a normal distribution. This distribution has mean $\mu = 0$ and variance $\sigma^2 = P/N$, where P is the power of the vector. In general, we assume unit vectors so $P = 1$, but the derivations will show P explicitly so other assumptions could be investigated.

The vector $\mathbf{f}$ can be represented as the doubly infinite vector

$$\mathbf{f} = (\cdots, f_{-2}, f_{-1}, f_0, f_1, f_2, \cdots).$$

The vector is centered at zero, and zeros flank the N nonzero feature values. The subscripts range from $-(N-1)/2$ to $(N-1)/2$, where N is odd.

For two vectors $\mathbf{f}$ and $\mathbf{g}$, their convolution is

$$(\mathbf{f}*\mathbf{g})_x = \sum_i f(i)g(x-i) = \sum_i f(x-i)g(i) \quad (2)$$

and their correlation is

$$(\mathbf{f}\#\mathbf{g})_x = \sum_i f(i)g(x+i) = \sum_i f(i-x)g(i) \quad (3)$$

for element x in the convolution or correlation. For convolution, the sum of the arguments is x, whereas for correlation the difference (second minus first) is x.

There are innumerable accounts of convolution and correlation in the scientific literature; a particularly detailed and well-illustrated account is given in Bracewell (1978). For accounts applicable to memory, see Eich (1982), Metcalfe and Murdock (1981), or Murdock (1979).

Convolution is commutative ($\mathbf{f}*\mathbf{g} = \mathbf{g}*\mathbf{f}$) whereas correlation is not ($\mathbf{f}\#\mathbf{g} \neq \mathbf{g}\#\mathbf{f}$). Some useful properties are

$$\mathbf{f}\#(\mathbf{f}*\mathbf{g}) = \mathbf{g}'$$

$$\mathbf{g}\#(\mathbf{f}*\mathbf{g}) = \mathbf{f}'$$

$$\mathbf{f}*(\mathbf{f}\#\mathbf{g}) = \mathbf{g}'$$

$$\mathbf{g}*(\mathbf{g}\#\mathbf{f}) = \mathbf{f}', \quad (4)$$

where, as above, $\mathbf{f}'$ and $\mathbf{g}'$ are similar to $\mathbf{f}$ and $\mathbf{g}$. The similarity of any two vectors ($\mathbf{f}$ and $\mathbf{g}$, or $\mathbf{f}$ and $\mathbf{f}'$) is here defined as their dot product.

For convolution followed by correlation,

$$(\mathbf{f}\#(\mathbf{f}*\mathbf{g})_w)_x = (\mathbf{f}\# \sum_{i'} f(w-i')g(i'))_x$$

$$= \sum_i f(i) \sum_{i'} f(x+i-i')g(i')$$

$$= \sum_i \sum_{i'} f(i)f(x+i-i')g(i'). \quad (5)$$

The similarity of $\mathbf{g}'$ to $\mathbf{g}$ as given by the $\mathbf{g} \cdot \mathbf{g}'$ dot product is

$$\mathbf{g} \cdot (\mathbf{f}\#(\mathbf{f}*\mathbf{g})_w)_x$$

$$= \sum_x g(x) \sum_i \sum_{i'} f(i)f(x+i-i')g(i')$$

$$= \sum_x \sum_i \sum_{i'} f(i)f(x+i-i')g(x)g(i'). \quad (6)$$

For moments, the expected value E of the dot-product similarity is

$$E[\mathbf{g} \cdot \mathbf{g}']$$

$$= E[\sum_x \sum_i \sum_{i'} f(i)f(x+i-i')g(x)g(i')]$$

$$= E[\sum_x \sum_i f^2(i)g^2(x)] = N^2 E[Z^2 W^2]$$

$$= N^2 \sigma^4 = N^2 P^2/N^2 = P^2 = 1, \quad (7)$$

where Z and W (and later X, Y, U, and V) are identically distributed independent random variables (with $\mu = 0$ and $\sigma^2 = P/N$). For the $\mathbf{g} \cdot \mathbf{g}'$ comparison, there are N^2 cases in which pairs of features match

and each case has $E[Z^2W^2]$. These N^2 cases come from the $N \times N$ cross products (e.g., $f_{-1}{}^2g_{-1}{}^2$, $f_{-1}{}^2g_0{}^2$, ..., $f_1{}^2g_1{}^2$ if $N = 3$) of $\mathbf{g} \cdot \mathbf{g}'$. [It must be remembered that $\mathbf{g}' = \mathbf{f}\#(\mathbf{f}*\mathbf{g})$.] There are also $(2N^3 + N)/3 - N^2$ cases in which both pairs of features mismatch and $E[ZWXY] = 0$. Because the assumption is that $\sigma^2 = P/N$, if $P = 1$, then $E[\mathbf{g} \cdot \mathbf{g}'] = 1$.

For the variance,

$$\text{Var}[\mathbf{g} \cdot \mathbf{g}'] = N^2 \, \text{Var}[Z^2W^2] + 2N\binom{N}{2}$$

$$\times \, \text{Cov}[Z^2W^2, Z^2X^2] + 2N\binom{N}{2}$$

$$\times \, \text{Cov}[Z^2W^2, X^2W^2] + \left(\frac{2N^3 + N}{3} - N^2\right)$$

$$\times \, \text{Var}[ZWXY] = \left(\frac{14N^3 + 9N^2 + N}{3}\right)\sigma^8$$

$$= \frac{14N^2 + 9N + 1}{3N^3}. \quad (8)$$

Because the variance of a sum is the sum of the variances plus twice the covariance of all $\binom{N}{2}$ pairwise combinations, we have two nonzero covariance components for the matching pairs. For each $\mathbf{f}$ value $(\cdots, f_{-1}, f_0, f_1, \cdots)$ there are $\binom{N}{2}$ Z^2W^2, Z^2X^2 covariances on $\mathbf{g}$, and because there are N such $\mathbf{f}$ values, twice the covariance is $2N\binom{N}{2}$. Likewise, for each $\mathbf{g}$ value $(\cdots, g_{-1}, g_0, g_1, \cdots)$ there are $\binom{N}{2}$ Z^2W^2, X^2W^2 covariances on $\mathbf{f}$, and because there are N such $\mathbf{g}$ values, the coefficient again is $2N\binom{N}{2}$.

By use of the moment-generating function (e.g., Hoel, 1962) it may be shown that for a normal distribution with $\mu = 0$, $E[Z^2] = \sigma^2$, $E[Z^4] = 3\sigma^4$, $E[Z^6] = 15\sigma^6$, and $E[Z^8] = 105\sigma^8$. Consequently,

$$\text{Var}[Z^2W^2]$$

$$= E[Z^4W^4] - E[Z^2W^2]E[Z^2W^2]$$

$$= E[Z^4]E[W^4] - E[Z^2]E[W^2]E[Z^2]E[W^2]$$

$$= 9\sigma^8 - \sigma^8 = 8\sigma^8. \quad (9)$$

There is again the assumption that $\sigma^2 = P/N$ and that $P = 1$.

These derivations, then, give the first two moments of the similarity distribution when a single pair of items is convolved, one of the items is correlated with the convolution, and the retrieved information is compared to the target item by means of the dot product. Even though $E[\mathbf{g} \cdot \mathbf{g}'] = E[\mathbf{g} \cdot \mathbf{g}] = 1$ (i.e., on average the retrieved information is a perfect match, as good as the item itself), the variance is not zero. Over trials there will be a distribution of similarity values around 1, and this must be taken into account in predicting recall performance.

So far the derivations have been restricted to a single pair. Now we must consider a list of pairs in which the information stored in memory includes both item and associative information as described by Equation 1.

For item recognition, the probe item is compared to the memory vector, and for an old-item probe $\mathbf{f}_k$ or $\mathbf{g}_k$ the expected values are

$$E[\mathbf{f}_k \cdot \mathbf{M}] = \gamma_1 c_k$$

$$E[\mathbf{g}_k \cdot \mathbf{M}] = \gamma_2 c_k, \quad (10)$$

where c_k is the serial-position constant for serial position k. For a list of p paired associates each presented once, $c_k = \alpha^{p-k}$. For a new item (one not on the list) $E[\mathbf{h} \cdot \mathbf{M}] = 0$.

For associative recognition, the expected value of a correct pairing is

$$E[(\mathbf{f}_k * \mathbf{g}_k) \cdot \mathbf{M}] = \gamma_3 c_k, \quad (11)$$

whereas for an incorrect pairing,

$$E[(\mathbf{f}_k * \mathbf{g}_l) \cdot \mathbf{M}] = 0. \quad (12)$$

So, item and associative expectations depend only on their weighting parameters and the serial position of the probe.

These results indicate one of the main advantages of a distributed-memory system. Expected values for item recognition are not affected by the presence of other items or by associative information in the memory vector. Expected values for associative recognition are not affected by the presence of other associations or by item information in the memory vector.

The reason is not hard to find. The elements of each vector are assumed to be nor-

mally distributed independent random variables, and $E[ZW] = E[Z]E[W] = 0$. Because $E[Z^2W] = E[Z^2]E[W] = E[Z]E[Z]E[W] = 0$, items are unaffected by associations and associations are unaffected by items.

For variance, basically nothing new is involved if one remembers that α and $\gamma_1-\gamma_3$ function as constants and $\text{Var}[cZ] = c^2 \text{Var}[Z]$, where c is a constant. What we need to predict performance are expectations and variances for item recognition (correct and incorrect pairings), and cued recall (correct responses and intralist intrusions).

For recognition of item information, for an old-item probe

$$\text{Var}[\mathbf{f}_k \cdot \mathbf{M}] = \text{Var}[\mathbf{g}_k \cdot \mathbf{M}]$$

$$= \left(\frac{1}{N} \gamma_i^2 + \frac{N+1}{N^2} \gamma_3^2\right) c_k^2 + A, \quad (13)$$

where

$$A = \left[\frac{1}{N}(\gamma_1^2 + \gamma_2^2) + \frac{3N^2+1}{4N^3} \gamma_3^2\right] \sum_j c_j^2 \tag{14}$$

and $i = 1$ if the probe is f (i.e., the A term of the A–B pair), but $i = 2$ if the probe is g (i.e., the B term of the A–B pair). For a new-item probe f,

$$\text{Var}[\mathbf{f}_n \cdot \mathbf{M}] = A. \tag{15}$$

For recognition of associative information, for a correct pairing $\mathbf{f}_k$–$\mathbf{g}_k$,

$$\text{Var}[(\mathbf{f}_k * \mathbf{g}_k) \cdot \mathbf{M}] = \left[\frac{N+1}{N^2}(\gamma_1^2 + \gamma_2^2)\right.$$

$$\left. + \frac{4N+3}{N^2} \gamma_3^2\right] c_k^2 + B, \quad (16)$$

where

$$B = \left[\frac{3N^2+1}{4N^3}(\gamma_1^2 + \gamma_2^2)\right.$$

$$\left. + \frac{2N^2+1}{3N^3} \gamma_3^2\right] \sum_j c_j^2. \quad (17)$$

For an incorrect pairing $\mathbf{f}_k$–$\mathbf{g}_l$,

$$\text{Var}[(\mathbf{f}_k * \mathbf{g}_l) \cdot \mathbf{M}] = \frac{N+1}{N^2}[(\gamma_1^2 + \gamma_3^2)c_k^2$$

$$+ (\gamma_2^2 + \gamma_3^2)c_l^2] + B. \quad (18)$$

For recall, the variance of the similarity of the retrieved information $\mathbf{g}'_k$ to the target item $\mathbf{g}_k$ given the probe $\mathbf{f}_k$ is

$$\text{Var}[\mathbf{g}_k \cdot \mathbf{g}'_k] = \frac{N+1}{N^2}(\gamma_1^2 + \gamma_2^2)$$

$$+ \frac{4N+3}{N^2} \gamma_4^2 + B. \quad (19)$$

For an intralist intrusion $\mathbf{g}_l$ given the probe $\mathbf{f}_k$,

$$\text{Var}[\mathbf{g}_l \cdot \mathbf{g}'_k] = \frac{N+1}{N^2}(\gamma_1^2 + \gamma_3^2)c_k^2$$

$$+ (\gamma_2^2 + \gamma_3^2)c_l^2. \quad (20)$$

To predict recall performance, we need to know the probability that the retrieved information $\mathbf{g}'$ is more similar to the target item $\mathbf{g}$ than to any other item and that the retrieved information $\mathbf{g}'$ is within criterial range of the target. The expression for the probability of correct recall is

$$PC = \int_a^b \phi_0(s) \prod_j \left[1 - \int_{P-s}^{P+s} \phi_{n,j}(s)\right] ds, \quad (21)$$

where the product $\prod$ runs from $j = 1$ to $j = m$. In this equation Criteria a and b determine the critical range; $\phi_0(s)$ is the old-item similarity distribution (the distribution of the g–g′ dot product), which is approximately normal with mean μ and variance σ^2, and $\phi_{n,j}(s)$ are the m new-item (i.e., intrusion) similarity distributions, which are also approximately normal, with mean zero and different variances depending on serial position.

Equation 21 is like the standard forced-choice expression of signal-detection theory (e.g., Green & Swets, 1966), which is why the terms old and new are used. The forced-choice expression is the probability that one observation drawn at random from the old-item distribution exceeds all $m - 1$ competitors from the new-item distributions. Here, each competitor can have a different variance and instead of exceeding the old-item sample must be closer to P.

The number of competitors would depend on the makeup of the list. If A and B terms were categorically different (e.g., digits and letters) then $m = p - 1$. That is, the number of competitors would be all B items other

than the target. If A and B terms were randomly sampled from the same pool (e.g., common English words), the $m = 2(p - 1)$. That is, the number of competitors would be the number of A and B items from all list pairs except the tested pair.

From this description it might seem as though each competitor somehow existed in memory waiting, as it were, to be compared. In fact, nothing like this is intended. There is the retrieved information $\mathbf{g'}$ and nothing more. The comparison with all competitors is for the experimenter's benefit—that is, to predict performance—but the subject only has the single vector $\mathbf{g'}$ as the basis for producing the target item $\mathbf{g}$.

I have now presented the necessary derivations to show how the general concepts of this theory may lead to explicit predictions about recall and recognition performance. It is still necessary to show how these equations may actually be applied to experimental data. This is the topic of the next section.

Application

At this point, the theory may seem rather complicated. In fact, it is really quite simple. With not unreasonable assumptions (i.e., that $\gamma_1 = \gamma_2$ and that a and b are equidistant from P), there are five parameters for recall and 5 ± 1 for recognition (-1 for forced

choice recognition, $+1$ for yes–no recognition). These parameters are N, the number of elements in the vector; α, the forgetting parameter; γ_1 and γ_3, the weighting parameters for item and associative information; and one, zero, or two criterion parameters for recall, forced-choice recognition, and yes–no recognition.

A convenient way to compute the variance is as follows: Set up what may be called $\mathbf{P}$ and $\mathbf{S}$ terms. (They are vectors, but unfortunately do not have quite the computational elegance one might wish). For the $\mathbf{P}$ term, first let $a = \gamma_1^2$, $b = \gamma_2^2$, $c = \gamma_3^2$, and let $A = c_k^2$, $B = c_l^2$, $C = \sum_{j \neq k} c_j^2$, $D = \sum_{j \neq l} c_j^2$, and $E = \sum_{j \neq k, l} c_j^2$. Then, $\mathbf{P0} = aA$, $\mathbf{P1} = bA$, $\mathbf{P2} = bB$, $\mathbf{P3} = cA$, $\mathbf{P4} = bB$, $\mathbf{P5} = aC$, $\mathbf{P6} = bC$, $\mathbf{P7} = bD$, $\mathbf{P8} = cC$ and $\mathbf{P9} = cE$.

The $\mathbf{S}$ terms are shown in Table 2. Then the necessary variances are the sum of selected $\mathbf{S} \times \mathbf{P}$ products, the selection varying with the case.

These components of variance are shown in Table 3 for the four cases of interest (item and associative recognition, old and new). Each individual term is an $\mathbf{S} \times \mathbf{P}$ product, and the variance is the sum of these $\mathbf{S} \times \mathbf{P}$ products. Thus, 00 is shorthand for $\mathbf{S0} \times \mathbf{P0}$, 51 is shorthand for $\mathbf{S5} \times \mathbf{P1}$, and so forth.

A numerical example to show the item and associative variance for a list of four pairs as a function of k and l is given in Table 4. In this example, $N = 15$, $\alpha = .9$, $\gamma_1 = \gamma_2 = .95$, $\gamma_3 = .8$, and $P = 1$.

Given that we now have the expectations (as given by Equations 10–12) and the vari-

Table 2

The S Terms to Calculate Variance

$$S0 = \frac{2P}{N}$$

$$S1 = \frac{P}{N}$$

$$S2 = P^3 \frac{7N^2 + 4N + 1}{4N^3}$$

$$S3 = P^3 \frac{3N^2 + 1}{4N^3}$$

$$S4 = P^4 \frac{14N^2 + 9N + 1}{3N^3}$$

$$S5 = P^4 \frac{5N^2 + 3N + 1}{3N^3}$$

$$S6 = P^4 \frac{2N^2 + 1}{3N^3}$$

Table 3

Components of Variance for Item Recognition and for Associative Recognition and Recall

Probe	Variance component
	Item recognition
Old	00 + 51 + 11 + 61 + 32 + 83
New	01 + 51 + 11 + 61 + 33 + 83
	Associative recognition and recall
Old	02 + 53 + 12 + 63 + 34 + 86
New	02 + 53 + 22 + 73 + 35 + 45 + 96

ances (as given by the **S** × **P** summed products), everything else is straightforward. We either want to estimate parameters from a given set of data or, given the parameters, want to know what the theory predicts. For accuracy measures, one simply uses the appropriate expression for probability correct and does the necessary computation. For recall, the appropriate expression is Equation 21, which, as stated, can be evaluated by numerical integration. For forced-choice item or associative recognition, I simply use the standard forced-choice expression from signal-detection theory; namely, the probability that an observation drawn at random from the old-item distribution exceeds the observations drawn at random from the m new-item distributions (one from each). Again, only numerical integration is required. For yes–no item recognition, it is generally sufficient to assume a single criterion and compute d' in the usual way.

Latency is not quite so simple. I do not know if analytic expressions can be developed, especially with the converging criteria. One approach is to run Monte Carlo simulations using the desired parameter values to determine the distributions and specify the criteria. An example will be given in the next section. To fit data, I would suggest the two-stage normal-exponential model suggested in Ratcliff and Murdock (1976). From this, one would extract the three parameters μ, σ, and τ (the mean, the standard deviation of the normal component, and the rate constant of the exponential component, respectively). However, it requires additional assumptions to relate μ, σ, and τ to the parameters of the present model, and so far this is an unexplored aspect of the theory.

With varying degrees of specificity, then, it is possible to apply the theory to experimental data and test its adequacy. So, I now turn to predictions.

Predictions

Recall and recognition accuracy and latency depend on the values of the model parameters: N, α, γ_1–γ_3, the Recall Criteria a and b, or, for yes–no recognition, the starting values and convergence rate of the criteria. The model may be fit to any given set of data

Table 4

Item and Associative Variance for Probe k, Response l, and p = 4

		Associative variance l			
k	Item variance	1	2	3	4
1	.513	.520	.487	.504	.524
2	.526	.487	.558	.517	.538
3	.542	.504	.517	.606	.555
4	.562	.524	.538	.555	.664
New	.457				

Note. Parameter values: $N = 15$, $\alpha = .9$, $\gamma_1 = \gamma_2 = .95$, $\gamma_3 = .8$, and $P = 1$.

by using standard parameter-estimation techniques, and predictions given the parameters may also be obtained. All this has been detailed in the preceding section.

Although detailed parameter estimation is possible, it may be premature. To avoid losing sight of the forest for the trees, it is necessary first to see if the theory is qualitatively compatible with the broad range of data that already exists in the field (see Estes, 1975). The possible argument that quantitative derivations are unnecessary for qualitative predictions is strongly denied. In theories of any appreciable complexity, intuitive predictions are not to be trusted. One needs quantitative derivations even to know what the qualitative predictions are. Computer simulations are useful in this regard if the derivations are not available.

What predictions does the theory make? Recall and recognition performance will both improve as N—the number of elements or attributes—increases. Recall and recognition will both fall off as α—the forgetting parameter—decreases. Although these two results are similar, they occur for different reasons. Increases in N have no effect on means; instead, they decrease variance. Changes in α, however, affect both means and variance.

The relative weighting of item and associative information can vary. This is determined by parameters γ_1, γ_2, and γ_3. In principle, the weighting could vary across tasks, nature of the material, and between and within subjects.

One of the strongest types of evidence consistent with this theory comes from studies

of intralist similarity in paired associates. Similarity is easily and naturally accommodated by the theory yet would seem to pose very serious problems for any kind of a network model. These matters are extensively reviewed by Eich (1982) so will not be discussed further here.

Recognition failure of recallable words also provides strong support for the theory. As studied by Tulving and his colleagues (see, e.g., Flexser & Tulving, 1978; Tulving & Thomson, 1973), this phenomenon shows that there are cases in which an item cannot be recognized but can be recalled. More specifically, following the presentation of a list of A–B pairs, there are cases where the B item cannot be recognized when presented by itself (or in forced-choice recognition) but can be recalled given A as the cue.

This phenomenon demonstrates the independence of item and associative recognition. Testing for recognition is testing item information, and testing for recall is testing associative information. As described in the Derivations section, $E[g \cdot (f*g)] = 0$. Thus, item information is independent of associative information, and the recognition failure is to be expected. Actually, the theory predicts complete independence, but Flexser and Tulving (1978) report that a large amount of data is best fit by a parabolic function about 10% above the chance line. This 10% effect could be due to averaging across subjects with slightly different parameter values.

The model is particularly applicable to studies of short-term memory for single paired associates. General results such as effects of type of material, list length, or presentation rate would not be hard to predict, but any theory could probably do so too. Instead, I will cite three more specific results that mesh nicely with the theory.

First, in probe paired-associates studies there are generally no primacy effects at all, only recency effects (Murdock, 1974). This is predicted because the forgetting is basically geometric (see Equation 1). Second, under a wide variety of conditions, there is associative symmetry (Murdock, 1974). That is, recall of A given B as the probe is as accurate as recall of B given A as the probe at all serial positions. This is predicted because convolution is commutative (see Equation 4). (If different material is used for A terms and for B terms—as in Lockhart, 1969, or Wolford, 1971—then of course the situation is different.) Third, intrusion analyses generally show that at any probe position, correct responses are more numerous than intrusions (Murdock, 1963). That is, no incorrect position contributes as many responses as the correct position does. (As correct responses and intrusions become less frequent, omissions become more frequent.) This is predicted because the expected $g' \cdot g$ similarity is higher than any $g' \cdot h$ similarity (which is always zero) and the variances are comparable (Equations 19 and 20). (Here is a case where predictions cannot be made unless theoretical variances are known. It would be quite possible for a distribution with a lower mean to have more area between Criterion a and Criterion b than a distribution with a higher mean if its variance were greater.)

A rather surprising prediction is that forgetting functions for cued recall and associative recognition should be roughly parallel. As shown in the section on derivations, the moments of the $g \cdot g'$ similarity distribution and the moments of the $(f*g) \cdot (f*g)$ similarity distribution are exactly the same. A priori, there is no particular reason why this should be so, but the derivations work out that way. I said "roughly" parallel because, even though the moments are exactly the same, probability of correct recall depends on the variance of all intralist intrusions whereas probability of correct recognition does not. Data that support this prediction may be found in Murdock (1970).

Latency and accuracy of recognition will vary inversely with location of Criteria a and b (see Figure 3). Within limits, as the criteria move further apart, the responses become more accurate but slower. This, of course, is the well-known speed–accuracy trade-off. As a demonstration, I ran a Monte Carlo simulation varying the random-noise magnitude—the initial difference between Criteria a and b—and their rate of convergence with indecisive outcomes.

In particular, I assumed a standard signal-detection case with old and new item distributions of unit variance and $d' = 1.5$. The extraneous noise variance was 0, .25, 1.0, 2.25, or 4.0 (so standard deviations of 0, .5,

1.0, 1.5, or 2.0, respectively). Criteria a and b were always centered on $d'/2$ (i.e., at .75) but had a starting difference (in z-score units) of 0, .75, 1.5, 2.25, or 3.0. The criteria adjustments were .95, .90, or .85. That is, after each observation that fell between Criteria a and b, the difference between them was reduced by the appropriate fraction before the decision about the next observation was made.

Good results were obtained for a criterion adjustment of .9. The obtained d' values and the count are given in Table 5 for the five noise and the five (criterion) start values. There were 1,000 trials for each of the 25 conditions. For each condition the obtained d' value was computed from the standard 2×2 table (old–new by yes–no) in the usual way (see, e.g., McNicol, 1972). The count value was simply the total number of samples over the 1,000 trials.

Obviously, when the criterion starting difference was zero, increasing the noise variance impaired performance. Conversely, when the noise variance was zero, increasing the starting difference slowed down the reaction time but had no effect on the obtained d' values. The noise variance did not seem to have much of an effect until it was at least as large as the new-item variance. Once it was as large or larger, the expected speed–accuracy results occurred.
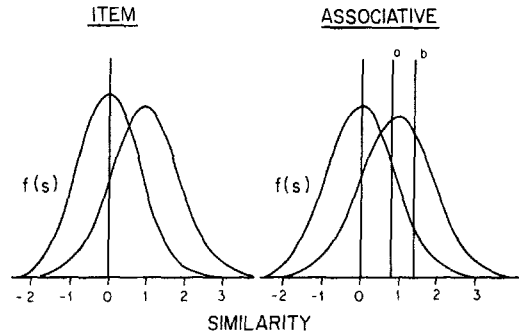
In general, then, this simulation showed

Figure 4. Old and new similarity distributions for item recognition and associative recognition and recall. (The parameter values used were $N = 15$, $\alpha = .98$, $\gamma_1 = \gamma_2 = \gamma_3 = 1$, $P = 1$, $a = .8$, and $b = 1.2$. [$p = 4$, $k = 1$, and $l = 3$.])

the desired speed–accuracy trade-off pattern. However, accuracy did not go to zero. Even with no waiting, d' was still about .75. Some data (e.g., Wickelgren, Corbett, & Dosher, 1980) do show the function all the way to $d' = 0$. In such cases, I would have to say that other factors were coming into play.

Let me close this section with what I think is a surprising result. Suppose we consider the relative levels of item and associative information in a short-term-memory list. In general, associative information is quite poor (e.g., Murdock, 1970), but item information is quite good (e.g., Murdock & Anderson, 1975). How can this be? A look at Equation 1 suggests one obvious explanation; let α and γ_1 be large, but let γ_3 be small. This will clearly work, but it seems a bit ad hoc. Anything could be "explained" this way. Is there any alternative?

One might think that, depending on the value of α, item and associative performance could both be quite good or both be quite poor, but not one good and one poor. After all, forgetting depends on α, and α operates on both item and associative information in the same way. However, it turns out that the answer to this puzzle is surprisingly simple and illustrates an important point about recall and recognition comparisons. An example is given in Figure 4, which shows the underlying distributions for item and associative information. That is, these are the similarity distributions for old and new items (item information) and correct recalls and intrusions (associative information).

Table 5

The d' and Count Values for Computer Simulation With Criterion Adjustment of .9

SD (noise)	Criterion starting difference				
	0	.75	1.50	2.25	3.00
			d'		
0	1.60	1.38	1.49	1.50	1.36
.5	1.34	1.17	1.33	1.72	1.31
1.0	1.06	1.19	1.16	1.35	1.34
1.5	.76	.94	.88	1.09	1.03
2.0	.73	.68	.72	.78	.98
			Count value		
0	1,000	3,502	5,400	7,741	9,863
.5	1,000	1,331	2,032	3,114	4,578
1.0	1,000	1,244	1,678	2,083	2,718
1.5	1,000	1,182	1,412	1,726	2,119
2.0	1,000	1,153	1,309	1,545	1,813

The parameter values here were $N = 15$, $\alpha = .98$, $\gamma_1 = \gamma_2 = \gamma_3 = 1.0$, $a = .8$, and $b = 1.2$. A four-pair list was assumed with the probe from the first serial position. For item information, the old-item distribution had a mean of .941 and a variance of .813. These are the first two moments of the similarity distribution when an item from the first pair is compared to the memory vector by means of the dot product, given these parameters. The new-item distribution had a mean of 0 and a variance of .691. These are the moments of the comparable similarity distribution when a new-item probe is compared to the memory vector. For recall, the $\mathbf{g} \cdot \mathbf{g}'$ similarity distribution had a mean of .941 and a variance of .919. These are the first two moments of the similarity distribution of the retrieved information $\mathbf{g}'$ compared to the target item $\mathbf{g}$, given the probe $\mathbf{f}$. The $\mathbf{h} \cdot \mathbf{g}'$ similarity distribution had a mean of 0 and a variance of .808. This is actually the variance for an intrusion from the third serial position, but Positions 2–4 were virtually identical, so the third was selected.

For recognition, assume a two-alternative forced-choice procedure. An old and a new item are presented, and the subject must identify the old item. For the distributions shown in Figure 4, the predicted percentage correct would be 77.5%. Probability of correct recall is given by Equation 21 and turns out to be .143. In the same way, the probability of an intrusion from Serial Position 3 is .080.

Thus, we have a situation where forced-choice recognition is relatively good (about 78%) but probability of correct recall is quite low (about .14). This happens despite the fact that γ_1, γ_2, and γ_3—the weighting factors for item and associative information—are the same. Basically, the reason is very simple. Item recognition requires only that the old-item similarity be greater than the new-item similarity. Associative recall requires that $\mathbf{g}'$—the retrieved information—be close enough to the target item $\mathbf{g}$ to permit recall.

It could be argued that performance here is not very good. Generally, one would expect both better recognition and better recall. To remedy this, I recalculated the predictions with $N = 63$ and $\alpha = .98$. Then, predicted

recognition accuracy rose to .94, whereas predicted correct recall was .32. These are very much what one might expect, and certainly substantiate the conclusion. Recognition can be quite good whereas recall is quite poor even though both item and associative information are stored in a common memory vector and operated on in the same way by the forgetting parameter α.

Alternative Assumptions

So far the assumptions have been stated rather baldly, without much justification or explanation. The reason is that I have started with the most difficult case first. Having worked through the model and obtained predictions, it is very easy to consider alternative assumptions and derive predictions for them. Had I started with an easier case, the initial presentation would have been somewhat briefer, but there would still be much left to do.

The main assumptions are that items can be represented by random vectors, that convolution and correlation are the storage and retrieval operations, that item and associative information are stored in a common memory vector, that the decision system works on the output from the memory comparison or retrieval process, and that uncertain outcomes are not acted on until a more decisive result occurs. The assumption I would like to discuss further is the third assumption, that item and associative information are stored in a common memory vector.

Consider three alternatives. The first alternative is that item and associative information are stored in separate memory vectors. That is, there are two memory vectors, one for item information and one for associative information. The second alternative is that there is a single memory vector for each pair of items, but this vector holds both the item and the associative information for that pair. Thus, there would have to be as many separate memory vectors as pairs in the list. The third alternative is that there is a separate memory vector for each item and for each pair. Thus, there is no distributed memory at all, and there would have to be three times as many memory vectors as pairs in the list (one for each A item, one for each B item, and one for each A–B pair).

Let the original assumptions be Model 1, the first alternative be Model 2, the second alternative be Model 3, and the third alternative be Model 4. The four models are illustrated in Figure 5. The components of variance for Model 1 have already been presented in Table 3. The components of variance for Models 2–4 are shown in Table 6. It is assumed that for an old-item probe, one can directly access the relevant memory vector, but one must interrogate all appropriate memory vectors for a new-item probe. They can almost be derived by inspection. That is, some of the variance components of Model 1 are absent in the other models, but which components are absent depends on the model. The beauty of this approach is that one has a simple and direct way to compare these various models that are the same in all respects except the assumption about the format of the store.

It is quite clear that Models 2–4 all have smaller variance than Model 1. The expectations are identical for all four models. Consequently, Models 2–4 all give better performance than Model 1. Table 7 shows predicted performance for parameters $N = 15$, $\alpha = .98$, $\gamma_1 = \gamma_2 = \gamma_3 = 1$, $P = 1$, $a = .8$, and $b = 1.2$. The three columns show predicted percentage correct on two-alternative forced-choice item recognition, probability of a correct recall, and probability of an intrusion. (As before, $p = 4$, $k = 1$, and $l = 3$). Figure 6 shows the distributions drawn to scale. Going from Model 1 to Model 4, performance improves appreciably, and Model 4 is clearly the best.
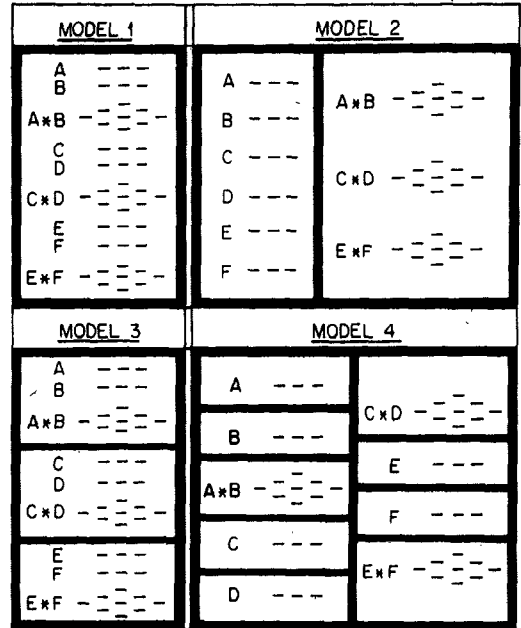

Figure 5. Illustration of the memory vectors for Models 1–4. (For each model, all the item or associative information enclosed in the heavy margins is stored in a single memory vector.)

However, what is missing from this analysis is the cost of a search mechanism. Particularly for Models 3 and 4, if there is localized storage, there must be some mechanism to examine the relevant locations. How does the mechanism know where to look? How much time does it take? Does it make a decision after each interrogation, and what are the decision rules? None of these problems bother Models 1 or 2 because no search

Table 6
Components of Variance for Models 2–4

	Variance component		
Probe	Model 2	Model 3	Model 4
	Item recognition		
Old	00 + 51 + 11 + 61	00 + 11 + 32	00
New	01 + 51 + 11 + 61	01 + 51 + 11 + 61 + 33 + 83	01 + 51
	Associative recognition and recall		
Old	34 + 86	02 + 12 + 34	34
New	35 + 45 + 96	02 + 22 + 35 + 45	35 + 45

Note. Compare to Table 3.

Table 7

Predicted Performance on Item Recognition and Associative Recall for Models 1–4

Model	Recognition	Recall	
		Correct	Intrusion
1	.775	.143	.080
2	.816	.223	.048
3	.827	.197	.064
4	.937	.274	.030

Note. Parameter values: $N = 15$, $\alpha = .98$ $\gamma_1 = \gamma_2 = \gamma_3 = 1$, $P = 1$, $a = .8$, and $b = 1.2$ ($p = 4$, $k = 1$, $l = 3$).

process is required. (Model 2 has two memory vectors, but which one is queried would depend on the nature of the test.)

Search models and network representations are currently quite popular, but unfortunately, their proponents seldom tell us enough about the processes so meaningful comparisons can be made. In the present case, there is clearly a trade-off; Models 1 and 2 do not give as good performance as do Models 3 and 4, but neither do they need any search mechanism to accomplish their task.

It is, however, possible to make a somewhat more decisive comparison of Models 1 and 2. We can determine the S/N (signal to noise) ratio of each model and compare them. The S/N ratio is simply the squared expectation over the variance (see, e.g., Wozencraft & Jacobs, 1965). (The familiar d' of signal-detection theory is the square root of the S/N ratio). Let us define the efficiency of Model 1 as its S/N ratio compared to the S/N ratio of Model 2. Because the expectations are the same, the efficiency is simply the variance ratio (Model 2 to Model 1). Using the P and S notation described previously, the efficiency for item information (denoted Eff_I) is

$$\text{Eff}_I = \frac{01 + 51 + 11 + 61}{01 + 51 + 11 + 61 + 33 + 83}, \quad (22)$$

and the efficiency for associative information (denoted Eff_A) is

$$\text{Eff}_A = \frac{35 + 45 + 96}{02 + 53 + 22 + 73 + 35 + 45 + 96}. \quad (23)$$

In both cases I am using the new-item variances not the old-item variances.

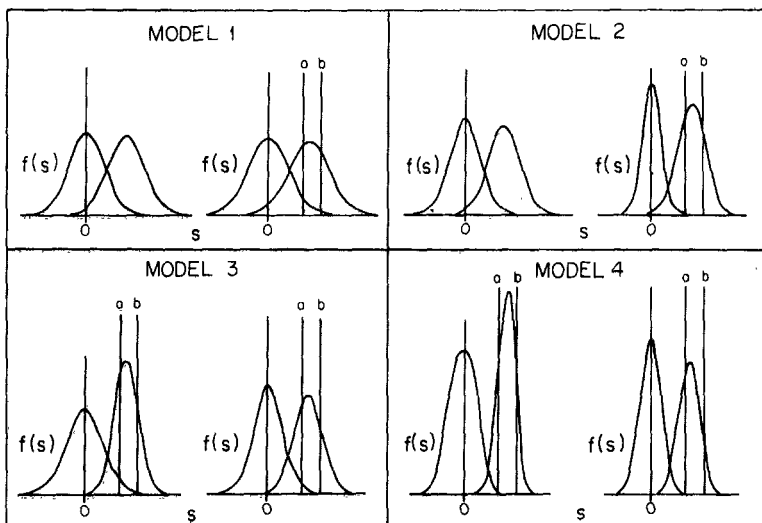If we assume that $\gamma_1 = \gamma_2 = \gamma_3 = 1$ and that $P = 1$, then as an approximation we find

Figure 6. Old and new similarity distributions for item recognition and associative recall for Models 1–4. (The parameter values used were $N = 15$, $\alpha = .98$, $\gamma_1 = \gamma_2 = \gamma_3 = 1$, $P = 1$, $a = .8$, and $b = 1.2$. [$p = 4$, $k = 1$, and $l = 3$.])

that

$$\text{Eff}_I \approx \frac{8N^2}{11N^2 + 1}$$

and

$$\text{Eff}_A \approx \frac{4N^2 + 2}{13N^2 + 5}. \qquad (24)$$

Finally, the limit of Eff_I as N approaches infinity is 8/11, or 73%, and the limit of Eff_A as N approaches infinity is 4/13, or 31%. Thus, it turns out that there is a price to pay for combining item and associative information in the same memory vector, and the price is much greater for associative information than for item information.

What should we conclude? Obviously, it would be rash to opt for Model 1 when it is so much less efficient (has so much lower an S/N ratio) than Model 2. On the other hand, Model 1 is simpler than Model 2 in that there is only one memory vector rather than two. Perhaps it would be prudent not to take a strong stand here yet. We know the strengths and weaknesses of the two models, and we can compare them quantitatively; additional work will be needed to reach a decision.

In concluding this section, I should emphasize that Model 1 and Model 2 are both distributed-memory models. In both cases, item information from all pairs in the list is stored in a single memory vector, and associative information from all pairs in the list is stored in a single memory vector. What differs is whether the item-information and the associative-information memory vectors are the same. By contrast, Models 3 and 4 are really not distributed-memory models at all. They use localized, not common, storage.

Additional Mechanisms

The theory as described will work as indicated, but some additional mechanisms or processes are necessary for it to be complete. First, the question may be raised about where the information is being stored while the convolution and correlation operations are being carried out. Taken to an extreme, the notion of a distributed memory with only a single memory vector might preclude the very operations whose end products it was designed to store. Clearly, something like the notion of a working memory is needed. Even if the two items f and g and their convolution f∗g
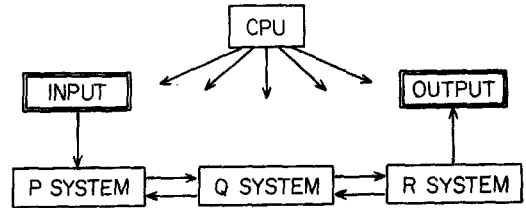

Figure 7. Block diagram to indicate the role of the P system, the Q system, and the R system. (The central processing unit [CPU] oversees all of them.)

were all stored in working memory before being added to the memory vector, only three buffers would be needed. Working memory is generally assumed to be at least this big (see, e.g., Baddeley & Hitch, 1974). However, each buffer must be able to store an N-dimensional random vector.

A more serious problem is how to go from the retrieved information g′ to recall of the target item g. So far all I have said is that successful recall depends on this step. Its probability is given by Equation 21, but no details have been provided about mechanisms or processes. I am not able to give a satisfactory answer to this problem, but let me at least indicate what sort of answer might be worth exploring.

When an item is presented, a perceptual system processes the sensory information. This processed information is the input to memory (e.g., Bower, 1967). We take this for granted and seldom bother even to mention it. Likewise, there must be a comparable response system that takes the information from memory and uses it to make some overt responses. This interaction between the memory system and the response system is responsible for going from g′ to g. Generally, the interaction between the perceptual system and the memory system is indifferent to the mode of presentation (e.g., auditory or visual), yet surely, different sensory systems come into play depending on presentation modality. In the same way, the interaction between the memory system and the response system is largely indifferent to whether the response is to be written or spoken, but here too the motor processes in writing and speaking are quite different.

To be slightly more specific, let me illustrate these ideas with a highly oversimplified block diagram. This is shown in Figure 7

where we have the P (perceptual) system, the Q (query) system, and the R (response) system. These three systems communicate with one another. The P system receives input from the environment and the R system puts out responses to the environment. The CPU (central processing unit) oversees all. I have called the Q system the query system because it is designed to answer inquiries. It is here that the stored information—the memory vector—resides. (I think it would be misleading to call the Q system a "memory" system, because memory is not unique to it. After all, neither the P system nor the R system could work if they did not include stored information too.)

This block diagram is so oversimplified it does not even show working memory. Obviously, it could be embellished considerably, but I would rather not do so. My only purpose here is to highlight a very simple notion, namely, that execution of a response is not the job of the Q system. It still remains to specify how the CPU is able to map g'—the retrieved information—into an acceptable response g.

How the storage and retrieval operations (summation, dot product, convolution and correlation) are carried out is also not specified. However, now we are talking about much more basic mechanisms and processes and they are simply assumed, not explained. From a computational point of view, none of these operations are very difficult to implement, and surely, it is not unreasonable to assume the brain has these capabilities. In fact, Anderson (1973) and Anderson, Silverstein, Ritz, and Jones (1977) have been quite specific about possible neural mechanisms for a distributed-memory model. However, the aim here has been to present an abstract model that characterizes the operation of the system; the detailed realization is at a different level of discourse.

Relation to Other Theories

CADAM, a model for content-addressable distributed associative memory proposed by Liepa (Note 1), was the starting point for the present work. (This model is discussed in Murdock, 1979.) What I have done here is to make a number of additional assumptions and explore some of their implications. The rehearsal model of Metcalfe and Murdock (1981) is an application of these concepts to free recall. CHARM (Eich, 1982) is another application of convolution and correlation but deals only with associative information, not item information. CHARM and the present work were parallel developments. Each fills in some of the gaps of the other.

This work is most closely related to recent developments in distributed-memory models (see Hinton & Anderson, 1981). Although the same concepts are used, the aim is somewhat different. In general, these distributed-memory models are neural models of one sort or another and are concerned with underlying mechanisms of memory storage and retrieval. My intention is to develop an abstract model that can be applied to experimental data on recall and recognition in a fairly detailed manner. The underlying neural mechanisms are obviously very important, but interpretation and understanding of experimental data may be another means to the same end, namely, understanding how human memory works.

The decision system suggested for recognition is not completely novel either. The recognition models of Atkinson and Juola (1973) and Mandler (1980) are well-known examples of models that use normal distributions and two criteria. Actually, as pointed out by Vickers (1980), two-criterion models have been suggested and tested in the psychophysics area; see Cartwright and Festinger (1943) and Swets and Green (1961).

Evaluation

In closing, let me attempt a brief evaluation of this approach. A distributed-memory system using convolution and correlation permits efficient storage of item and associative information and provides direct access to stored information. There is no need for search, and the complications of a network model are completely avoided. Starting from basic assumptions, one can work out explicit expressions to predict performance, and quantitative comparisons of models are possible. The model is broadly consistent with much of the short-term-memory data from relatively simple tasks, though detailed and extensive parameter estimation lies ahead.

Rather than depending on ad hoc as-

sumptions, this theory derives and provides a rationale for several well-established phenomena in the area of human memory. One is the use of a signal-detection analysis for recognition memory. Rather than simply assume a signal and noise distribution, this theory suggests what their origin might be. They could be the similarity distributions for the dot product of new- and old-item probes with the memory vector. Another is the distributional analysis of Ratcliff and Murdock (1976). They showed that reaction-time distributions are well described by a two-stage model that is characterized by a normal distribution with parameters μ and σ and an exponential distribution with parameter τ. Here, the exponential or waiting-time distribution arises from the decision system. No response occurs until an observation falls below Criterion a or above Criterion b. The time course of a stochastic process like this is exponential (e.g., Feller, 1968). Third, a rationale for the speed–accuracy trade-off is also provided. Accuracy and speed of responding will vary inversely as the location of the two criteria, a and b, are varied.

The theory suggested here has somewhat the same viewpoint as the theory of the ideal observer in psychophysics (see, e.g., Green & Swets, 1966). Both are designed to provide the specifications of an ideal system necessary to perform at the obtained level. The fact that a subject has a $d' = 2.0$ means that an ideal observer with this capability could do as well; the real observer might be an imperfect observer with a higher d'. In the same sense, here too we are specifying ideal conditions for specified performance levels. In fact, the real subject might have a more powerful system that was being used inefficiently.

I should also mention some problems. First, I have dealt only with item and associative information. The simplest model for serial order (a chaining model) is much more complex. Even though I have worked out some of the expressions, the parameter-estimation programs are unbelievably cumbersome and frustrating. Second, the question of capacity is not to be lightly dismissed. Although one can make memory performance as good as desired by increasing N— the number of elements in the memory vector—what are these elements? If it takes 50, 100, 1000, or 10,000 elements to produce the

necessary results, they are certainly not the cognitive features that others have in mind. At the neurological or biochemical level, these numbers would not pose any problem, but then the interpretation might be somewhat different.

Third, and most serious, is the learning problem. I have said nothing about it in this article, but I do not want to hide it. If more than a single presentation is used, it seems reasonable to assume simply that additional representations are added to the memory vector. In fact, this idea must be wrong. If one works out the predictions, one finds that although the expected values increase, so do the variances. In the cases I have examined, the variances increase so as to offset exactly the increases in the means. As a result, learning would not be expected. I have explored other possibilities and some look promising. However, the unwary reader is warned that extension from the single-presentation case to the multipresentation case is not completely without surprises.

Finally, to end on a more positive note, let me point out a few aspects of this approach that might have some general implications for our understanding of human memory. First, the convolution and correlation algorithm makes very explicit one possible way information could be stored and retrieved. What is stored is not a "wax tablet" or graven image; in fact, what is stored is not in any sense any sort of an item at all. An association is not a link or a path or a bond between two items. It is an unrecognizable aggregate of the two items it relates.

What is retrieved is not the target itself but information sufficient to retrieve the target. Thus, we do not have to retrieve perfect replicas to recall or recognize. The P system handles perception, the R system handles responding, and the function of the Q system is to mediate the two. It allows information received at one point in time by the P system to be output at a later point in time by the R system without storing the actual stimuli or responses of either.

Reference Note

1. Liepa, P. *Models of content addressable distributed associative memory (CADAM)*. Unpublished manuscript, University of Toronto, 1977.

References

Anderson, J. A. A theory for the recognition of items from short memorized lists. *Psychological Review*, 1973, *80*, 417–438.

Anderson, J. A., Silverstein, J. W., Ritz, S. R., & Jones, R. S. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 1977, *84*, 413–451.

Anderson, J. R., & Bower, G. H. *Human associative memory*. Washington, D.C.: V. H. Winston, 1973.

Asch, S. E. A reformulation of the problem of associations. *American Psychologist*, 1969, *24*, 92–102.

Atkinson, R. C., & Juola, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973, pp. 583–612.

Baddeley, A. D., & Hitch, G. Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8). New York: Academic Press, 1974, pp. 47–89.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 1). New York: Academic Press, 1967, pp. 229–325.

Bracewell, R. *The Fourier transform and its applications* (2nd ed.). New York: McGraw-Hill, 1978.

Cartwright, D., & Festinger, L. A quantitative theory of decision. *Psychological Review*, 1943, *50*, 595–621.

Cavanagh, P. Holographic and trace strength models of rehearsal effects in the item recognition task. *Memory & Cognition*, 1976, *4*, 186–199.

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

Estes, W. K. Some targets for mathematical psychology. *Journal of Mathematical Psychology*, 1975, *12*, 263–282.

Feller, W. *An introduction to probability theory and its applications* (Vol. 1; 3rd ed.). New York: Wiley, 1968.

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

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

Greeno, J. G., James, C. T., DaPolito, F., & Polson, P. G. *Associative learning: A cognitive analysis*. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Hinton, G. E., & Anderson, J. A. (Eds.). *Parallel models of associative memory*. Hillsdale, N.J.: Erlbaum, 1981.

Hoel, P. G. *Introduction to mathematical statistics* (3rd ed.). New York: Wiley, 1962.

Horowitz, L. M., & Manelis, L. Toward a theory of redintegrative memory: Adjective–noun phrases. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 6). New York: Academic Press, 1972, pp. 193–224.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970, pp. 334–374.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. *Journal of Experimental Psychology*, 1969, *79*, 12–17.

McNicol, D. *A primer of signal detection theory*. London: Allen & Unwin, 1972.

Mandler, G. Recognizing: The judgment of previous occurrence. *Psychological Review*, 1980, *87*, 252–271.

Metcalfe, J. A., & Murdock, B. B., Jr. An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 1981, *20*, 161–189.

Murdock, B. B., Jr. Short-term retention of single paired associates. *Journal of Experimental Psychology*, 1963, *65*, 433–443.

Murdock, B. B., Jr. Short-term memory for associations. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970, pp. 285–304.

Murdock, B. B., Jr. *Human memory: Theory and data*. Potomac, Md.: Erlbaum, 1974.

Murdock, B. B., Jr. Convolution and correlation in perception and memory. In L.-G. Nilsson (Ed.), *Perspectives in memory research: Essays in honor of Uppsala University's 500th anniversary*. Hillsdale, N.J.: Erlbaum, 1979, pp. 105–119.

Murdock, B. B., Jr., & Anderson, R. E. Encoding, storage, and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, N.J.: Erlbaum, 1975, pp. 145–194.

Pribram, K., Nuwer, M., & Baron, R. The holographic hypothesis of memory structure in brain function and perception. In D. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: Freeman, 1974, pp. 416–457.

Ratcliff, R., & Murdock, B. B., Jr. Retrieval processes in recognition memory. *Psychological Review*, 1976, *83*, 190–214.

Swets, J. A., & Green, D. M. Sequential observations by human observers of signals in noise. In C. Cherry (Ed.), *Information theory: Proceedings of the fourth London symposium*. London: Butterworths, 1961, pp. 177–195.

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

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, *76*, 559–573.

van Heerden, P. J. A new optical method for storing and retrieving information. *Applied Optics*, 1963, *2*, 387–392.

Vickers, D. Discrimination. In A. T. Welford (Ed.), *Reaction times*. London: Academic Press, 1980, pp. 25–72.

Wickelgren, W. A., Corbett, A. T., & Dosher, B. A. Priming and retrieval from short-term memory: A speed-accuracy trade-off analysis. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 387–404.

Wickens, D. D. Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: V. H. Winston, 1972, pp. 191–215.

Wolford, G. Function of distinct associations for paired-associate performance. *Psychological Review*, 1971, *78*, 303–313.

Wozencraft, J. M., & Jacobs, I. M. *Principles of communication engineering*. New York: Wiley, 1965.