Psychological Review
1997, Vol. 104, No. 4, 839–862

Context and Mediators in a Theory of Distributed Associative Memory (TODAM2)

Bennet B. Murdock
University of Toronto

Recent findings on the interactions of item and associative information have necessitated a revision of TODAM (a theory of distributed associative memory). Item information underlies the memory for individual items or events, and associative information allows one to relate or associate 2 separate items or events. The troublesome findings are the differential forgetting of item and associative information (item recognition falls off over retention intervals, whereas associative information does not) and the differential emphasis results (greater attention to items hurts pair recognition, but greater attention to pairs does not affect item recognition). The addition of context and mediators enables TODAM to account for these interactions.

Item information is the information in memory that enables one to recognize individual objects or events ("items"). Associative information is the information in memory that enables one to remember pairs of items (e.g., king–queen or bread–butter) so that one can recognize or recall them. Associative information is what relates two items so that, for instance, a person can recall one item given the other item as the cue.

Early evidence for the distinction between item information and associative information was presented in Murdock (1974). Recent evidence has been summarized by Hockley (1991) and Clark and Gronlund (1996). Much of this evidence is discussed here. TODAM (a theory of distributed associative memory) is an attempt to provide a general theory for the storage of item, associative, and serial-order information (Murdock, 1993); in this article, I deal only with item information and associative information.

This article is divided into three main sections, followed by discussion. The first section reviews TODAM and explains how it deals with item and associative information. The second section deals with some recent experimental findings that turn out to be problematic for TODAM. The third section presents a revision of TODAM that can account for these problematic data. Then I discuss recent experimental findings on context, the mirror effect, and receiver operating characteristic (ROC) slopes and end with a comparison of the revision with other global-matching models.

TODAM

TODAM (Murdock, 1979, 1982) deals with the encoding, storage, and retrieval of item and associative information. It uses the convolution–correlation formalism of Borsellino and Poggio (1973) and was an extension of the CADAM model of Liepa (1977). CADAM (content-addressable distributed associative memory) was really three models in one, dealing separately with item, associative, and serial-order information. TODAM combined them into a single theoretical framework.

Representation Assumption

Following Anderson (1973), TODAM assumes that items can be represented as random vectors. That is, they are vectors of random variables such that each element in the item vector is a random sample from a specified feature distribution. Another way to say this is that an item can be represented as a list of attributes or features (e.g., Bower, 1967; Underwood, 1969), and each feature is a random variable. The feature distribution is assumed to be a normal distribution with mean zero and variance P/N, where P (the power of the vector) is 1 and N is the dimensionality (number of features) of the item vectors.

An immediate consequence of this representation assumption is that item vectors are automatically normalized to 1 (i.e., P) regardless of the numerical value of N. That is, for any item vector $\mathbf{f}$, the expected value E of the dot product of the item vector with itself is

$$E(\mathbf{f}\cdot\mathbf{f}) = 1,$$

where the dot (or inner) product is the sum of the squared features; that is,

$$E(\mathbf{f}\cdot\mathbf{f}) = \sum_{i=1}^{N} f(i)f(i).^{1}$$

The normalization is statistical; $E(\mathbf{f}\cdot\mathbf{f}) = 1$, but $\mathrm{Var}(\mathbf{f}\cdot\mathbf{f}) > 0$. That is, the expected value is 1.0, but there is some variability. The item vectors are independent but not orthogonal. If they were orthogonal, the expected value would also be 1, but there would be no variability; the dot product of every item with itself

This work was supported by Natural Sciences and Engineering Research Council of Canada Grant APA146.

I am most appreciative of the many comments and questions Kevin Murnane and Steven Clark made about the article.

Correspondence concerning this article should be addressed to Bennet B. Murdock, Department of Psychology, Room 4020, 100 St. George Street, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Electronic mail may be sent via Internet to murdock@psych.toronto.edu.

[1] For a more general account of the dot product in particular and vector spaces in general, see Jordan (1986).

would always be exactly 1.0. For independent item vectors, the dot product of an item with itself is not always exactly 1.0 because the elements of the item vectors are random variables. For orthogonal item vectors, all pairwise dot products are exactly 1.0.[2]

To say that item vectors are independent means that, for any two item vectors **f** and **g**,

$$E(\mathbf{f}\cdot\mathbf{f}) = E(\mathbf{g}\cdot\mathbf{g}) = 1 \text{ but } E(\mathbf{f}\cdot\mathbf{g}) = 0.$$

Again, there is variability; thus, whereas $E(\mathbf{f}\cdot\mathbf{g}) = 0$, here too $\mathrm{Var}(\mathbf{f}\cdot\mathbf{g}) > 0$. These ideas should not be unfamiliar; for unit-normal random variables Z and W, $E(ZZ) = E(WW) = \sigma^2$, but $E(ZW) = 0$ and $\mathrm{Var}(ZW) > 0$. Thus, random variables are similar to random vectors, and these properties result from the basic representation assumption of Anderson (1973) adopted by TODAM.

Convolution and Correlation

Convolution (*) is a way of combining two item vectors, and correlation (#) is a way to unpack them. From Aristotle to Pavlov to interference theory to semantic-memory models, associations have always been viewed as connections; an association is a link between two nodes. With the advent of holographic models in the 1960s and 1970s, a different possibility (convolution) was suggested, and associations as convolutions seemed more appropriate for distributed memory models.

This possibility is also consistent with the Gestaltist view of an association as a "wholistic entity" (e.g., Asch, 1969; Kohler, 1941; Rock & Ceraso, 1964), but no mathematical formalism was suggested. This formalism was provided by Borsellino and Poggio (1973), who showed how the mathematical operation of convolution could be used to form an association and correlation could be used to retrieve one of the items given the other item as the cue or probe.

As a formal definition, for item vectors **f** and **g** the xth element of their convolution is

$$(f*g)(x) = \sum_i f(i)g(x-i). \tag{1}$$

One can form a matrix from the cross products of the elements of **f** and **g** and sum the elements of the matrix along the negative diagonals (Figure 1). The sum of these elements is the convolution. If **f** and **g** are three-element vectors, then **f*g** will be a five-element vector. More generally, if **f** and **g** are N-dimensional vectors, then **f*g** will be of dimension $2N - 1$.

What is the point of convolution? It makes for a fault-tolerant memory. Going back to the matrix step, for an A–B pair each feature of A (the rows) is distributed over all features of B (the columns), and vice versa. Thus, if there is local damage, there will not be complete memory loss. Instead, the loss will be proportional to the amount of damage. This is still true when the negative diagonals are summed to form the convolution. From a neural-network point of view, all of the elements (weights) of A are represented in all of the elements (weights) of B, and vice versa.

An association (i.e., convolution) does not "look like" an item; it is not even of the same dimensionality. In effect, element i of **f** is attenuated (multiplied) by all of the elements of **g** and

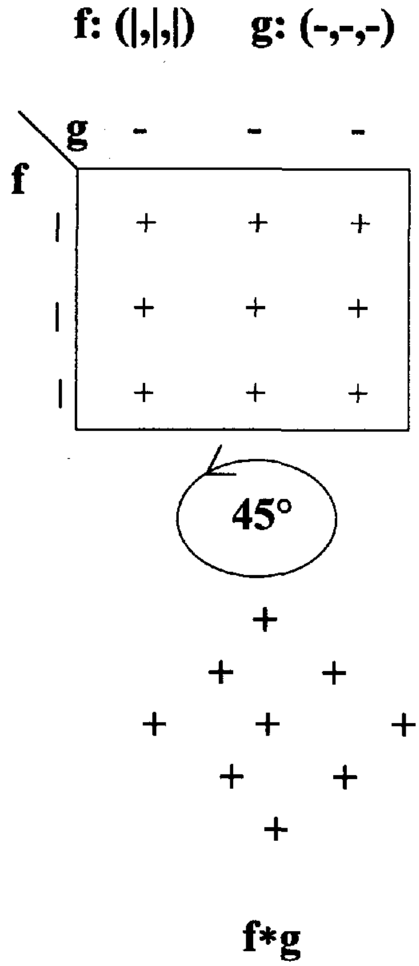

Figure 1. Illustration of convolution. The cross (or outer) product matrix of **f** and **g** is formed (top), rotated 45° counterclockwise, and summed vertically to form **f*g** (bottom). The result (i.e., **f*g**) is here a five-element vector, and the elements of this vector are the sum of the five columns.

spread or dispersed over about half of the convolution, but the locations shift systematically with i. Connectionist models use a similar formalism, but they describe it in terms of the weight matrix. The weight matrix is the cross-product (technically, outer product) matrix before the negative diagonals are summed.

If associations are formed by convolution, they can be unpacked by correlation. If $\boldsymbol{\delta}$ is the delta vector (a vector with center element 1 flanked by zeros) and primes (e.g., $\boldsymbol{\delta}'$ and $\mathbf{g}'$) are approximations (to $\boldsymbol{\delta}$ and $\mathbf{g}$), then

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

Each element in the approximation ($\mathbf{g}'$) matches the corre-

[2] Orthogonal item vectors would result in a slightly greater signal-to-noise ratio, but they seem quite unrealistic for any model of human memory.

sponding element in the target item (**g**), but there is a small plus–minus.[3]

The fact that correlation retrieves only an approximation to the target item has important consequences for TODAM. In particular, it requires a cleanup or deblurring process to map the approximation $\mathbf{g}'$ into the target item $\mathbf{g}$. This is not formally modeled by TODAM, but one possibility is a Hopfield net, as used by Lewandowsky and Li (1994). A Hopfield net (Hopfield, 1982) is an autoassociative matrix that can deblur a noisy approximation.

Storage Equation

So far I have discussed only single items and single associations; how are they combined? In TODAM, I assume a single all-purpose memory vector that can be denoted by **M**. If one has a list of paired associates consisting of items **f** and **g**, after the jth pair has been presented, the simplest storage equation is

$$\mathbf{M}_j = \alpha\mathbf{M}_{j-1} + \mathbf{f}_j + \mathbf{g}_j + \mathbf{f}_j * \mathbf{g}_j, \tag{2a}$$

where α is the forgetting parameter, $0 \leq \alpha \leq 1$. That is, the memory vector before the current pair is presented is attenuated (multiplied) by α, and then the current items ($\mathbf{f}_j$ and $\mathbf{g}_j$) and the association ($\mathbf{f}_j * \mathbf{g}_j$) are added. If $\alpha = 0$ there is complete forgetting, and if $\alpha = 1.0$ there is no forgetting; generally α is close to 1.0. Thus, the memory vector contains both item and associative information, and α is the forgetting parameter.

Because items and associations are of different dimensionality, how can they be combined? Both items and associations (convolutions) are center justified, and one can think of item vectors as padded with zeros at both ends to make their dimensionality equal to that of the associations (Figure 2).

Equation 2a assumes equal attention to the items and the association, but this is unnecessarily restrictive. Subjects can surely pay differential attention to items or their association. I parameterize this by γ. The limited-capacity assumption (that attention is limited, so the separate γ coefficients must sum to 1.0) yields

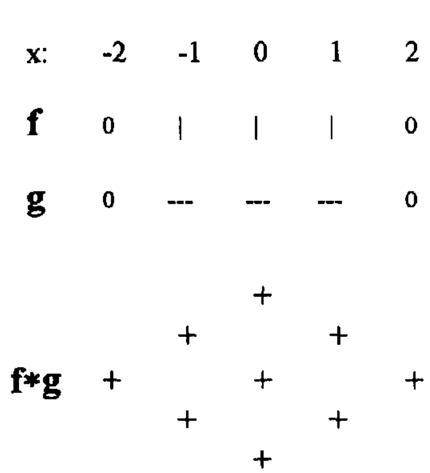

Figure 2. Illustration to show how the end elements of **f** and **g** are padded with zeros at both ends so that the dimensionality of the item vectors is the same as that of the convolution.

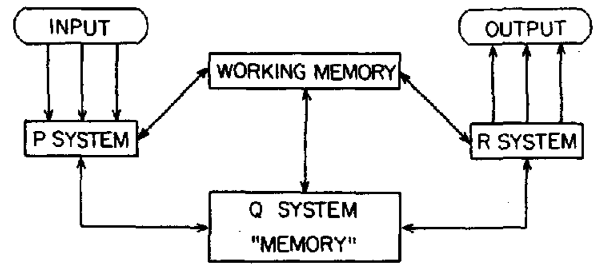

Figure 3. System architecture. The P system (perceptual) is for encoding, the Q system (queries) is where the memory vector (**M**) is located, and the R (response) system is for output. The TODAM operations are assumed to go on in working memory.

$$\mathbf{M}_j = \alpha\mathbf{M}_{j-1} + \gamma\mathbf{f}_j + \gamma\mathbf{g}_j + (1 - 2\gamma)(\mathbf{f}_j * \mathbf{g}_j). \tag{2b}$$

Because $0 \leq \gamma \leq 0.5$, there is a seesaw effect; the more attention paid to the items (association), the less attention paid to the association (items).

It is also necessary to assume that only some of the features of the items are encoded on each presentation.[4] This is represented by probabilistic encoding; each feature of an item vector is encoded with probability p or not encoded (set to zero) with probability $1 - p$, $0 \leq p \leq 1.0$. If $p = 0.0$, then no features are encoded; if $p = 1.0$, then all features are encoded; and, with intermediate values, some features are encoded. In previous applications p has been about 0.4. If $p\mathbf{f}$ is a shorthand for probabilistic encoding, then

$$\mathbf{M}_j = \alpha\mathbf{M}_{j-1} + \gamma p\mathbf{f}_j + \gamma p\mathbf{g}_j + (1 - 2\gamma)(p\mathbf{f}_j * p\mathbf{g}_j). \tag{2c}$$

Equation 2c is the basic storage equation for item and associative information with parameters α, γ, and p.

System Architecture

Where are these operations (convolution, correlation, scalar multiplication, and summation) carried out? I assume a P system (perceptual), a Q system (query), an R system (response), and working memory. The system architecture is shown in Figure 3, and the operations are carried out in working memory. The arrows indicate transfer of information from one system to another, and they are usually bidirectional.[5]

The main point of Figure 3 is to show the division of labor that must occur. I do not model the P system; that is where encoding occurs. Thus, my representation scheme is abstract, and one cannot set up any isomorphisms between real-world stimuli and TODAM item vectors. Nor do I model the R system; this is for response factors (e.g., deblurring). The memory vector (**M**) resides in the Q system (the Q system is designed to

[3] The error in the approximation (magnitude of the plus–minus values) is inversely proportional to N.

[4] If one assumes that all of the features are encoded on each presentation, performance, as measured by d', will not improve after the first trial, so learning will not occur; see Murdock and Lamon (1988) for details.

[5] A detailed account of working memory has been presented elsewhere (e.g., Murdock, 1993); it is not necessary for the present article.

answer queries), and the operations are carried out in working memory.

Comparison Process

When an item or associative probe is presented on a recognition list, I assume that the comparison process is the dot product. Item probes are either old or new (i.e., members of the list or not), and associative probes are either intact ("old") or rearranged, mixed (one old and one new item), or new (two new items). The latter three are mismatches, so the correct response is no.

For an item probe, I assume that the subject takes the dot product of the probe item with the memory vector, and the result of this comparison process is the basis for decision. For an associative probe, I assume that the subject convolves the two items and then takes the dot product of the convolution with the memory vector. One can derive explicit expressions for the means and variances of the dot product of the probe with the memory vector in terms of the parameters of the model and so compare the predicted d' values with experimental results.

Experimental Data

Four main sets of experimental findings are reported here. They are the length–difficulty relation, forgetting of associative information, item and associative forgetting, and instructional effects. The first study (length–difficulty relation) dealt with paired-associates learning. It constituted the starting point for the entire set of experiments. The remaining three (forgetting of associative information, item and associative forgetting, and instructional effects) were all short-term recognition-memory studies that followed up one of the implications of the length–difficulty study. The results of the first two sets of studies support TODAM, but the results of the second two pose serious problems for TODAM.

Length–Difficulty Relation

The length–difficulty relation describes the function relating number of pairs in a list to trials to learn to criterion. According to a two-process model proposed by Thurstone (1930), trials to criterion should be a function of the square root of the number of pairs in the list.

An experimental test of the length–difficulty relation was reported in Murdock (1989). There were 2 subjects, and each subject learned lists of 9, 16, 25, 36, 49, 64, 81, and 100 pairs to a criterion of one perfect trial. Each subject learned one list a day, and there were five replications of each list length in counterbalanced order.

The results for three of the conditions (9, 49, and 100 pairs) are shown in Figure 4. The triangles are the data points averaged over replications, and the smooth curves are the fit of the model for each condition for each subject. The fits are quite good, especially when it is considered that, for each subject, the parameter values were the same for all eight conditions. None of the parameters varied with list length. Thus, it is the model that is doing the work, not the parameters.

The surprise was the very high value of α in the fits (Subject 1, α = .996; Subject 2, α = .998). This means that there was very little forgetting. In retrospect, it had to be this way; otherwise, the subjects would never have learned the longest lists. What is surprising is that forgetting is so common in the interference-theory literature, even with an A–B, C–D design. Perhaps these were exceptional subjects. In any event, this conclusion certainly needed verification.

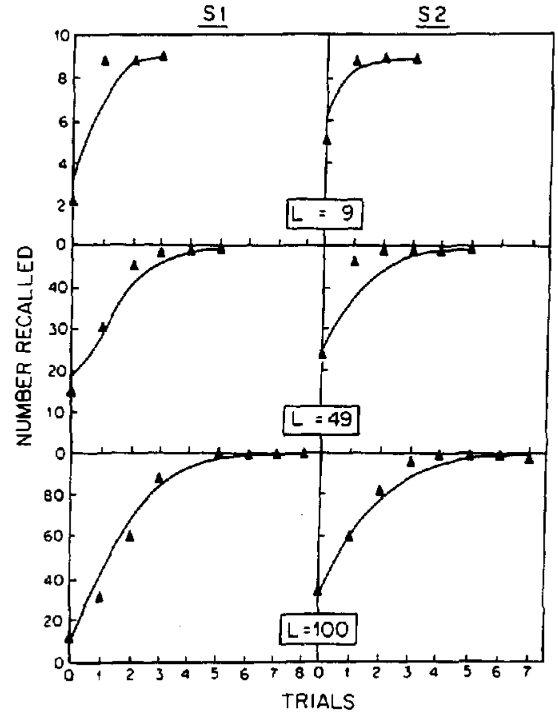

Figure 4. Results of the length–difficulty study. Number of items recalled as a function of number of trials is shown in the three left-hand panels for Subject 1 and in the three right-hand panels for Subject 2. Adapted from "Learning in a Distributed Memory Model," by B. B. Murdock, 1989, in C. Izawa, *Current Issues in Cognitive Processes: The Tulane Floweree Symposium on Cognition* (p. 92), Hillsdale, NJ: Erlbaum. Copyright 1989 by Erlbaum. Adapted with permission.

Forgetting of Associative Information

The purpose of the next set of experiments (reported in Murdock & Hockley, 1989) was to test whether, in fact, there was little or no forgetting of associative information over short periods of time. We used a continuous task in which subjects studied pairs of unrelated words, and our independent variable was lag. Lag is the number of pairs (study or test) intervening between study and test of a given pair. Each pair was intact ("old") or rearranged ("new"), so we were able to use d' as our dependent variable.

The results of four experiments are shown in Figure 5. Experiment 2 used two kinds of pairs, nouns and nonnouns. Experiment 3 tested each pair twice, and the results are shown separately for the first test and the second test. Experiment 4 presented each pair three times (spaced repetition), and the results are shown separately for the first presentation, the second presentation, and the third presentation. The results are very clear; both visually and statistically, there was no forgetting.

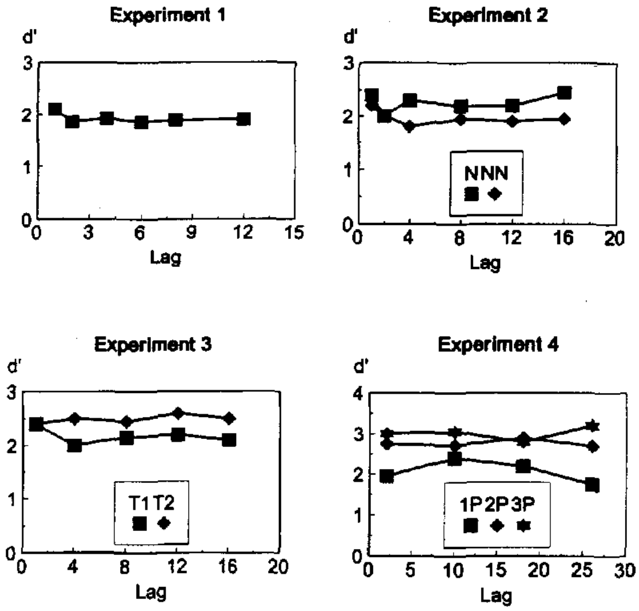

Figure 5. Results of the four experiments in Murdock and Hockley (1989). Each panel shows d' as a function of lag (number of pairs intervening between study and test). Experiment 2 used nouns (N) or nonnouns (NN), Experiment 3 tested each pair twice (T1 and T2), and Experiment 4 had 1, 2, or 3 presentations (P) of each pair.

This was a clear confirmation of the α prediction; as noted by Lewandowsky (1993), it was a case in which a fit of the model to data led to a prediction that was confirmed in subsequent experiments. These results seem even more surprising when one considers that there is much evidence (e.g., see Murdock, 1974) that item information is forgotten over comparable lag ranges. Unfortunately for TODAM, the story does not end here. Most of the data on item and associative forgetting are between experiments, so the next set of studies examined item and associative forgetting within the same experiments.

Item and Associative Forgetting

Using a continuous task, Hockley (1991) presented pairs but tested either items or pairs at comparable lags. For item recognition, the items were either old or new; for associative recognition, the pairs were either intact or rearranged. Item information was controlled in tests of associative recognition because all items in both intact and rearranged pairs were old. The tests were mutually exclusive, so tested items were never included in tested pairs, and vice versa.

The results are shown in Figure 6. As before, there was no forgetting of associative information; in fact, visually (but not statistically), associative information improved with lag in Experiments 3 and 4. However, in all four experiments, the d' value for item information decreased with lag, and this was no surprise; such a result is almost always found in studies of item recognition. Thus, under comparable conditions, in within-experiment comparisons, item information is forgotten over lag ranges, whereas associative information is not forgotten. These results were replicated and generalized in Hockley (1992).

These results are clearly counter to TODAM. As can be seen in Equation 2a, the forgetting parameter α acts on the memory vector **M**, and, because items (**f** and **g**) and associations (**f*****g**) are both stored in **M**, item information and associative information should both be forgotten (or not forgotten) at the same rate. Introduction of the parameters γ (Equation 2b) and p (Equation 2c) does not change this prediction.

Instructional Effects

Before discussing these results further, I present the fourth set of findings, those dealing with instructional effects. The experiments were designed by Hockley and Cristi (1996) to test the limited-capacity assumption (see Equation 2b). If subjects are instructed to remember the items, they should do worse if tested on the associations; if they are instructed to remember the associations, however, they should do worse if tested on the items. This is the seesaw effect mentioned earlier.

The experiments used a study–test procedure, and lists of paired associates were presented. One group of subjects was given item-emphasis instructions ("Remember the items"), whereas the other group was given associative-emphasis instructions ("Remember the associations"). Each list was followed by appropriate recognition tests (new and old items for the item-emphasis group and intact and rearranged pairs for the associative-emphasis group), and neither group was tested for the other kind of information. However, there was a final recognition test for both groups that involved both types of information, and these data constituted the main finding of the experiment.

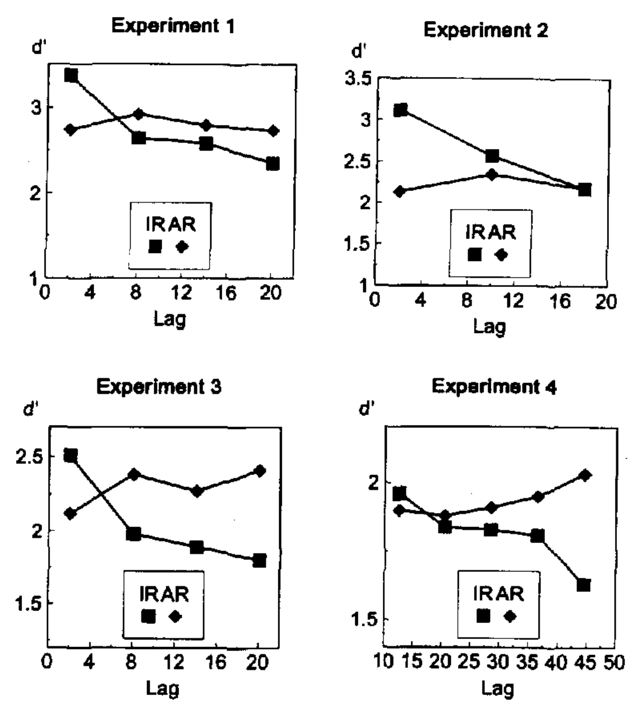

Figure 6. Forgetting of item (squares) and associative (diamonds) information. Results of four experiments from Hockley (1991). IR = item recognition; AR = associative recognition.

The results of the final recognition test are shown in Figure 7. As can be seen, performance on pair recognitions dropped considerably when associative emphasis was changed to item emphasis. However, the converse was not true; performance on the item tests was visually and statistically the same for both emphasis groups. Thus, counter to the TODAM prediction, the half seesaw effect was obtained, and half a seesaw is not better than none at all.

TODAM Revision

It is clearly necessary to modify TODAM if there is to be any hope of explaining these data. The general idea is as follows: Embed items in context, assume context drifts over time, assume context affects item information more than it affects associative information (to explain the differential forgetting of item and associative information), and have item information included in associative information so that there are two bases for item recognition. In this way, item but not associative information will be independent of γ, so in principle this could explain the half seesaw effect.

Context

In keeping with the representation assumption of TODAM, I assume that context is a random vector, just as items are random vectors. The elements are random variables, and I do not specify what they are. The dimensionality of item vectors is N, and the dimensionality of the context vector is N_c. Items are embedded in context, so $N_c > N$.

Because items and associations are center justified, so is context. Thus, context flanking elements extend the item vectors to the left and right, and this is shown in Figure 8. For the context vector $\mathbf{c}$, $\mathbf{a}$ denotes the left-hand side, $\mathbf{b}$ denotes the right-hand side, and their elements are represented by xs. Embedding is a concatenation operation in which concatenation is symbolized by $\oplus$. I define the embedded items $\mathbf{F}$ and $\mathbf{G}$ as

$$\mathbf{F} = \mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b}$$

and

$$\mathbf{G} = \mathbf{a} \oplus \mathbf{g} \oplus \mathbf{b}.$$

That is, I use uppercase letters to denote item plus context. Because the A and B items are presented at the same time, they are assumed to have the same context; thus, $\mathbf{a}$ and $\mathbf{b}$ are the same for $\mathbf{F}$ and $\mathbf{G}$.

Binding

If one simply summed $\mathbf{F}$ and $\mathbf{G}$ so that

$$\mathbf{M} = \alpha\mathbf{M}_{j-1} + \mathbf{F}_j + \mathbf{G}_j, \tag{3a}$$

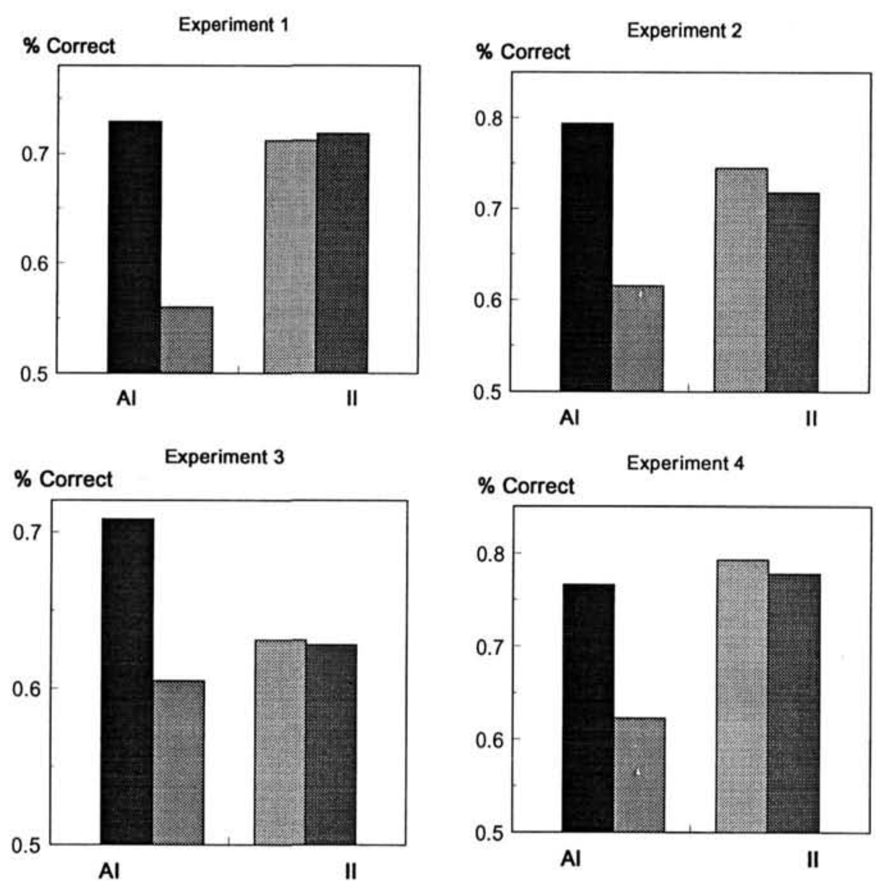

Figure 7. Results on the final recognition test that examined associative information (two left-hand bars in each panel) or item information (two right-hand bars in each column). For both types of tests, the associative-emphasis (A1) condition is shown on the left and the item-emphasis (II) condition is shown on the right. Results from Hockley and Cristi (1996).

the items would be free floating; they would not be bound to context. Consequently, I autoassociate **F** and **G** to bind the items to their context. If one denotes the autoconvolution $\mathbf{F}*\mathbf{F}$ as $\mathbf{F}^{*2}$ or $\mathbf{G}*\mathbf{G}$ as $\mathbf{G}^{*2}$ and considers only the item information, one would have

$$\mathbf{M}_j = \alpha\mathbf{M}_{j-1} + \mathbf{F}_j^{*2} + \mathbf{G}_j^{*2}. \tag{3b}$$

Autoconvolution here plays a role similar to activation in the MINERVA model of Hintzman (1986, 1988). Activation is the summed strength (number of matching features) cubed, and raising it to the third power serves to bind the items to context.[6]

Associations

To have two bases for item recognition, I sum the *A* and *B* items before embedding them in context and doing the autoassociation. This is illustrated in Figure 9, and I denote the embedded pair of items as **FG**. The context elements **a** and **b** are still the same as in **F** and **G**. If one adds the attention weights (γ), one has

$$\mathbf{M}_j = \alpha\mathbf{M}_{j-1} + \gamma\mathbf{F}_j^{*2} + \gamma\mathbf{G}_j^{*2} + (1 - 2\gamma)(\mathbf{FG}_j)^{*2}. \tag{3c}$$

It is to be understood that probabilistic encoding (p) applies to the item vectors (both **f** and **g**) but not to the context. This is the new storage equation for the revised version of TODAM.[7]

Weights

Although not shown explicitly in Equation 3c, it is necessary to use weights for **f** + **g** in the **FG** term. The reason is that

$$E[(\mathbf{f} + \mathbf{g})\cdot(\mathbf{f} + \mathbf{g})] = E(\mathbf{f}\cdot\mathbf{f}) + E(\mathbf{g}\cdot\mathbf{g}) = 2,$$

and this violates the principle that all item vectors should be normalized to 1. If, instead, I say that

$$\mathbf{FG} = \mathbf{a} \oplus \upsilon(\mathbf{f} + \mathbf{g}) \oplus \mathbf{b} \tag{4}$$

and let $\upsilon = 1/\sqrt{2}$, then

$$\mathbf{E}[(\upsilon\mathbf{f} + \upsilon\mathbf{g})\cdot(\upsilon\mathbf{f} + \upsilon\mathbf{g})] = \upsilon^2 E(\mathbf{f}\cdot\mathbf{f}) + \upsilon^2(\mathbf{g}\cdot\mathbf{g}) = 1,$$

and the principle is satisfied. That is, the **FG** chunk is normalized to 1.0, even though the item and associative components in the chunk are not. So it is to be understood that **FG** is as defined by Equation 4, where $\upsilon = 1/\sqrt{2}$.

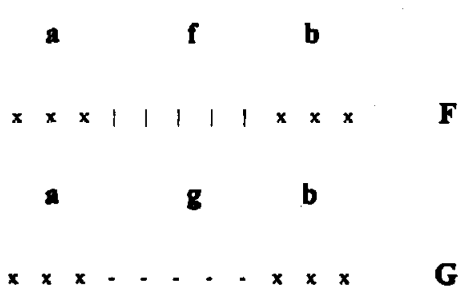

Figure 8. Concatenation of **a**, **f**, and **b**. The context elements of **a** and **b** are denoted by *x*s, whereas the content elements are denoted by vertical bars for **f** and dashes for **g**.

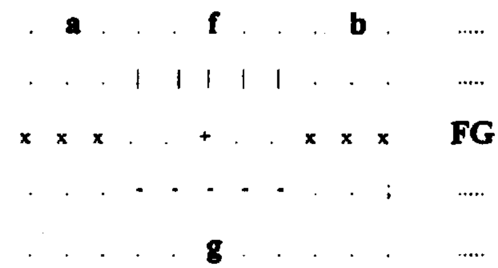

Figure 9. **FG** shown as the sum of item vectors **f** and **g** embedded in the context vectors **a** and **b**. The context elements of **a** and **b** are denoted by *x*s, whereas the content elements are denoted by vertical bars for **f** and dashes for **g**.

Context Drift

As in Estes (1955), I assume that context changes slightly from moment to moment. More specifically, if $\mathbf{c}_j$ is the context when the jth pair is presented, and **z** is random noise, then

$$\mathbf{c}_j = \rho\mathbf{c}_{j-1} + \sqrt{(1 - \rho^2)}\mathbf{z},$$

where ρ is the parameter that determines the magnitude of the context drift, $0 \le \rho \le 1$.

This formalization is exactly the same as the one I have used previously to specify the similarity of an exemplar to a prototype (Murdock, 1995). Just as ρ is the similarity of an exemplar to a prototype, so ρ here is the similarity of the interpair context, that is, the context similarity between the jth and $(j - 1)$th pairs. The noise coefficient $\sqrt{(1 - \rho^2)}$ serves to keep the context normalized to 1 because, for any j, $E(\mathbf{c}_j\cdot\mathbf{c}_j) = E\{[\rho\mathbf{c}_{j-1} + \sqrt{(1 - \rho^2)}\mathbf{z}]\cdot[\rho\mathbf{c}_{j-1} + \sqrt{(1 - \rho^2)}\mathbf{z}]\} = E[\rho^2(\mathbf{c}_{j-1}\cdot\mathbf{c}_{j-1})] + E[(1 - \rho^2)(\mathbf{z}\cdot\mathbf{z})] = \rho^2 + (1 - \rho^2) = 1$.

Derivations

It is necessary to work out the expected d' values for a recognition task when old or new items or intact or rearranged pairs are used as the probe. According to TODAM, the comparison process is the dot product, so one needs the old-item and new-item means and variances for these four kinds of probes when the probe item or pair is dotted with the memory vector. However, it simplifies the task considerably if one uses the continuous memory assumption, because then one does not need the variances; rather, the one needs only the expectations.

According to the continuous memory assumption (Murdock & Kahana, 1993), memory carries over not only from list to list but also from the preexperimental to the experimental session. Obviously, it must unless one assumes that **M** is a tabula rasa that is reset to zero at the start of each list, and

---

[6] MINERVA2 also has context elements, but they are list specific rather than item specific. Comparisons with other models are considered in the Discussion section.

[7] The assumption that probabilistic encoding is applied to the items but not the context is arbitrary, and other assumptions could be investigated. There are more content features than context features, so this assumption tends to equalize their effects.

this seems rather unlikely. The continuous memory assumption has been used before in some TODAM simulations (Murdock & Lamon, 1988), and it provides a simple explanation for the list-strength (non) effect of Ratcliff, Clark, and Shiffrin (1990).[8]

The derivations for the expectations are presented in the Appendix. With a few small exceptions, the derivations are purely algebraic and are no more difficult than multiplying two series together and collecting terms. In fact, they are easier because there are no cross-product terms. That is,

$$(x + y + z)(x + y + z) = x^2 + y^2 + z^2 + 2xy + 2xz + 2yz;$$

however, for independent random vectors $\mathbf{f}$, $\mathbf{g}$, and $\mathbf{h}$, the expected value (E) is

$$E[(\mathbf{f} + \mathbf{g} + \mathbf{h})\cdot(\mathbf{f} + \mathbf{g} + \mathbf{h})] = E(\mathbf{f}\cdot\mathbf{f}) + E(\mathbf{g}\cdot\mathbf{g}) + E(\mathbf{h}\cdot\mathbf{h})$$

because

$$E(\mathbf{f}\cdot\mathbf{g}) = E(\mathbf{f}\cdot\mathbf{h}) = E(\mathbf{g}\cdot\mathbf{h}) = 0.$$

This also works for convolutions; for example,

$$E\{[(\mathbf{f}*\mathbf{g} + \mathbf{g}*\mathbf{h})]\cdot[(\mathbf{f}*\mathbf{g}) + (\mathbf{g}*\mathbf{h})]\} = E[(\mathbf{f}*\mathbf{g})\cdot(\mathbf{f}*\mathbf{g})] + E[(\mathbf{g}*\mathbf{h})\cdot(\mathbf{g}*\mathbf{h})]$$

because

$$E[(\mathbf{f}*\mathbf{g})\cdot(\mathbf{g}*\mathbf{h})] = 0.$$

Components

The many terms in the dot products can be partitioned into four components. These components are pure context ($\mathbf{a}*\mathbf{a}$, $\mathbf{b}*\mathbf{b}$, and $\mathbf{a}*\mathbf{b}$), pure item ($\mathbf{f}*\mathbf{f}$ and $\mathbf{g}*\mathbf{g}$), context item ($\mathbf{a}*\mathbf{f}$, $\mathbf{a}*\mathbf{g}$, $\mathbf{f}*\mathbf{b}$, and $\mathbf{g}*\mathbf{b}$), and pure associative ($\mathbf{f}*\mathbf{g}$). There are no single items here because, regardless of the probe type, all terms are convolutions (e.g., $\mathbf{F}*\mathbf{F}$, $\mathbf{G}*\mathbf{G}$, or $\mathbf{FG}*\mathbf{FG}$).

These components are similar to those in the SAM (search of associative memory) model of Shiffrin and Raaijmakers (1992). The difference is that the components are derived from the revision of TODAM when context is added, whereas they are simply assumed in the SAM model, in which context has always been included. SAM is a computational model (i.e., given the parameter values, one can compute the predicted results), and TODAM is a process model, but perhaps there is more similarity than would first appear. SAM is capable of fitting a large amount of data at a quantitative level, and perhaps with these modifications TODAM will be able to do so as well.

Predictions

Each of the four components (pure context, pure item, context item, and pure associative) has weighting parameters (α, p, ρ, and N_a, and N_b, where $N_a = N_b$ is the number of elements in $\mathbf{a}$ and $\mathbf{b}$). If one considers only the pure-item component, neglecting the encoding probabilities (which function only as a scale factor), and writes Equation 3c as

$$\mathbf{M}_j = \alpha\mathbf{M}_{j-1} + \gamma(\mathbf{a}_j \oplus \mathbf{f}_j \oplus \mathbf{b}_j)^{*2} + \gamma(\mathbf{a}_j \oplus \mathbf{g}_j \oplus \mathbf{b}_j)^{*2} + (1 - 2\gamma)(\mathbf{a}_j \oplus v(\mathbf{f}_j + \mathbf{g}_j) \oplus \mathbf{b}_j)^{*2},$$

then, dropping the subscript, the pure-item component is

$$\gamma(\mathbf{f}*\mathbf{f}) + \gamma(\mathbf{g}*\mathbf{g}) + (1 - 2\gamma)v^2(\mathbf{f}*\mathbf{f} + \mathbf{g}*\mathbf{g}) = [\gamma + v^2(1 - 2\gamma)][\mathbf{f}*\mathbf{f} + \mathbf{g}*\mathbf{g}].$$

For an old-item probe, say $(\mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b})^{*2}$, the pure-item component is $\mathbf{f}*\mathbf{f}$, and the pure-item component of $E(\text{Probe}\cdot\text{Memory})$ is

$$\begin{aligned} &E(\text{Probe}\cdot\text{Memory}) \\ &= E\{[\gamma + v^2(1 - 2\gamma)](\mathbf{f}*\mathbf{f} + \mathbf{g}*\mathbf{g})\cdot\mathbf{f}*\mathbf{f}\} \\ &= [\gamma + v^2(1 - 2\gamma)]E[(\mathbf{f}*\mathbf{f})\cdot(\mathbf{f}*\mathbf{f})]. \end{aligned}$$

Because, for any realistic value of N (e.g., 100 or more), if Z and W are two identically distributed independent random variables,

$$\begin{aligned} E[(\mathbf{f}*\mathbf{f})\cdot(\mathbf{f}*\mathbf{f})] &= NE[Z^4] + \tbinom{N}{2}E[(2ZW)(2ZW)] \\ &= 3N\sigma^4 + 2N(N - 1)\sigma^4 \cong 2N^2(P/N)^2 \quad (5) \end{aligned}$$

(Weber, 1988); thus, because $P = 1$ (P is the power of the vector, and, as described, it is set to 1 for normalization purposes), the pure-item component of the dot product of the probe item with the memory vector is

$$2[\gamma + v^2(1 - 2\gamma)]\alpha_{kt},$$

where α_{kt} is a function of α that depends on serial position (k) and test position (t). Now v is the weighting coefficient in $\mathbf{FG}$ and $v = 1/\sqrt{2}$, and consequently

$$2[\gamma + v^2(1 - 2\gamma)] = 2[\gamma + 0.5(1 - 2\gamma)] = 1.$$

This means that the pure-item component for an item probe is independent of γ for any serial position (k) and any test position (t).

This is a remarkable result. It means the half seesaw effect falls directly out of the model without any parameters or ad hoc assumptions at all. That is, if the pure-item component is the same regardless of γ, then γ does not affect performance. Consequently, performance on an item-recognition test should be independent of the allocation of attention to items or pairs. Item information is stored redundantly. An increase in the γ (item) component is compensated for by a decrease in the $1 - 2\gamma$ (associative) component.

Unfortunately, that is not the whole story. There is also the item-context component,[9] which does show some dependence

[8] The concomitant list-length effect is more problematic; this analysis cannot explain data (e.g., Ohrt & Gronlund, 1995) showing that some effect of list length is obtained even when recency is controlled.

[9] There is no pure-associative component for an item probe, and because the probe is embedded in context at the time of test, the pure-context component is the same for old- and new-item probes. Consequently, the pure-context component does not affect the predicted d' value for an item probe, and this is also true in SAM.

on γ because the weights are v and w rather than v^2 and w^2 (see Appendix). Thus, only if γ is very small (maximizing the contribution of the pure-item component and minimizing the contribution of the context-item component) will one get this pure independence. But that immediately means one will lose the differential forgetting of item and associative information. This differential forgetting comes about because context changes affect item more than associative information. But if one minimizes the context-item component by using a small value of γ, item forgetting will depend on α, and one will lose the item–associative interaction.

Suppose that the assumption is, instead, that subjects do not use context on an associative recognition test. Rather than $(\mathbf{FG})^{*2}$, they use $(v\mathbf{f} + v\mathbf{g})^{*2}$. This is not unreasonable; with associative recognition tests that use intact and rearranged pairs, the context is, on average, the same for intact and rearranged pairs and thus provides no useful information. In item recognition, context is essential, because that is what differentiates old and new items. With the continuous memory assumption and α very close to 1, the d' value (old–new) would be very close to zero if subjects did not use context on an item-recognition test.[10]

One can have it both ways: The half seesaw effect is obtained, as shown earlier, because it is assumed that subjects do use context on item-recognition tests; however, it is assumed that subjects do not use context to discriminate between intact and rearranged pairs because context provides no useful information. For a pair recognition test, if one disregards probabilistic encoding, then the pure associative component when $\mathbf{f}_j$ and $\mathbf{g}_k$ are two items in the probe is

$$E\{[2(\mathbf{f}_j*\mathbf{g}_k)2(\mathbf{f}_j*\mathbf{g}_k)]\cdot\mathbf{M}\} = \begin{bmatrix} 4(1-2\gamma)v^4\alpha_{kt}, & j = k \\ 0, & j \neq k \end{bmatrix},$$

where again α_{kt} is a function of α that depends on serial (k) and test (t) position. Specifically, if L is list length, then $\alpha_{kt} = \alpha^{L-k+t-1}$. That is, α_{kt} depends on lag, and lag is a function of serial position k and test position t.

Combining this with the item analysis produces the item–associative interaction nicely but leaves out the main effects. The decrease in d' with lag for pair recognition is virtually unnoticeable, but the absolute level of d' for pair recognition is too low. In particular,

$$d'_I(t) = \mu_O(t) - \mu_N(t) = s1 + s2 + s3 - s1 = s2 + s3$$

(see Appendix), and, because $v = 1/\sqrt{2}$,

$$d'_I(t) = p_s^2 p_r^2(\alpha_{kt} + (N_a/N_c + N_b/N_c)p_s p_r(\alpha\rho)_{kt}.$$

For a pair probe,

$$d'_A(t) = t2I + t4I - t2R \cong t4I = (1-2\gamma)p_s^2 p_r^2(\alpha\rho)_{kt}$$

(see Appendix). Consequently, the difference (Δ) between $d'_I(t)$ and $d'_A(t)$ is

$$\Delta = d'_I(t) - d'_A(t)$$
$$\cong 2\gamma p_s^2 p_r^2 \alpha_{kt} + (N_a/N_c + N_b/N_c)p_s p_r(\alpha\rho)_{kt}.$$

Because Δ is positive for any serial position k or test position t, the item and associative retention curves will never cross. Hockley's item and associative forgetting curves intersected in all four experiments (Figure 6), and the model cannot predict this.

Mediators

The one remaining problem for the model, then, is to increase the absolute level of performance for pair recognition without changing anything else, and the obvious thing to try is mediators. By a mediator I mean a third item—call it **m**—that, by virtue of preexperimental associations, functions to relate **f** and **g**. Thus, if I have to remember the pair *ring–ladder* and I think of *rung* as a mediator, I have *ring–rung–ladder. Ring* rhymes with *rung*, and *rung* is part of a ladder. In fact, this was the basis of the keyword method of Atkinson and his colleagues (e.g., Atkinson & Raugh, 1975), and it was shown to be quite beneficial in foreign-language vocabulary learning. It is not unreasonable to assume that subjects in Hockley's experiments used mediators, because there was a slow presentation rate (one pair per 3 s) and subjects were instructed to use mnemonic aids to help them remember the pairs.

The simplest way to represent mediators is by the autoassociation of $\mathbf{f} + \mathbf{m} + \mathbf{g}$, which results in

$$(\mathbf{f}+\mathbf{m}+\mathbf{g})^{*2}$$
$$= (\mathbf{f}+\mathbf{m}+\mathbf{g})*(\mathbf{f}+\mathbf{m}+\mathbf{g})$$
$$= \mathbf{f}*\mathbf{f} + \mathbf{m}*\mathbf{m} + \mathbf{g}*\mathbf{g} + 2(\mathbf{f}*\mathbf{m}) + 2(\mathbf{f}*\mathbf{g}) + 2(\mathbf{m}*\mathbf{g}).$$

One can embed $\mathbf{f} + \mathbf{m} + \mathbf{g}$ in context, so, without a mediator, $(\mathbf{FG})^{*2}$ is

$$[\mathbf{a} \oplus (\mathbf{f}+\mathbf{g}) \oplus \mathbf{b}]^{*2};$$

with a mediator, $(\mathbf{FMG})^{*2}$ is

$$[\mathbf{a} \oplus (\mathbf{f}+\mathbf{m}+\mathbf{g}) \oplus \mathbf{b}]^{*2}.$$

Weighting coefficients are needed to normalize $(\mathbf{f}+\mathbf{g})$ and $(\mathbf{f}+\mathbf{m}+\mathbf{g})$ to 1. Recall that $v = 1/\sqrt{2}$ was used to normalize $(\mathbf{f}+\mathbf{g})$ to 1; if $w = 1/\sqrt{3}$ is used for $(\mathbf{f}+\mathbf{m}+\mathbf{g})$, then the triple will also be normalized to 1.

Why does a mediator improve recognition? Because the content part of **FMG** is $\mathbf{f}+\mathbf{m}+\mathbf{g}$; thus, with the autoassociation (i.e., $\mathbf{FMG}^{*2}$), one has not only the $\mathbf{f}*\mathbf{g}$ association as before but also the associations between the items and the mediators ($\mathbf{f}*\mathbf{m}$ and $\mathbf{m}*\mathbf{g}$). Even though the $\mathbf{f}*\mathbf{g}$ association is attenuated slightly (the weighting coefficient is now $w = 1/\sqrt{3}$ rather than $v = 1/\sqrt{2}$), this is more than compensated for by the addition of the item–mediator association.

In particular, these item–mediator associations add an additional term to the pure-associative component. The coefficient is 2 because there are two of them ($\mathbf{f}*\mathbf{m}$ and $\mathbf{m}*\mathbf{g}$). The encoding probability for the mediator (p_m) is much greater than the encod-

[10] This suggests a possible difference between item recognition and lexical decision. Perhaps subjects do use context on an item-recognition test but do not use context on a lexical-decision test.

ing probabilities for the items at study (p_s) and at test (p_r), so $p_s p_m^2 p_r > p_s^2 p_r^2$. Consequently, the additional terms contributed by the mediator benefit performance on a pair-recognition test—and hence increase $d_I'(t)$—but the amount depends on the probability of finding a mediator at study and at test.

A mediator may be more similar to an item than two items are to each other (e.g., *ring* and *rung* in comparison with *ring* and *ladder*); does this cause problems? Similarity in TODAM can increase the covariance, but that is similarity between items of different pairs, not similarity within pairs. Any two mediators are no more likely to be similar to one another than are any two items, so the similarity between a mediator and an item does not cause any troublesome covariance.

One must assume that subjects embed the pair (with or without a mediator) in context at the time of study because context might be useful for some kinds of pair-recognition tests (e.g., discriminating intact from mixed or new pairs); as long as subjects do not encode the context at the time of the tests however, the unwanted context-item component will be zero. Also, because there are pure-item components (i.e., $\mathbf{f}*\mathbf{f}$ and $\mathbf{g}*\mathbf{g}$) and context-item components (i.e., $\mathbf{a}*\mathbf{f}$, $\mathbf{a}*\mathbf{g}$, $\mathbf{f}*\mathbf{b}$, and $\mathbf{g}*\mathbf{b}$) in the memory vector, this will help performance on item-recognition tests.[11]

Parameters

One must add a few parameters for mediators. First, mediators are not guaranteed, so there must be some probability that a subject can generate a mediator at study and some probability that the subject can generate the same mediator at test. Call these σ and τ, respectively. The test probability (τ) is a conditional probability, and it must be zero if the subject does not generate a mediator at study.

There are, of course, other possibilities. A subject may generate a different mediator at test from study or a mediator at test when none was generated at study. Neither of these would help recognition; in fact, in comparison with no mediator, they would hurt recognition slightly because the normalization coefficient would be $w(1/\sqrt{3})$ rather than v $(1/\sqrt{2})$. These cases are not included here because they would complicate the analysis considerably and could probably be incorporated into the present simplified version by slight adjustments in the estimates of σ and τ.

To be consistent, one must allow for probabilistic encoding of the mediator as well as the study items. Call this p_m to distinguish it from p_s, where p_m is the encoding probability for the mediators and p_s is the encoding probability for the *A* and *B* items at the time of study. Because the mediator comes from semantic memory, it seems reasonable to assume that p_m is the same at study and test.

One must distinguish between p_s and p_r, where p_s is the encoding probability for items at study and p_r is the encoding probability for items at retrieval. Because p_s is a function of study time (Murdock & Lamon, 1988), so is p_r. This assumption is not without precedent; it was used by Hockley and Murdock (1992) to fit some speed–accuracy data that would otherwise have been problematic for their model of response times. Because a mediator is self-generated, it is assumed that p_m is considerably higher than p_s or p_r, which are the encoding probabilities for items at study or at test.

Table 1
Parameter Values for Fits Shown in Figures 10 and 11

Parameter	Value
α (forgetting)	.9998
ρ (context drift)	.97
γ (item attention at study)	.2 or (.45, .1)
κ (item attention at test)	.0
p_s (probability of encoding at study)	.2
p_m (probability of encoding for mediator)	1.0
p_r (probability of encoding at retrieval)	$\dfrac{1}{\sqrt{1+\dfrac{a}{time}}} = .95$
σ (probability of mediator at study)	.35
τ (probability of mediator at test)	.9 or .2
N_a (number of features in a)	10

Note. The value of N_a is based on a value of $N_c = 100$; all that matters is the ratio N_a/N_c ($= N_b/N_c$). Also, κ does not apply for an item probe; κ (and γ) is the amount of attention paid to single items when a test (or study) pair is presented.

Finally, just as one distinguishes between p_s and p_r, one must distinguish between attention at study (γ) and attention at test (κ). Thus, κ is the counterpart of γ when a test probe is presented. For an item probe, κ would not be relevant, and, if items rather than pairs were presented at study, γ would not be relevant.

In all there are 10 parameters, and they are listed in Table 1, along with the best fitting values. (These values are discussed shortly.) The function for p_r comes from Ratcliff (1978), and it was rationally derived from his diffusion model. Even though the underlying processes in the diffusion model are quite different from those in TODAM, encoding processes are assumed to occur in the P system and are not modeled explicitly in TODAM. It is not inconceivable that an accumulation of information that Ratcliff (1981) modeled by the mathematics of a diffusion process could be occurring in encoding.

d′ Values

The derivations necessary to compute predicted d' values as a function of the model parameters are shown in the Appendix. As explained shortly, the d' I am using is simply the mean difference between old and new items (or intact and rearranged pairs); there is no variance term in the denominator. For the differential forgetting of item and associative information, I derived the expressions for $d_I'(t)$ (item information) and $d_A'(t)$ (associative information), where t is test position. I averaged over study position to mimic the Hockley data. For the half seesaw effect, again I averaged over input (study) position but assumed a long lag (100) appropriate for a final recognition test. As noted earlier, I also assumed that the subjects used context in the test situation when the probes were single items but not when they were pairs.

If t is the test position, then, for item information in which O and N denote old and new item probes,

[11] The use of mediators is similar to that in Clark and Shiffrin (1987), who added an encoding-match assumption to SAM to boost the overall level of associative recognition performance.

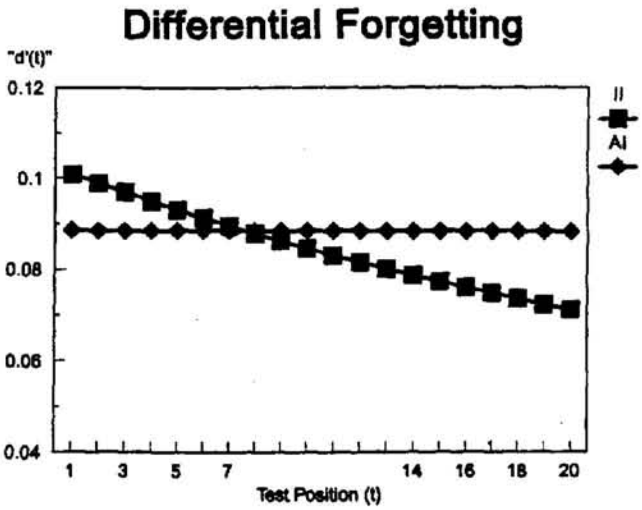

Figure 10. Differential forgetting of item (II; squares) and associative (AI; triangles) information. These data are plotted as a function of test position (t) averaged over study position, and the parameter values are shown in Table 1.

$$d_I'(t) = \mu_O(t) - \mu_N(t);$$

for associative information in which I and R denote intact and rearranged probe pairs,

$$d_A'(t) = \mu_I(t) - \mu_R(t).$$

The variance acts as a scale factor and is assumed not to differ between conditions. That is, by the continuous memory assumption, memory carries over from list to list and from the preexperimental session to the experimental session. Consequently, any difference between conditions will be trivial and can be ignored.

Results

The results are shown in Figures 10 and 11. Figure 10 shows the differential forgetting of item and associative information. Not only does the model capture the interaction, but the main effects are approximately accurate as well; overall, the absolute performance levels are similar for item and associative information. Figure 11 shows the half seesaw effect; there is little effect of instructions on the item information but a large effect on the associative information. With two exceptions, the parameters were the same for both figures and all conditions.

According to Atkinson, Bower, and Crothers (1965), parameter invariance is the sine qua non of mathematical models. The parameters should be the same for different experimental paradigms; otherwise, one is describing rather than explaining the results. Also, Hockley and Murdock (1987) noted the distinction between fixed and free parameters. Fixed parameters are constant across experimental conditions, whereas free parameters vary with experimental conditions. If one fits data with free parameters, then the parameters, rather than the model, are doing the work, and one must explain why the parameters vary as they do.

How do the present results shown in Figures 10 and 11 measure up by these two criteria? As reference to Table 1 shows, one parameter that differs between the two data sets is γ, the attention parameter at study. However, this is what the emphasis experiments manipulated, so of course it has to vary. I choose $\gamma = 0.2$ as a reasonable first approximation to how uninstructed subjects allocate attention between items ($\mathbf{f}$ and $\mathbf{g}$) and pairs ($\mathbf{f} * \mathbf{g}$) and then either 0.1 or 0.45 when they are told to remember the pairs or the items, respectively.

The other parameter is τ. As noted earlier, τ is a conditional probability, namely, the probability that a subject can retrieve a mediator at test given that a mediator was found at study. It was set to .9 for the differential-forgetting experiments in which the test phase directly followed the study phase. However, because the probability of generating the same mediator at study and test presumably decreases with lag, it must be considerably lower on a final recognition test. It is critical in determining the absolute level of performance for pairs on a final recognition test. It was set to .2 so that associative recognition for the pair-emphasis group would be comparable to item recognition for the item-emphasis group (compare Figure 11 with Figure 7). Different values of τ change the absolute level of associative recognition, but they do not seem to affect the interaction.

As a means of reducing the number of parameters somewhat, p_m was fixed at 1.0 for all conditions of both sets of experiments. Also, $p_s = N_a/N_c + N_b/N_c$, and this compensated for the fact that probabilistic encoding was applied to the study items but not the context. There were more item features than context features, but with this restriction the "functional" contribution of context and content to item recognition was roughly equated.

The value of α (.9996) is consistent with the values that were

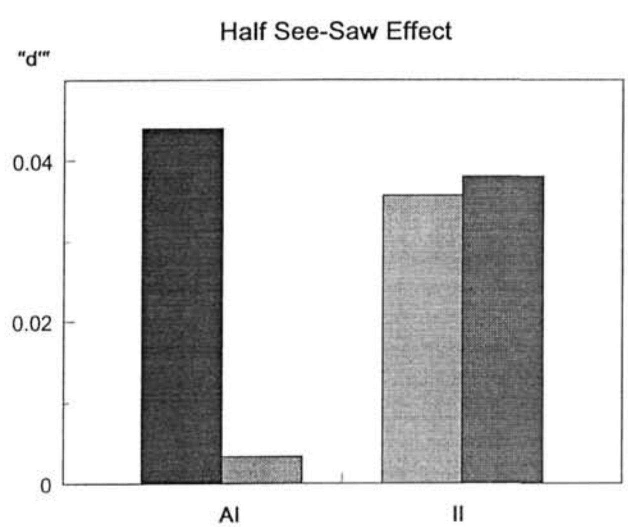

Figure 11. The half see-saw effect predicted by the model using the parameter values shown in Table 1. The coding is the same as in Figure 7. AI = associative information; II = item information.

obtained in the fits to the length–difficulty relation (Figure 4).[12] The value of ρ (.97) is still close to 1.0, but it is enough lower than α to generate the differential forgetting rates. The value of p_s is quite low (.2), but this is necessary to prevent saturation. As noted by Grossberg (1978), saturation is a major problem for nonadaptive models (i.e., models without feedback), and, with α so high, p_s must be low. This implies that N, the number of content features, must be very large, but I needed an N of 5,000 to fit the length–difficulty relation.[13]

The only other parameter that needs special mention is p_r, the encoding probability at the time of retrieval. The expression for p_r in Table 1 comes from Ratcliff (1978); it is explained further in the speed–accuracy portion of the Discussion section. It was set somewhat below 1.0 because of a recent finding by McElree and Dosher (1993). They found that subjects seem to trade accuracy for speed in a study–test paradigm; their d' values were somewhat lower than asymptotic d' values in a comparable response-deadline procedure. Presumably, subjects in the Hockley experiments could have done slightly better on their recognition tests had they taken more time to encode the probe items(s) completely.

Discussion

The revised version of TODAM is able to handle both the item–associative interaction and the half seesaw effect, albeit with a modest increase in the number of parameters. However, these parameters seem reasonable given the complexity of the experimental situations. With two necessary exceptions (γ and τ), the parameter values are the same for both data sets and were fixed across experiments and experimental conditions. Thus, for the most part, it is the model, rather than the parameters, that is doing the work.

Perhaps it should be pointed out that these two data sets were targeted because they seemed particularly problematic for TODAM. They constituted a real challenge for the model; by adding context and mediators, however, they could be solved. Of course, the model is now somewhat more complicated, but it is also more realistic. It has the potential to deal with such phenomena as list discrimination, judgments of frequency and recency, and state-dependent effects, obvious shortcomings of the original model.

$d_I'(t)$ and $d_A'(t)$

With the continuous memory assumption, the memory vector carries over from trial to trial, and, of course, so does the context. However, for both item and associative information, the predicted d' values do not change over lists. As shown in the Appendix, for item information, $d_I'(t)$ is the sum of the pure-item component ($s2$) and the context-item component ($s3$), where

$$s2 \cong 2\{\gamma + (1 - 2\gamma)[v^2(1 - \sigma) + w^2\sigma]\}p_s^2 p_r^2\, \alpha_{kt}$$

and

$$s3 \cong 4\{\gamma + (1 - 2\gamma)[v(1 - \sigma) + w\sigma]\}p_s p_r\left(\frac{N_a}{N_c} + \frac{N_b}{N_c}\right)(\alpha\rho)_{kt};$$

α_{kt} and $(\alpha\rho)_{kt}$ are functions of α and $\alpha\rho$ that vary with study position (k) and test position (t). Thus, the model correctly predicts that one should obtain lag effects that show pure recency and no buildup of proactive inhibition over trials. The same is true for associative information, but the recency effect should be much smaller because $d_A'(t)$ depends on α but not $\alpha\rho$, and α is very close to 1.0.

The numerical values of "d'" in Figures 10 and 11 are quite small; however, that is because what I am using for d' is the mean difference—that is, $\mu_o(t) - \mu_N(t)$—which is why the quotation marks are used. As noted, with the continuous memory assumption, the new- and old-item variances will be approximately equal, so "$d'(t)$" differs from a true $d'(t)$ by a scale factor. The scale factor, of course, is the variance (standard deviation).

The variance must be very small to generate realistic d' values with these parameters. However, the variance is inversely proportional to N, so one can make d' as large as one likes. The value of N (5,000) selected to fit the length–difficulty relation may be too small to give reasonable d' values, but that is not a problem. Even if it were an order or two of magnitude greater, it would still be small in comparison with the number of computational units in the brain. It means that one may have to start thinking on a very different scale. One cannot really simulate the model until desktop supercomputers are available, so one may have to depend on analytic derivations to derive testable predictions and to fit the model to data.

Encoding Probability

A problem pointed out by one of the reviewers is that the value of p_r (.95) is much greater than the value of p_s (.20). Why should encoding probability at study be so much less than encoding probability at test? Subjects know they are going to be tested on the material, so they should encode it as well as possible. Also, as this reviewer pointed out, reaction times for tests in his laboratory are generally faster than the time allotted for study.

By way of justification, a relatively low value of p_s is needed for learning (see Murdock & Lamon, 1988), so this is why p_s is set so low. Also, in the Hockley experiments, subjects were instructed to use mnemonic aids; thus, not all of the study time was devoted to encoding. Finally, for the item-recognition tests, full attention was devoted to the item, and, on associative-recognition tests, the study mediator was probably found more quickly than at study. However, it must be admitted that these counterarguments are not completely convincing, and this question certainly deserves further attention.

[12] It might be thought that, with α = .9996, a person would remember every association ever formed. However, if a person learned one new association every minute and one assumed a 16-hr day, an unrehearsed association would have dropped to 5% of its original strength in about 8 days' time.

[13] In case it is not clear, N is the number of content features or elements, N_c is the total number of context elements before the item is embedded in context, and N_a and N_b are the left- and right-hand subsets of the context vector. The $N_c - (N_a + N_b)$ center elements of the context vector are overwritten by the item elements to get $\mathbf{F} = \mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b}$. In Figure 8, $N_a = N_b = 3$, $N = 5$, and $N_c = 11$.

Forgetting

The retention curve for $d'_A(t)$ shown in Figure 10 indicates that there is essentially no forgetting of associative information over a lag range of 0–20 intervening items. Does this mean that there should be no decrease in cued recall over a comparable lag range? In fact, over lag ranges of the order of 0–10, except for a recency effect of 1 or 2 items, the retention curves for cued recall are flat (Murdock, 1963; Tulving & Arbuckle, 1963), and the recency effect could be due to working memory (Murdock, 1993).

However, this is not the whole story. Figure 10 shows the retention curve for associative recognition, but, according to TODAM, cued recall involves another stage. This additional stage is deblurring, and the probability of deblurring could well decrease with lag. Although deblurring is not formally modeled in TODAM, different effects on recognition and recall would not be inconsistent with my revision. A classic example of such differences would be the absence of retroactive inhibition in an A–B, A–D design when recognition tests are used (Dyne, Humphreys, Bain, & Pike, 1990).

Parameter Variation

TODAM2 can clearly handle the half seesaw effect (Figure 11), but what about a full seesaw effect (i.e., a crossover interaction)? One reviewer commented that he believed the revision undercut one of the chief advantages of TODAM1. As shown in Equation 2b, the attentional coefficient γ forces a full seesaw effect; the more attention paid to item information, the less attention paid to associative information. There are cases of crossover interactions; perhaps the strongest example is interactive imagery. McGee (1980) found that separate imagery instructions produced good item recognition but poor associative recognition, whereas interactive imagery instructions produced good associative recognition but poorer item recognition. However, as noted by Hockley and Cristi (1996), this is not a universal finding.

Another example might be the word-frequency effect. It has long been known that low-frequency words produce better results on an item-recognition test than high-frequency words (Shepard, 1967), and recent studies by Clark (1992) and Clark and Burchett (1994) found that high-frequency words produced better results than low-frequency words on an associative-recognition test. Clark (1992) and Clark and Burchett (1994) found a low-frequency advantage for item recognition combined with a high-frequency advantage for associative recognition; the important point is this crossover interaction. However, Hockley (1994) did not find any effect of word frequency on associative recognition.[14] In any event, can TODAM2 explain crossover interactions?

To answer this question and to provide a further understanding of TODAM2, Figure 12 shows univariate plots for the nine main parameters (α, ρ, γ, κ, σ, τ, p_s, p_r, and N_a). I have taken the parameter values shown in Table 1 and varied them one at a time, and the dependent variable is $d'_I(t)$ on the left, that is, $\mu_o(t) - \mu_N(t)$, and $d'_A(t)$ on the right, that is, $\mu_I(t) - \mu_R(t)$. I assumed a list length of 20 pairs and a probe test at $t = 20$ (i.e., Test Position 20) with the data averaged over all 20 study positions.

As can be seen, some parameters (α, p_s, and p_r) have essentially the same effect on $d'_I(t)$ and $d'_A(t)$. Other parameters (ρ, κ, τ, and N_a) have an effect on either item or associative information but not both. More specifically, two parameters (ρ and N_a) affect item information but not associative information, whereas two other parameters (κ and τ) affect associative but not item information. Finally, one parameter (σ) has opposite effects on item and associative information, and another (γ) affects item and associative information the same way but in different amounts. As σ increases, item information goes down, but associative information goes up. As γ increases, both item and associative information go down, but the latter does so much more than the former.

Thus, it would seem that TODAM2 can, in fact, predict or explain a crossover interaction. The parameter σ in TODAM2 acts like parameter γ in TODAM1, and it will provide a full seesaw effect. Although the effect is asymmetrical, this asymmetry is, in fact, consistent with some data (e.g., McGee, 1980). Furthermore, it is reassuring that the relevant parameter is σ, because this is the probability of a mediator. Assuming that interactive imagery operates something like a mediator, this means that the McGee (1980) result falls directly out of the model. With separate imagery instructions, subjects do not try to interrelate the A and B items, so σ is low if not zero; with interactive imagery, however, subjects do try to relate the A and B items, so σ is high. Consequently, item recognition is better under the separate imagery condition than under the interactive imagery condition, but associative recognition is better under the interactive imagery condition than under the separate imagery condition.

Not only does TODAM2 have the potential to explain crossover interactions, it can also explain "one-way" interactions (e.g., the Hockley word-frequency results) and main effects with no interactions. Of course, univariate analyses do not begin to capture the full complexity or power of a multiparameter model such as this, but at least they show that the model is capable of explaining a variety of different results. TODAM2 is more complicated than TODAM1, but unfortunately that seems necessary given the variety of experimental results in the field.

One could argue that with this varied parameter pattern, the model could explain any result; however, this is not true. TODAM is clearly falsifiable. The differential forgetting and the half seesaw effect were both counter to TODAM1, and these were what led to the revision. In TODAM2, parameter variation generally is not required to explain the results; the model, rather than the parameters, provides the explanations. Also, any parameter variation that does track an experimental manipulation must be logically consistent with the experimental manipulation (e.g., σ and the McGee, 1980, results). If the only parameters that allowed the model to fit data generated by some test manipulation were study parameters (e.g., ρ, σ, or N_a), this would hardly be confirmation of the model.

Finally, it should be stressed that I have not simply assumed what needs to be explained. Starting with the basic assumptions of TODAM (representation, storage, and retrieval), I added mediators and then worked out the basic derivations of the model

[14] A possible reason for the discrepancy is that Clark used very-low-frequency words (four per million).

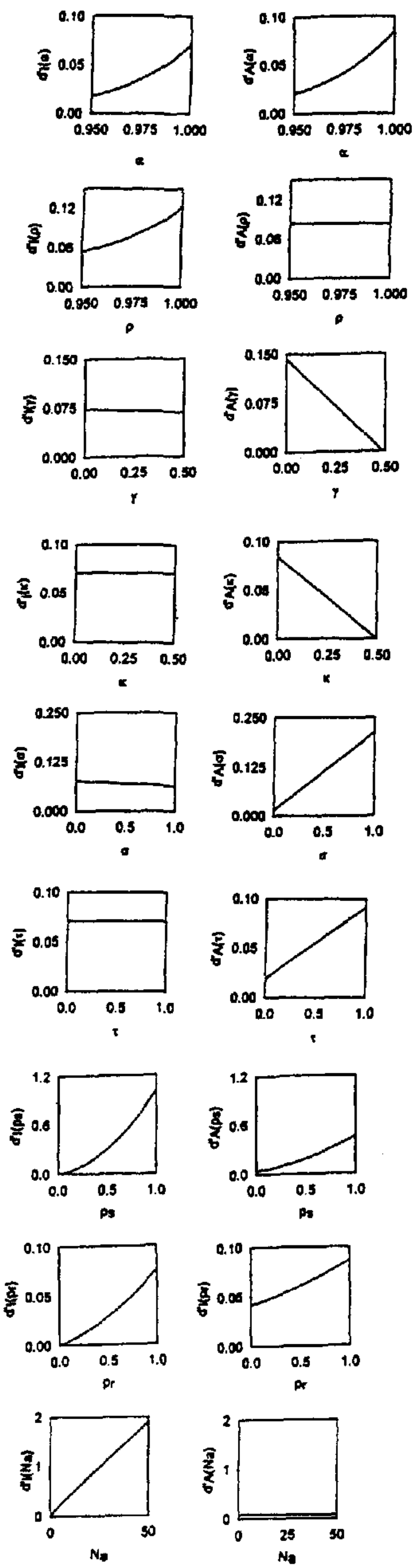

Figure 12. Univariate plots for the nine model parameters (α, ρ, γ, κ, σ, τ, p_s, p_r, and N_a) with $d'_I(t)$ on the left and $d'_A(t)$ on the right. All parameter values were fixed at the values shown in Table 1 except the one under analysis, and that was varied over a wide range.

in terms of the model parameters. I then estimated reasonable parameter values to fit the two problematic Hockley data sets. Finally, I conducted the univariate analysis (Figure 12) showing that the effects of each of these nine parameters varied one at a time. This univariate analysis came at the end, not at the beginning.

Recognition and Recall

The relation between item recognition and cued recall can be described by the Tulving–Wiseman function (Tulving & Wiseman, 1975), which plots the conditional probability of recognition given recall as a function of the unconditional probability of recognition. However, as noted by Hintzman (1992), there are constraints on the function, and a scatter plot shows that recognition and recall are largely independent. How does TODAM handle this?

In TODAM1 (Equation 2c), item and associative information are independent because

$$E(\mathbf{f}\cdot\mathbf{g}) = E[\mathbf{f}\cdot(\mathbf{f} * \mathbf{g})] = E[\mathbf{g}\cdot(\mathbf{f} * \mathbf{g})] = 0.$$

Consequently, one would not expect any correlation between recognition and recall. The small correlation implied by the Tulving–Wiseman function could be explained by parameter variation across subjects (but see Metcalfe, 1991, for a contrary view). The same interpretation could apply to TODAM2, although the addition of context might affect the relation somewhat. However, a deeper analysis by Kahana (1997) shows that variability in probabilistic encoding could have a large effect on the correlation not only in TODAM but in several other distributed memory models as well.

Speed–Accuracy Trade-Off

As can be seen in Table 1, encoding probability at retrieval (p_r), like encoding probability at study (p_s), is a function of encoding time. That is, up to a point, the more time one has at test, the more features of the probe one can encode before the comparison process; the function relating p_r to encoding time shown in Table 1 comes from Ratcliff (1978). The speed–accuracy trade-off curve shows accuracy as measured, say, by d' as a function of encoding time, and the predicted curve is shown in Figure 13. This curve is in reasonable agreement with experimental results reported by Gronlund and Ratcliff (1989) and Reed (1976).

To generate the predicted curve, I simply took the parameter values shown in Table 1 but computed p_r by the Ratcliff function shown there. Then I simply varied time to sweep out the curve. The slope of the function is determined by the value of the parameter a, and here $a = 0.1$. The time interrupt is shown here as 0, but this could easily be changed to allow for the time taken by other processes (e.g., sensory processes or decision and response processes) without otherwise affecting the speed–accuracy trade-off function.

An analysis of the trade-off function for associative information would be somewhat more complex, and more parameters would be involved. Not only is more time needed to encode two items than one, but there is also the mediator to consider. Consequently, the intercept, slope, and asymptote of the associative trade-off function could all differ from the item trade-off

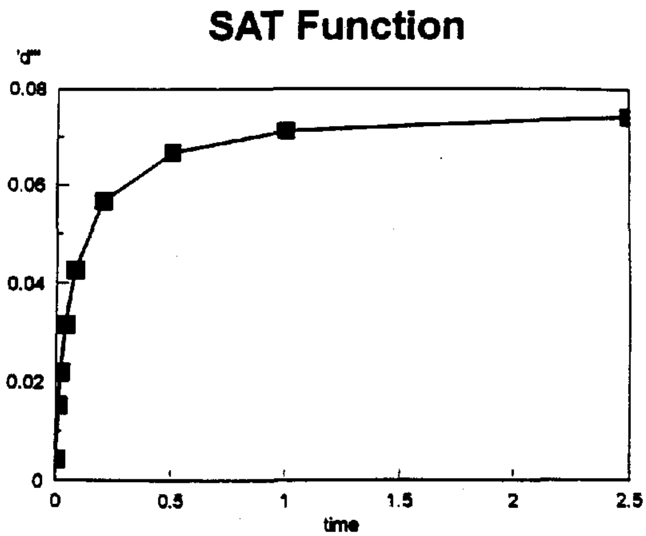

Figure 13. Speed–accuracy trade-off (SAT) function generated by the model when p_r is a function of encoding time at test. The parameters and the function relating p_r to time are shown in Table 1, and this function comes from Ratcliff (1978).

function. Obviously, this all has to be worked out and tested, but at least the revised version of TODAM is able to account qualitatively for the speed–accuracy trade-off function for item information without any additional modifications.

In the remainder of this section, I discuss some experimental studies of the effects of context, some problematic findings (the mirror effect and the slopes of ROC curves), and the relation between the revision and TODAM2. I end with a comparison of the revision of TODAM2 with the other global-matching models.

Experimental Effects of Context

Bjork and Richardson-Klavehn (1989) analyzed context effects in recall and distinguished four types of effects: (a) physical reinstatement of study context at test, $AA > AB$; (b) imaginal reinstatement of study context at test, $AB(A) > AB$; (c) varied contexts across study sessions and Recall List 1 in a neutral context, $ABN > AAN$); and (d) varied context across study sessions and Recall Lists 1 and 2 in a neutral context, $ABN > AAN$. The symbols denote same context (e.g., AA), varied context (e.g., AB), or neutral context (N); see Bjork & Richardson-Klavehn, 1989, p. 317, Table 9.1). These authors suggested that although the final three effects are relatively clear and not in dispute, the first effect is very much in dispute (e.g., Fernandez & Glenberg, 1985), and "these data suggest the possibility that a number of the significant effects of physical reinstatement in the literature represent Type I errors" (Bjork & Richardson-Klavehn, 1989, p. 319). If one accepts this suggestion, my context analysis may help to explain the pattern of results.

Even though the data analyzed here involved recall rather than recognition, given TODAM's recall-recognition identity (e.g., Weber, 1988), similar results would be expected for recall and recognition. If $\mathbf{g}'$ is the retrieved information when the probe ($\mathbf{f}$) is correlated with the memory pair ($\mathbf{f}*\mathbf{g}$), the identity is

$$E(\mathbf{g}\cdot\mathbf{g}) = E\{\mathbf{g}\cdot[\mathbf{f}\#(\mathbf{f}*\mathbf{g})]\} = E[(\mathbf{f}*\mathbf{g})*(\mathbf{f}*\mathbf{g})].$$

The similarity of the retrieved information to the target item ($\mathbf{g}$), as measured by the dot product (for recall), is the same as what one would have in a recognition test in which one compared the convolution of the two probe items ($\mathbf{f}$ and $\mathbf{g}$) with the memory pair. Recall and recognition are not identical because there is another process involved in recall (deblurring, bringing the retrieved information $\mathbf{g}'$ into focus); however, this analysis shows that recognition results should probably follow the same pattern as recall.

I have already suggested that the use of context at the time of test is optional; for the Hockley data, it is used for item recognition because it enhances the discriminability of old and new items, but it is not used in pair recognition because, in this particular case, it does not provide any useful information. If, for instance, the pair-recognition tests had included mixed and new pairs (as in Clark & Shiffrin, 1987, or Humphreys, 1976), then presumably context would have been used because it could have helped to discriminate mixed and new pairs from intact pairs. It does not seem unreasonable to suggest that, in recall, subjects will use context if it is helpful but will not do so if it is harmful.

For the first Bjork and Richardson-Klavehn effect, context would not be useful because roughly the same context is associated with all pairs, and thus item discrimination (i.e., discriminating one possible response from another) would be harder with context. If the pairs were A–B, C–D, and E–F, there would be cross coupling based on context among B, D, and F; thus, if the probe was A, the discriminability of B from D and F would be adversely affected. On the other hand, the final three effects all involve list discrimination, and there context could be useful. Therefore, by this analysis, beneficial effects of context would be expected for the later three cases but not for the first.

As Bjork and Richardson-Klavehn (1989) pointed out, it is also necessary to distinguish integrated, influential, and incidental aspects of context. An example of the beneficial effects of integrated context would be cued versus uncued recall of word pairs (Thomson & Tulving, 1970). An example of influential context would be the money–lend–bank versus river–stream–bank example of Light and Carter-Sobell (1970). Both of these effects are consistent with the analysis here, because they are examples of the encoding-specificity principle discussed shortly. Incidental context effects (e.g., same or different experimental room) are much more variable, but that is because the use of background context at the time of test is optional rather than obligatory.

I have assumed that context is not used in associative recognition, but this would seem to be flatly contradicted by the data of Murnane and Phelps (1993), who showed, in several experiments, that changes in context lower both the hit rate and the false-alarm rate.[15] Their context manipulation was within lists, and there were four different learning context conditions involving differences in foreground color, background color, and loca-

[15] A criterion change seems unlikely; see the original article for details.

tion on a computer screen (see Murnane & Phelps, 1993, p. 884, Table 1). For associative recognition (Experiments 1–3), intact and rearranged pairs were tested in the same or a different context; the different context was yet another combination of foreground color, background color, and location on the computer screen. For rearranged pairs, both members of the pair had been presented in the same study context, so same or different context was not a useful cue for the intact–rearranged decision.

How can TODAM2 explain this? One possibility is that subjects thought context was a useful cue, even though it was not. Subjects were informed that screen colors and locations might change, that the display characteristics were probably quite salient (see Murnane & Phelps, 1993, p. 884, Table 1), and that the different context condition was quite unlike the study context conditions, which could reinforce this mistaken idea. Another possibility is that perhaps these display characteristics should be considered as part of the stimulus (or content) features and not as context features. If so, then one would expect lower hit rates and false-alarm rates because the test items would be less similar to the encoded study items under the different context conditions than under the same context condition, and, in TODAM, it is these encoded test items that go into the convolution to form the association that is compared with the memory vector to form the basis for decision.

In his review of the effects of environmental context (EC), Smith (1988) noted that "EC-dependent memory [is] generally found with recall, but not with recognition testing" (p. 19). This would seem to be exactly backward to what I have suggested here (i.e., that context is used in recognition but not in recall). However, this is not quite right. I have stated that, in the Hockley experiments, this is what happened, but it is not a general principle. As I have just pointed out, context will or will not be used depending on whether it is useful. Thus, in the Murnane and Phelps experiment, context should have been (and apparently was) used in associative recognition (and also in later experiments by the same authors; see Murnane & Phelps, 1994, 1995).

Still, the fact that EC effects in recognition memory are conspicuous by their absence is puzzling. One possibility (mentioned by Smith in another context) was that the context was not encoded at the time of original learning, so the fact that context did not have an effect does not necessarily mean it was not used at the time of test. Also, as Smith pointed out, there is abundant anecdotal evidence that context affects recognition memory. The question is why the experimental tests have been so negative. Perhaps experiments are poor simulations of the real world (Bjork & Richardson-Klavehn, 1989). In any event, as Smith pointed out, there is much that needs to be known about context, and how to produce reliable and dramatic EC effects in the laboratory heads his list (Smith, 1988, p. 28).

Problematic Findings

The mirror effect. The mirror effect is a major problem for global-matching models in general and TODAM in particular (global-matching models are discussed later). The mirror effect (Glanzer & Adams, 1990) refers to the fact that hit rates and correct rejection rates mirror one another across conditions. Thus, if condition A has a higher hit rate than condition B, condition A will also have a higher correct-rejection rate than condition B. The interpretation suggested by Glanzer and his colleagues (e.g., Glanzer, Adams, Iverson, & Kim, 1993) is that the four underlying strength distributions are aligned as shown in Figure 14, and this alignment would certainly produce a mirror effect.

The strongest evidence for the distributional analysis shown in Figure 14 is the forced-choice inequalities. If N is new and O is old, then the predicted percentage correct on a forced-choice test pitting one condition against another should be $AN/AO > AN/BO = BN/AO > BN/BO$, where the slash indicates the comparison. Also, assuming perfect symmetry, the null choices (AN/BN and BO/AO) should be equal and greater than chance (.5). As shown in several articles by Glanzer and his colleagues (e.g., Glanzer et al., 1993), this is typically the way the data work out.

Actually, it is the distributional alignment, not the mirror effect per se, that is the problem for the global-matching models. All of the models assume a common new-item distribution that should be the same for both conditions, and, if this is so, how could the mirror effect come about? One possibility (differential learning on test trials) was suggested by Hintzman, Caulton, and Curran (1994), but unfortunately it was not supported by their data.

Another possibility (a criterion or bias effect) was suggested by Greene (1994, 1996). In response to a question about the forced-choice data, he wrote that

> One can explain the forced-choice data in about the same way that one explains the yes-no data. On yes-no tests, subjects try to equate the number of times they say "yes" to each class of stimuli. On forced-choice tests, subjects try to equate the number of times they pick items from each category; this introduces a bias in favor of the less-memorable category on trials containing an item from each of the two categories. (R. L. Greene, personal communication, November 19, 1994)

To follow up Greene's suggestion, I worked out the following analysis: Let A be the probability that the subject remembers the stronger item (i.e., the item from condition A) and B be the probability that the subject remembers the weaker item, so $A > B$. Assume a response bias Δ to pick the less memorable category (i.e., B) on mixed trials. Let y_i be the probability of choos-

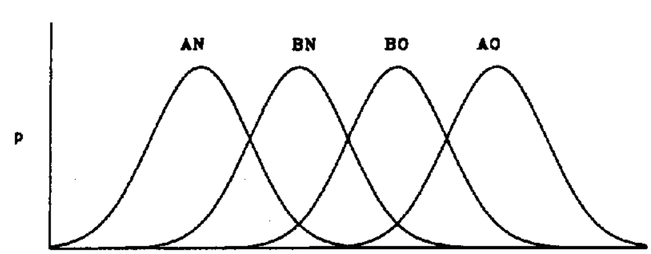

Figure 14. Strength distributions to illustrate the mirror effect. A and B are the stronger and weaker conditions, respectively, and the probe types are new (N) or old (O). Adapted from "The Regularities of Recognition Memory," by M. Glanzer, J. K. Adams, G. J. Iverson, and K. Kim, 1993, *Psychological Review, 100,* p. 547. Copyright 1993 by the American Psychological Association. Adapted with permission of the author.

ing the stronger item (i.e., A) in condition i, where i is one of the six forced-choice conditions. Then, with the six conditions ordered as shown in Table 2,

$$y_1 = A + .5(1 - A),$$

$$y_2 = B + (.5 + \Delta)(1 - B),$$

$$y_3 = A + (.5 - \Delta)(1 - A),$$

$$y_4 = B + .5(1 - B),$$

$$y_5 = .5 + \Delta,$$

and

$$y_6 = AB(.5 - \Delta) + A(1 - B) + (1 - A)(1 - B)(.5 + \Delta).$$

Note that there is nothing here stipulating that $AN < BN$; for old items $A > B$, and there is a guessing bias (Δ), but that is all.

The predicted forced-choice probabilities (i.e., y_i) for four values of Δ are shown in Table 2 for $A = .7$ and $B = .3$. (These values of A and B were suggested by Greene.) The values shown below each column are the overall proportion of times subjects pick the less memorable condition. It is easy to show that subjects would equalize the responses overall if they set $\Delta = .297$, but that would not produce a very good mirror effect (the ordering would be $1 = 2 > 3 > 4$). However, suppose that subjects are conservative (as it is known they are) and set $\Delta = .2$; this would be a perfect mirror effect, and their overall bias would be .467, not far from the ideal .5.

One could object to this analysis on the grounds that it assumes subjects either remember an item or they do not, and countless applications of signal-detection theory to episodic memory tasks have shown that this is wrong. But signal-detection theory assumes a criterion, and a given item is either above ("remembered") or below ("not remembered") the criterion. Obviously many more details need to be worked out, but I think this bias interpretation is certainly a possible explanation of the data.

A conservative conclusion would be that it may be possible to explain the mirror effect by augmenting strength differences for old items with a bias parameter to equalize response choices, and this analysis shows one way that it could be done. It is not the case that the mirror effect mandates the underlying distributions shown in Figure 14. There are other ways the mirror effect could come about. So for the present, at least, one does not have to revise global-matching models (including TODAM) to conform to the situation depicted in Figure 14.

Table 2
Predicted Values of the Six Forced-Choice Conditions as a Function of the Bias Parameter (Δ)

i	Condition	Stronger	$\Delta = 0$	$\Delta = .1$	$\Delta = .2$	$\Delta = .297$
1	*AN/AO*	*A*	.85	.85	.85	.85
2	*AN/BO*	*B*	.65	.72	.79	.858
3	*BN/AO*	*A*	.85	.82	.79	.761
4	*BN/BO*	*B*	.65	.65	.65	.65
5	*AN/BN*	*B*	.5	.6	.7	.797
6	*BO/AO*	*A*	.7	.7	.7	.7
	Weaker		.40	.433	.467	.5

ROC slopes. Recent studies by Ratcliff, Sheu, and Gronlund (1992) and Ratcliff, McKoon, and Tindall (1994) have shown that, in item-recognition studies, the slopes of Z-transformed ROC functions are generally less than 1.0. On average, they are about 0.8, but they do not seem to vary with item strength when item strength is manipulated in a variety of ways. This is a problem for TODAM, because these slopes give the ratio of standard deviations (new to old) of the underlying distributions, and, as pointed out by these authors, the variances should be approximately equal, so the slopes for both item and pair recognition should be close to 1.0. (Unlike other global-matching models, TODAM does predict that the slopes should not vary with strength, but here I am primarily concerned with the slope ratio.)

Their analyses apply to the original version of TODAM (Equation 2c); what about the revision (Equation 3c)? Unfortunately, the revision does not seem to handle this problem any better than the original version did. As far as I can tell, the revised model also predicts that the slope should be about 1.0. The reason is the continuous memory assumption. Any differences between the old- and the new-item variances will be trivial if the contents of memory are not restricted to the current list, and the addition of context would not seem to affect this.

To check these intuitions, I ran a fairly extensive simulation with the parameter values shown in Table 1 for a number of study–test trials in which the memory vector carried over from trial to trial. The hope was that I had overlooked something and, perhaps, over blocks of trials, the old-item variance would become somewhat larger than the new-item variance. For a reasonable list length, if one runs many trials, one needs a large value of N. The reason is that, in a distributed memory model, the total number of items stored in memory should be less than the dimensionality of the space because a basic principle of linear algebra is that there can be N independent non-null vectors in an N-dimensional vector space. To be safe, N should probably be an order of magnitude larger than the total number of items in the simulation. The simulation I ran used an N of 3,840, an N_c of 4,800, and 20 study–test trials with a list length of 20. This is still a rather small-scale simulation, yet it took more than 5 days on a SPARC2 using the TODAM program library (a set of routines written in C by David Mitchell and available from the author on request). There was nothing in the results that gave me any reason to believe I could reject the null hypothesis.

So, it would seem that the ROC problem is still present. There undoubtedly are patches that would fix this particular problem, but then the differential forgetting or the half seesaw effect might be lost. The problem is to find a solution that will preserve the gains made.

TODAM1 and TODAM2

The revised version of TODAM (Equation 3c) is clearly more elaborate than the original version (Equation 2c), but that is to be expected. The original version is almost 15 years old, and many new findings have emerged. It would have been nice if all of these new findings could be accommodated by the original version, but that is rather unrealistic. TODAM was an early distributed memory model and, although these distributed memory models had many advantages over extant models (see Anderson, Silverstein, Ritz, & Jones, 1977), their application to experimental data was far from clear.

TODAM2 (Murdock, 1993) is a general version of TODAM that is designed to deal with item, associative, and serial-order information in a unified manner. The basic building block is the chunk. A chunk is the sum of n grams, and an n gram is the n-way autoconvolution of the sum of n item vectors. That idea is exploited here for associative information; **FG** is $(\mathbf{f} + \mathbf{g})^{*2} = (\mathbf{f} + \mathbf{g})*(\mathbf{f} + \mathbf{g})$ embedded in context. With mediators, **FMG** $= (\mathbf{f} + \mathbf{m} + \mathbf{g})^{*2} = (\mathbf{f} + \mathbf{m} + \mathbf{g})*(\mathbf{f} + \mathbf{m} + \mathbf{g})$ embedded in context. Concatenated vectors, such as $\mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b}$ or $\mathbf{a} \oplus (\mathbf{f} + \mathbf{g}) \oplus \mathbf{b}$, were not used in TODAM2, but context was not included in TODAM2, and the autoassociation of concatenated item vectors is consistent with the use of autoassociation in TODAM2.

Autoassociated concatenated item vectors provide an efficient and powerful mechanism to implement the encoding-specificity principle of Tulving (e.g., Tulving & Thomson, 1973) in a distributed-memory model. If $\mathbf{F} = \mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b}$, then $\mathbf{F}^{*2} = \mathbf{F}*\mathbf{F}$ is the autoconvolution or autoassociation of concatenated item vectors, and the dot product of $\mathbf{F}*\mathbf{F}$ with $\mathbf{F}*\mathbf{F}$ is the (unlikely) case in which the probe matches an item in the memory vector exactly. Now the expected value of this dot product is

$$\begin{aligned} E(\mathbf{F}^{*2}\cdot\mathbf{F}^{*2}) &= E[(\mathbf{F}*\mathbf{F})\cdot(\mathbf{F}*\mathbf{F})] \\ &= \{[(\mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b}) * (\mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b}] \\ &\quad \cdot[(\mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b}) * (\mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b})]\} \\ &= E\{[\mathbf{a}*\mathbf{a} + \mathbf{f}*\mathbf{f} + \mathbf{b}*\mathbf{b} + 2(\mathbf{a}*\mathbf{f}) \\ &\quad + 2(\mathbf{a}*\mathbf{b}) + 2(\mathbf{f}*\mathbf{b})]\cdot[\mathbf{a}*\mathbf{a} + \mathbf{f}*\mathbf{f} \\ &\quad + \mathbf{b}*\mathbf{b} + 2(\mathbf{a}*\mathbf{f}) + 2(\mathbf{a}*\mathbf{b}) + 2(\mathbf{f}*\mathbf{b})]\} \\ &= E[(\mathbf{a}*\mathbf{a})\cdot(\mathbf{a}*\mathbf{a})] + E[(\mathbf{f}*\mathbf{f}) \\ &\quad \cdot(\mathbf{f}*\mathbf{f})] + E(\mathbf{b}*\mathbf{b})\cdot(\mathbf{b}*\mathbf{b})] \\ &\quad + 4E[(\mathbf{a}*\mathbf{f})\cdot(\mathbf{a}*\mathbf{f})] + 4E[(\mathbf{a}*\mathbf{b}) \\ &\quad \cdot(\mathbf{a}*\mathbf{b})] + 4E[(\mathbf{f}*\mathbf{b})\cdot(\mathbf{f}*\mathbf{b})], \end{aligned}$$

because there are no cross products. More generally, if we symbolize concatenated item vectors by K so just as

$$\sum_{i=1}^{n} \mathbf{f}_i = \mathbf{f}_1 + \mathbf{f}_2 + \cdots + \mathbf{f}_n$$

we have for concatenated item vectors

$$\mathop{\mathrm{K}}_{i=1}^{n} \mathbf{f}_i = \mathbf{f}_1 \oplus \mathbf{f}_2 \oplus \cdots \oplus \mathbf{f}_n;$$

the general principle is

$$E\left[\left(\mathop{\mathrm{K}}_{i=1}^{n} \mathbf{f}_i\right)^{*2}\cdot\left(\mathop{\mathrm{K}}_{i=1}^{n} \mathbf{f}_i\right)^{*2}\right] = \sum_{i=1}^{n} E[(\mathbf{f}_i*\mathbf{f}_i)\cdot(\mathbf{f}_i*\mathbf{f}_i)] + 4\sum_{i<j}^{n} E[(\mathbf{f}_i*\mathbf{f}_j)\cdot(\mathbf{f}_i*\mathbf{f}_j)].$$

This principle is illustrated in Figure 15 with the matrix representation of convolution. The probe and the memory vector are shown in the two partitioned matrices, and their dot products are their projection on the diagonal line. Even though elements of adjacent components in the convolution of the concatenated item vectors will overlap in $\mathbf{F}*\mathbf{F}$ (see Figure 1 as a reminder that the matrix representation is only an intermediate step in the formation of a convolution), this overlap is sorted out by the comparison process (dot product).

Now the probe and the memory vector are never identical; if nothing else, the context vectors **a** and **b** are always at least slightly and perhaps considerably different. However, this does not invalidate the general principle, because if **a** is similar to $\mathbf{a}'$ by ρ, then, for instance,

$$E[(\mathbf{a}*\mathbf{a})\cdot(\mathbf{a}'*\mathbf{a}')] = 2\rho^2$$

(see Equation 4). Thus, ρ provides a quantitative measure of the context similarity, and the same approach can be used with item vectors (Murdock, 1995). So, in addition to implementing the encoding-specificity principle, one has a measure of similarity that varies continuously between 0 and 1. Because this is a distributed memory model, however, it indicates the similarity between the probe item and all of memory, not just the target item.

Finally, all of this applies to associative information as well as item information. Without a mediator, one has $(\mathbf{f} + \mathbf{g})^{*2} = \mathbf{f}*\mathbf{f} + \mathbf{g}*\mathbf{g} + 2(\mathbf{f}*\mathbf{g})$, and these all coexist in the center submatrix of Figure 15. With mediators, one has $(\mathbf{f} + \mathbf{m} + \mathbf{g})^{*2} = \mathbf{f}*\mathbf{f} + \mathbf{m}*\mathbf{m} + \mathbf{g}*\mathbf{g} + 2(\mathbf{f}*\mathbf{m}) + 2(\mathbf{f}*\mathbf{g}) + 2(\mathbf{m}*\mathbf{g})$, and these too all coexist in the center submatrix. This applies only to a single pair, and, by Equation 3c, all pairs are combined in the memory vector **M**. If one wonders how this can be, it is reassuring to remember that simulation shows that it works. It is not surprising that people's memory is not perfect; the wonder is that it is as good as it is.

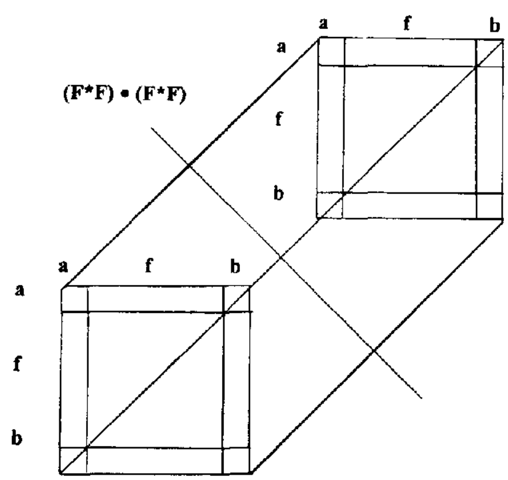

Figure 15. The dot product of two concatenated item vectors $(\mathbf{F}*\mathbf{F}\cdot\mathbf{F}*\mathbf{F})$ as the projection on the diagonal line (really a plane) of the diagonals $(\mathbf{a}*\mathbf{a}\cdot\mathbf{a}*\mathbf{a}$, $\mathbf{f}*\mathbf{f}\cdot\mathbf{f}*\mathbf{f}$, and $\mathbf{b}*\mathbf{b}\cdot\mathbf{b}*\mathbf{b})$ plus four times the off-diagonals $(\mathbf{a}*\mathbf{f}\cdot\mathbf{a}*\mathbf{f}$, $\mathbf{a}*\mathbf{b}\cdot\mathbf{a}*\mathbf{b}$, and $\mathbf{f}*\mathbf{b}\cdot\mathbf{f}*\mathbf{b})$.

Comparison With Other Models

In this final section, I compare the revised version of TODAM with other global-matching models, in particular the CHARM model of Metcalfe-Eich (1982, 1985), the SAM model of Gillund and Shiffrin (1984), the MINERVA2 model of Hintzman (1986, 1988), and the matrix model of Humphreys, Pike, Bain, and Tehan (1989). These four models are, like TODAM, global-matching models; that is, one way or another, all list items enter into the comparison process.

CHARM. CHARM (a composite holographic associative recall model) and TODAM are both offshoots of the CADAM model of Liepa (1977), so they are quite similar. They use the same random-vector item representation, and they both use convolution to represent associative information and correlation to retrieve it. However, they differ in how they model item information. The CHARM counterpart to Equation 2a is

$$\mathbf{M}_j = \mathbf{M}_{j-1} + \mathbf{f}_j*\mathbf{f}_j + \mathbf{g}_j*\mathbf{g}_j + \mathbf{f}_j*\mathbf{g}_j.$$

Thus, CHARM models item information by autoconvolution, whereas the original version of TODAM used single-item vectors (see Equation 2a).

Now that the revised version of TODAM models item information by autoconvolution, are CHARM and TODAM the same? No; the revised version of TODAM models associative information by $(\mathbf{f} + \mathbf{g})^{*2}$ (Equation 3c), and this is different. Although the expansion of $(\mathbf{f} + \mathbf{g})^{*2}$ includes an $\mathbf{f}*\mathbf{g}$ term, it also includes $\mathbf{f}*\mathbf{f}$ and $\mathbf{g}*\mathbf{g}$ terms, which constitute the dual bases for item recognition necessary to explain the half seesaw effect. Also, CHARM does not have a forgetting parameter (α), but it could easily be added.

SAM. Like TODAM and CHARM, the SAM (search of associative memory) model of Gillund and Shiffrin (1984) is a model for the storage and retrieval of item and associative information. An extension of the buffer model of Atkinson and Shiffin (1968), it represents information in memory by images, and list presentation sets up strength values between potential retrieval cues (including context) and these images. The four basic parameters are context strength (a), associative strength (b) between pairs of items in the buffer, self-strength (c) for each individual item, and background strength (d) which functions essentially as noise. The basic familiarity or strength equation can be represented as

$$\text{Strength} = \sum\prod(\text{Cues}\cdot\text{Images}),$$

where the summation (Σ) is over all images and the product (Π) is over all (retrieval) cues. The summed strength reflects the parameter values and serves as the basis for decision in a recognition task.

Although SAM keeps the images separate in storage and so, unlike TODAM, is a "localist" model, all of the images are summed in the comparison process; thus, SAM, like TODAM, is a global-matching model. Also, the revised version of TODAM has the pure-context, pure-item, and pure-associative components that seem similar to the first three parameters of SAM. SAM does not have a parameter corresponding to the context-item component of TODAM, but the item and context parameters of SAM do combine multiplicatively at the time of retrieval. Perhaps the continuous memory assumption could underlie the background strength parameter of SAM.

Although there are clear similarities between SAM and the revised version of TODAM, there are many differences as well. SAM is a computational model (given the numerical values of the parameters, the experimenter can compute the predicted values; however, except for the rehearsal-buffer notion, the model does not specify what the underlying storage and retrieval processes are). It does not have the detailed representation assumptions of TODAM, nor is there any counterpart to the storage (convolution) and retrieval (correlation) processes of TODAM. The TODAM notion that the item vectors are embedded in an ever-changing context has no counterpart in SAM, nor does the autoassociation operation that binds items to context and also underlies the item-to-item association.[16]

MINERVA2. The MINERVA2 model of Hintzman (1986, 1988), like the revised version of TODAM, uses partitioned vectors; the fact that the features are binary ($+1$ or -1) is a minor difference, because they are still random variables.[17] In MINERVA2, strength ("summed activation") is given by

$$\text{Strength} = \textstyle\sum_i(\text{Probe}\cdot\text{Memory})^3.$$

That is, the probe item is compared with each item in memory by a normalized dot product, this dot product is cubed ("echo intensity"), and the echo intensity is summed over all items in memory. The cubing operation binds the *A* and *B* members of a pair together and, as such, is the counterpart to autoassociation in the revised version of TODAM.

MINERVA2, like SAM, is a localized memory model because each item (or pair of items) is stored in a separate memory location but, like SAM and TODAM, is a global-matching model because all items (or pairs of items) enter into the comparison process. The representation of similarity is different; with binary features, MINERVA2 must use Haming distance (number of mismatching features; see, e.g., Abramson, 1963) as a measure of (dis) similarity, whereas TODAM (but not CHARM) uses correlated random vectors. Unlike TODAM, MINERVA2 has an explicit deblurring algorithm (see Hintzman, 1988), but this would probably not work in TODAM because of the normal (as opposed to binary) feature distribution.

MINERVA2, like TODAM, is a process rather than a computational model. Although it does not have the four explicit parameters of SAM, it has context features and item features, so one could probably obtain the pure-context, pure-item, context-item, and pure-associative components by the appropriate retrieval operations. What is not clear is how one would get mediators. Is it the concatenation of context, *A* and *B* item vectors, and the mediators? How would the model handle the single-item case at the same time? In the revised version of TODAM, this is

[16] The idea of changing context is very clearly included in an extension of the original SAM model proposed by Mensink and Raaijmakers (1988).

[17] In MINERVA2, features can also have a value of zero as a result of encoding failures or forgetting, and probabilistic encoding in TODAM seems to be the counterpart to encoding failures in MINERVA2. However, the forgetting mechanisms in TODAM and MINERVA2 are different; all features are attenuated by α in TODAM, but some features are set to zero at the time of study in MINERVA2.

handled by summation (of items, associations, and mediators) in a common memory system (the Q system), so proliferation is no problem. How this might work in MINERVA2 remains to be seen.

Matrix model. The matrix model of Humphreys and his colleagues (Humphreys, Pike, et al., 1989) is formally very similar to the linear-associator model of Anderson (e.g., Anderson et al., 1977), but it combines item and associative information in a common memory system. For pairs of items, the memory matrix holds both the outer product ($\otimes$) of the two items and the outer product of each item separately with a vector of ones ($\mathbf{r}$). Successive pairs are combined by superposition, so

$$\mathbf{M}_j = \mathbf{M}_{j-1} + \mathbf{f}_j \otimes \mathbf{r} + \mathbf{r} \otimes \mathbf{g}_j + \mathbf{f}_j \otimes \mathbf{g}_j.$$

This is similar to the original TODAM formulation (Equation 2a) except that it uses the outer product rather than convolution, and TODAM has no vector of ones ($\mathbf{r}$).

For context ($\mathbf{x}$), the matrix model uses a three-way outer product $\mathbf{x} \otimes \mathbf{f} \otimes \mathbf{g}$, and this is quite different from the revised version of TODAM. In TODAM, I embed (concatenate) items in context and then autoassociate them for binding, so the result is a two-way convolution ($\mathbf{F}*\mathbf{F}$, $\mathbf{G}*\mathbf{G}$, or $\mathbf{FG}*\mathbf{FG}$) rather than a three-way outer product (i.e., $\mathbf{x} \otimes \mathbf{f} \otimes \mathbf{g}$). The need for binding items to context was clearly documented in Humphreys, Bain, and Pike (1989), but it is done quite differently in the revised version of TODAM and the matrix model of Humphreys and his colleagues.

As in MINERVA2, it is not clear how one would add mediators. Would it require a four-way outer product? For an n-way outer product of N-dimensional item vectors, the number of elements required is N^n; thus, if $N = 1{,}000$ (and TODAM may need more than that), a four-way outer product would require 10^{12}, and that is about the number of neurons in the brain.[18] Also, unlike convolution, matrix multiplication is not commutative, so that might be a problem too because backward associations exist and can be as strong as forward associations.

Considered together, can these models (CHARM, SAM, MINERVA2, and the matrix model) handle the differential forgetting of item and associative information and the half seesaw effect? None of them can handle both effects without some modification, but then neither could TODAM. There would seem to be nothing that in principle would prevent a successful revision of any of these models, but this remains to be seen. When I started this revision, I had no idea whether or not it would be successful, even though in principle I thought it might work. But there is a very big difference between in principle and in fact.

Summary

The differential forgetting of item and associative information and the half seesaw effect necessitate a revision of TODAM. The revised version includes a general context vector and the autoconvolution of item vectors concatenated with context. For associative information, the two members of the pairs (the *A* and *B* items) are summed before the autoconvolution. Mediators are added to improve performance on tests of associative recognition.

The revision can handle these two effects at a quantitive level with the same parameter values if certain conditions are present:

1. Items are embedded in context.
2. There is context drift over time.
3. Context has more effect on item information than on associative information.
4. Item information is embedded in associative information so that there are two bases for item recognition.
5. Context is used for item recognition but not for pair recognition.
6. Mediators are used for associative information.
7. The study mediators are less available on a final recognition test.

References

Abramson, N. (1963). *Information theory and coding.* New York: McGraw-Hill.

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

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

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

Atkinson, R. C., Bower, G. H., & Crothers, E. J. (1965). *An introduction to mathematical learning theory.* New York: Wiley.

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology, 104,* 126–133.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium on Cognition* (pp. 313–342). Hillsdale, NJ: Erlbaum.

Borsellino, A., & Poggio, T. (1973). Convolution and correlation algebra. *Kybernetik, 122,* 113–122.

Bower, G. H. (1967). 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, pp. 229–325). New York: Academic Press.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20,* 231–243.

Clark, S. E., & Burchett, R. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition, 22,* 55–62.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

Clark, S. E., & Shiffrin, R. M. (1987). Recognition of multiple-item probes. *Memory & Cognition, 15,* 367–378.

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 813–824.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62,* 145–154.

Fernandez, A., & Glenberg, A. M. (1985). Changing environmental con-

[18] By comparison, with multiple convolutions, the number of elements required is approximately nN, so that is only several thousand.

text does not reliably affect memory. *Memory & Cognition, 13,* 333–345.

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

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Theory and data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567.

Greene, R. L. (1994, November). *Mirror effect in order and associative information.* Paper presented at the Psychonomics Meeting, St. Louis, MO.

Greene, R. L. (1996). Mirror effect in order and associative information: The role of response strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 687–695.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 846–858.

Grossberg, S. (1978). Do all neural models really look alike? A comment on Anderson, Silverstein, Ritz, & Jones. *Psychological Review, 85,* 592–596.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hintzman, D. L. (1992). Mathematical constraints and the Tulving-Wiseman law. *Psychological Review, 99,* 536–542.

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 275–289.

Hockley, W. E. (1991). Recognition memory for item and associative memory: A comparison of forgetting rates. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 227–248). Hillsdale, NJ: Erlbaum.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1321–1330.

Hockley, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition, 22,* 713–722.

Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition, 24,* 202–216.

Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review, 94,* 341–358.

Hockley, W. E., & Murdock, B. B. (1992). Speed-accuracy tradeoff and item recognition: A reply to Gronlund and Ratcliff. *Journal of Mathematical Psychology, 36,* 461–467.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79,* 2554–2558.

Humphreys, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition, 4,* 221–232.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96,* 208–233.

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology, 33,* 36–67.

Jordan, M. I. (1986). An introduction to linear algebra in parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Foundations* (Vol. 1, pp. 365–422). Cambridge, MA: MIT Press.

Kahana, M. J. (1997). *An analysis of the recognition-recall relations in four distributed memory models.* Manuscript submitted for publication.

Kohler, W. (1941). On the nature of association. *Proceedings of the American Philosophical Society, 84,* 489–502.

Lewandowsky, S. (1993). The rewards and hazards of computer simulations. *Psychological Science, 4,* 236–243.

Lewandowsky, S., & Li, S-C. (1994). Memory for serial order revisited. *Psychological Review, 101,* 534–538.

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

Light, L. L., & Carter-Sobel, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 1–11.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes and the recovery of order information. *Journal of Experimental Psychology: General, 122,* 291–315.

McGee, R. (1980). Imagery and recognition memory: The effects of relational organization. *Memory & Cognition, 8,* 394–399.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455.

Metcalfe, J. (1991). Recognition failure and the composite memory trace in CHARM. *Psychological Review, 98,* 529–553.

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

Metcalfe-Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review, 92,* 1–38.

Murdock, B. B. (1963). Interpolated recall in short-term memory. *Journal of Experimental Psychology, 66,* 525–532.

Murdock, B. B. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

Murdock, B. B. (1979). 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* (pp. 105–119). Hillsdale, NJ: Erlbaum.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B. (1989). Learning in a distributed memory model. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium on Cognition* (pp. 69–106). Hillsdale, NJ: Erlbaum.

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

Murdock, B. B. (1995). Similarity in a distributed memory model. *Journal of Mathematical Psychology, 39,* 251–264.

Murdock, B. B. (1996). Item, associative, and serial-order information in TODAM. In S. Gathercole (Ed.), *Models of short-term memory* (pp. 239–266). Hove, England: Psychology Press.

Murdock, B. B., & Hockley, W. E. (1989). Short-term memory for associations. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 71–108). New York: Academic Press.

Murdock, B. B., & Kahana, M. J. (1993). An analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 689–697.

Murdock, B. B., & Lamon, M. (1988). The replacement effect: Repeating some items while replacing others. *Memory & Cognition, 16,* 91–101.

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 882–894.

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition, 22,* 584–590.

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 158–172.

Ohrt, D. D., & Gronlund, S. D. (1995, August). *List-length effect: Control and theoretical implications.* Paper presented at the Mathematical Psychology Society meetings, Irvine, CA.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R. (1981). Parallel-processing mechanisms and processing of organized information in human memory. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 265–283). Hillsdale, NJ: Erlbaum.

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

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 763–785.

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC slopes. *Psychological Review, 99,* 518–535.

Reed, A. V. (1976). List length and the time-course of recognition in immediate memory. *Memory & Cognition, 4,* 16–30.

Rock, I., & Ceraso, J. (1964). Toward a cognitive theory of associative learning. In C. Scheerer (Ed.), *Cognition: Theory, research, promise* (pp. 110–146). New York: Harper & Row.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163.

Shiffrin, R. M., & Raaijmakers, J. (1992). The SAM retrieval model: A retrospective and prospective. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (Vol. 2, pp. 69–86). Hillsdale, NJ: Erlbaum.

Smith, S. M. (1988). Environmental context-dependent memory. In D. M. Thomson & G. M. Davies (Eds.), *Memory in context: Context in memory* (pp. 13–34). New York: Wiley.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86,* 255–262.

Thurstone, L. L. (1930). The relation between learning time and length of task. *Psychological Review, 37,* 44–58.

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior, 1,* 321–334.

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

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

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

Weber, E. U. (1988). Expectation and variance of item resemblance distributions in a convolution-correlation model of distributed memory. *Journal of Mathematical Psychology, 32,* 1–43.

Appendix

TODAM2 Derivations

For an item $\mathbf{f}$ embedded in context, where $\mathbf{a}$ and $\mathbf{b}$ are the context vectors flanking the item vector ($\mathbf{a}$ on the left and $\mathbf{b}$ on the right), define the embedded item vectors $\mathbf{F}$ and $\mathbf{G}$ as

$$\mathbf{F} = \mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b},\ \mathbf{G} = \mathbf{a} \oplus \mathbf{g} \oplus \mathbf{b},$$

where $\oplus$ denotes concatenation. With probabilistic encoding, in which each element of the item vector is encoded with probability p or not encoded (set to zero) with probability $1 - p$, this becomes

$$\mathbf{F} = \mathbf{a} \oplus p\mathbf{f} \oplus \mathbf{b},\ \mathbf{G} = \mathbf{a} \oplus p\mathbf{g} \oplus \mathbf{b}.$$

For an A–B pair, assume that the two item vectors are summed and the summation is embedded in context. Denote this as $\mathbf{FG}$, and, with probabilistic encoding,

$$\mathbf{FG} = \mathbf{a} \oplus \nu(p\mathbf{f} + p\mathbf{g}) \oplus \mathbf{b},$$

where $\nu = 1/\sqrt{2}$ is the normalization coefficient, so $E[\nu(\mathbf{f} + \mathbf{g}) \cdot \nu(\mathbf{f} + \mathbf{g})] = 1$.

For mediators, let σ be the probability that a subject finds a mediator and $1 - \sigma$ be the probability that he or she does not. I use the following notation: Let (A,p) denote outcome A with probability p and $(A,p) \cup (B,q)$ denote outcome A with probability p or outcome B with probability q. Then, if there are two study outcomes A and B with probabilities p and $1 - p$ and two test outcomes C and D with probabilities q and $1 - q$, the "expected utility" of the two study outcomes crossed with the two test outcomes is

$$[(A,p) \cup (B,1 - p)] \times [(C,q) \cup (D,1 - q)]$$
$$= ACpq + ADp(1 - q) + BC(1 - p)q + BD(1 - p)(1 - q).$$

An embedded pair with a mediator ($\mathbf{m}$) can be denoted as $\mathbf{FMG}$, where

$$\mathbf{FMG} = \mathbf{a} \oplus w(p\mathbf{f} + p\mathbf{m} + p\mathbf{m}) \oplus \mathbf{b};$$

$w = 1/\sqrt{3}$ is the normalization coefficient, so $E[w(\mathbf{f} + \mathbf{m} + \mathbf{g}) \cdot w(\mathbf{f} + \mathbf{m} + \mathbf{g})] = 1$.

For an item vector $\mathbf{f}$, denote autoconvolution as $\mathbf{f}^{*2} = \mathbf{f} * \mathbf{f}$; thus, with an attention weight γ, one can write the storage equation for the revised version of TODAM as

$$\mathbf{M}_j = \alpha\mathbf{M}_j + \gamma\mathbf{F}_j^{*2} + \gamma\mathbf{G}_j^{*2} + (1 - 2\gamma) \times \{[(\mathbf{FG}_j)^{*2}, 1 - \sigma] \cup [(\mathbf{FMG}_j)^{*2}, \sigma]\}.$$

Assume that there are attention weights γ at study and κ at test; encoding probabilities p_s at study, p_r at test, and p_m for the mediator; and mediator probabilities σ at study and τ at test. Ignoring item subscripts, one can then write any memory pair as

$$\gamma(\mathbf{a} \oplus p_s\mathbf{f} \oplus \mathbf{b})^{*2} + \gamma(\mathbf{a} \oplus p_s\mathbf{g} \oplus \mathbf{b})^{*2} + (1 - 2\gamma)\{[\mathbf{a} \oplus \nu(p_s\mathbf{f} + p_s\mathbf{g}) \oplus \mathbf{b}]^{*2}, 1 - \sigma\} \cup \{[\mathbf{a} \oplus w(p_s\mathbf{f} + p_m\mathbf{m} + p_s\mathbf{g}) \oplus \mathbf{b}]^{*2}, \sigma\}$$

and any probe pair as

$$\kappa(\mathbf{a} \oplus p_r\mathbf{f} \oplus \mathbf{b})^{*2} + \kappa(\mathbf{a} \oplus p_r\mathbf{g} \oplus \mathbf{b})^{*2} + (1 - 2\kappa)\{[\mathbf{a} \oplus \nu(p_r\mathbf{f} + p_r\mathbf{g}) \oplus \mathbf{b}]^{*2}, 1 - \tau\} \cup \{[\mathbf{a} \oplus w(p_r\mathbf{f} + p_m\mathbf{m} + p_r\mathbf{g}) \oplus \mathbf{b}]^{*2}, \tau\}.$$

For the old- and new-item (or intact- and rearranged-pair) means, E(Probe·Memory), the expected value of the probe dotted with the memory vector, is needed. For an item probe, there will always be three components: pure context (PC), pure item (PI), and context item (CI). For a pair probe, there will be these three components along with a fourth, pure associative (PA). To illustrate, consider an item probe $\mathbf{F} = (\mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b})^{*2}$ dotted with the corresponding item component $\mathbf{F}'$ in the memory vector, where $\mathbf{F}' = (\mathbf{a}' \oplus \mathbf{f}' \oplus \mathbf{b}')^{*2}$.

$$\begin{aligned} E(\mathbf{F}\cdot\mathbf{F}') &= E[(\mathbf{a} \oplus \mathbf{f} \oplus \mathbf{b})^{*2}\cdot(\mathbf{a}' \oplus \mathbf{f}' \oplus \mathbf{b}')^{*2}] \\ &= E\{[\mathbf{a}*\mathbf{a} + \mathbf{f}*\mathbf{f} + \mathbf{b}*\mathbf{b} + 2(\mathbf{a}*\mathbf{f}) + 2(\mathbf{a}*\mathbf{b}) \\ &\quad + 2(\mathbf{f}*\mathbf{b})]\cdot[\mathbf{a}'*\mathbf{a}' + \mathbf{f}'*\mathbf{f}' + \mathbf{b}'*\mathbf{b}' \\ &\quad + 2(\mathbf{a}'*\mathbf{f}') + 2(\mathbf{a}'*\mathbf{b}') + 2(\mathbf{f}'*\mathbf{b}')]\} \\ &= E[(\mathbf{a}*\mathbf{a})\cdot(\mathbf{a}'*\mathbf{a}')] + E[(\mathbf{f}*\mathbf{f})\cdot(\mathbf{f}'*\mathbf{f}')] \\ &\quad + E[(\mathbf{b}*\mathbf{b})\cdot(\mathbf{b}'*\mathbf{b}')] + 4E[(\mathbf{a}*\mathbf{f})\cdot(\mathbf{a}'*\mathbf{f}')] \\ &\quad + 4E[(\mathbf{a}*\mathbf{b})\cdot(\mathbf{a}'*\mathbf{b}')] + 4E[(\mathbf{f}*\mathbf{b})\cdot(\mathbf{f}'*\mathbf{b}')] \end{aligned}$$

because the cross products, such as $E[(\mathbf{a}*\mathbf{f})\cdot(\mathbf{a}'*\mathbf{b}')]$, are zero. The components for PC, PI, and CI, respectively, are

$$(\mathbf{a}*\mathbf{a})\cdot(\mathbf{a}'*\mathbf{a}') + (\mathbf{b}*\mathbf{b})\cdot(\mathbf{b}'*\mathbf{b}') + 4(\mathbf{a}*\mathbf{b})\cdot(\mathbf{a}'*\mathbf{b}'),\ (\mathbf{f}*\mathbf{f})\cdot(\mathbf{f}'*\mathbf{f}'),$$

and

$$4[(\mathbf{a}*\mathbf{f})\cdot(\mathbf{a}'*\mathbf{f}') + (\mathbf{f}*\mathbf{b})\cdot(\mathbf{f}'*\mathbf{b}')].$$

For a pair probe, there will be an $\mathbf{f}*\mathbf{g}$ term, so the PA component will be

$$4[(\mathbf{f}*\mathbf{g})\cdot(\mathbf{f}'*\mathbf{g}')]$$

or, with a mediator,

$$4[(\mathbf{f}*\mathbf{m})\cdot(\mathbf{f}'*\mathbf{m}') + 4(\mathbf{f}*\mathbf{g})\cdot(\mathbf{f}'*\mathbf{g}') + 4(\mathbf{m}*\mathbf{g})\cdot(\mathbf{m}'*\mathbf{g}')].$$

To obtain explicit expressions for these components, one must expand and collect terms in E(Probe·Memory). The following are some useful principles.

1. For any set of item vectors $\mathbf{f}$, $\mathbf{g}$, $\mathbf{u}$, and $\mathbf{v}$ (or their convolution) and constants a, b, c, and d,

$$E(a\mathbf{f}\cdot b\mathbf{g}) = abE(\mathbf{f}\cdot\mathbf{g}),$$

$$E[(a\mathbf{f}*b\mathbf{g})\cdot(c\mathbf{u}*d\mathbf{v})] = (ab)(cd)E[(\mathbf{f}*\mathbf{g})\cdot(\mathbf{u}*\mathbf{v})],$$

and the same holds true for autoconvolution; for example,

$$E[(a\mathbf{f}*a\mathbf{f})\cdot(b\mathbf{g}*b\mathbf{g})] = a^2b^2E[(\mathbf{f}*\mathbf{f})\cdot(\mathbf{g}*\mathbf{g})].$$

2. For any random vectors $\mathbf{u}$ and $\mathbf{v}$ that are normalized to P_u and P_V (i.e., if N_u and N_V are the number of elements in $\mathbf{u}$ and $\mathbf{v}$, the variance of the feature distribution is p_u/N_u and P_V/N_V), but only subsets of the features N_u' and N_V' enter into the convolution, if Z and W are random variables from the $\mathbf{u}$ and $\mathbf{v}$ feature distributions,

$$\begin{aligned} E[(\mathbf{u}*\mathbf{v})\cdot(\mathbf{u}*\mathbf{v})] &= N_u'N_V'E(Z^2W^2) \\ &= N_u'N_V'\sigma_u^2\sigma_v^2 = N_uN_v(p_uN_u)(P_VN_V); \end{aligned}$$

thus, if

$$P_u = P_V = 1,$$

then

$$E[(\mathbf{u}*\mathbf{v})\cdot(\mathbf{u}*\mathbf{v})] = (N_u'/N_u)(N_v'/N_v).$$

3. With normalized item vectors and a reasonable value of N, the expected value of the dot product of the autoconvolution of two item vectors $E[(\mathbf{f}*\mathbf{f})\cdot(\mathbf{f}*\mathbf{f})]$ is

$$E[(\mathbf{f}*\mathbf{f})\cdot(\mathbf{f}*\mathbf{f})] \cong 2,$$

but, for two independent item vectors $\mathbf{f}$ and $\mathbf{g}$,

$$E[(\mathbf{f}*\mathbf{g})\cdot(\mathbf{f}*\mathbf{g})] = 1$$

(Weber, 1988).

4. For k, a constant, using the expected utility notation,

$$\begin{aligned} k + [(A,p)\cup(B,q)] &= (k + A,p)\cup(k + B,q) \\ &= (k + A)p + (k + B)q. \end{aligned}$$

5. For τ, a conditional probability,

$$\begin{aligned} &[(A,\sigma)\cup(B,1-\sigma)] \times [(C,\tau)\cup(D,1-\tau)] \\ &\quad = AC\sigma\tau + AD\sigma(1-\tau) + BD(1-\sigma), \end{aligned}$$

because BC cannot happen.

For an item probe, the PC, PI, and CI components are necessary; call them $s1$, $s2$, and $s3$, respectively. For a new-item probe, $s2 = s3 = 0$ because there is no match, and $E(\mathbf{f}\cdot\mathbf{h}) = E[(\mathbf{f}*\mathbf{h})\cdot(\mathbf{g}*\mathbf{h})] = 0$. Consequently, at any output position t,

$$d_I'(t) \approx \mu_o(t) - \mu_I(t) = s1 + s2 + s3 - s1 = s2 + s3. \quad \text{(A1)}$$

The scale factor (proportionality factor) is the standard deviation, and $s1$, $s2$, and $s3$ are shown in Table A1.

For a pair probe, the PC, PI, CI, and PA components are necessary; call them $t1 = s1$, $t2$, $t3$, and $t4$. One needs to distinguish intact (I) and rearranged (R) probes because, unlike $s2$ and $s3$, $t2$, $t3 > 0$. For an intact probe,

$$\begin{aligned} d_A'(t) &\approx \mu_I(t) - \mu_R(t) \\ &= t1 + t2I + t3I + t4I - (t1 + t2R + t3R) \\ &= t2I + t3I + t4I - (t2R + t3R), \end{aligned} \quad \text{(A2)}$$

because $t4I$ is zero for a rearranged probe. The various t terms are shown in Table A1. If context is not used at the time of retrieval, then $t3I = t3R = 0$ and $d_A'(t)$ reduces to

$$d_A'(t) \approx t2I + t4I - t2R. \quad \text{(A3)}$$

Depending on the experimental design, it may be that $t2I \approx t2R$, in which case

$$d_A'(t) \approx t4I. \quad \text{(A4)}$$

The expressions for $\mu_I(t)$ and $\mu_A(t)$ contain fixed and variable terms. The fixed terms are independent of the study–test lag, whereas the variable terms depend on lag. For the variable terms, I assume that the context changes with each new study or test item or pair. Also, the memory vector is decremented (by α) with every new study or test item or pair. (These follow from the storage equation, Equation 4.) I also assume that the context changes before each test item or pair is presented but that the memory vector is decremented after the comparison process.

I denote the dependence of the variables by subscripts; thus, α_{kt} means α as a function of study position (k) and test position (t). Lag is the difference between k and t; if L is list length, then lag $= L - k + t - 1$. Also, $(\alpha\rho)_{kt}$ means the product of α and ρ as a function of lag, and

Table A1
Variable Terms, New- or Old-Item Probe Components (s1, s2, and s3), and Intact (I) and Rearranged (R) Probe Components (t2, t3, and t4)

Variable terms needed to compute the pure-context, pure-item, context-item, and pure-associative components

$$\sum_{kt}(\alpha\rho^2)_{kt} = \rho^2\,\frac{1-(\alpha\rho^2)^{L+t-1}}{1-\alpha\rho^2}$$

$$\alpha_{Lt} = \frac{1}{L}\,\frac{1-\alpha^L}{1-\alpha}\,\alpha^{t-1}$$

$$(\alpha\rho)_{Lt} = \frac{\rho}{L}\,\frac{1-(\alpha\rho)^L}{1-\alpha\rho}\,(\alpha\rho)^{t-1}$$

Pure-context ($s1$), pure-item ($s2$), and context-item ($s3$) components for a new or old item probe

$$s1 = \left[2\left(\frac{N_a}{N_c}\right)^2 + 2\left(\frac{N_b}{N_c}\right)^2 + 4\left(\frac{N_a}{N_c} + \frac{N_b}{N_c}\right)\right]\sum_{kt}(\alpha\rho^2)_{kt}$$

$$s2 = \{\gamma + (1-2\gamma)[v^2(1-\sigma) + w^2\sigma]\}p_s^2 p_r^2 \alpha_{Lt}$$

$$s3 = 4\{\gamma + (1-2\gamma)[v(1-\sigma) + w\sigma]\}p_s p_r\left(\frac{N_a}{N_c} + \frac{N_b}{N_c}\right)(\alpha\rho)_{Lt}$$

Pure-item ($t2I$ and $t2R$), context-item ($t3I$ and $t3R$), and pure-associative ($t4I$) components for intact (I) and rearranged (R) pair probes

$$t2I = 2\{2[A'C'\sigma\tau + A'D'\sigma(1-\tau) + B'D'(1-\sigma)]p_s^2 p_r^2 + (1-2\gamma)(1-2\kappa)\sigma\tau w^4 p_m^4\}\alpha_{kt}$$

$$t3I = 4\{2[AC\sigma\tau + AD\sigma(1-\tau) + BD(1-\sigma)]p_s p_r + (1-2\gamma)(1-2\kappa)\sigma\tau w^2 p_m^2\}\left(\frac{N_a}{N_c} + \frac{N_b}{N_c}\right)(\alpha\rho)_{kt}$$

$$t4I = 4(1-2\gamma)(1-2\kappa)\{[\sigma w^2 + (1-\sigma)v^2][\tau w^2 + (1-\tau)v^2]p_s^2 p_r^2\} + (2\sigma\tau w^4 p_s p_m^2 p_r)\alpha_{kt}$$

$$t2R = 2\{2[\gamma + (1-2\gamma)(\sigma w^2 + (1-\sigma)v^2)][\kappa + (1-2\kappa)v^2]p_s^2 p_r^2\}\alpha_{kt}$$

$$t3R = 4\{2[\gamma + (1-2\gamma)(\sigma w + (1-2\sigma)v)][\kappa + (1-2\kappa)_v]p_s p_r\}\left(\frac{N_a}{N_c} + \frac{N_b}{N_c}\right)(\alpha\rho)_{kt}$$

$A = \gamma + (1-2\gamma)w$; $B = \gamma + (1-2\gamma)v$; $C = \kappa + (1-2\kappa)w$; $D = \kappa + (1-2\kappa)v$;

$A' = \gamma + (1-2\gamma)w^2$; $B' = \gamma + (1-2\gamma)v^2$; $C' = \kappa + (1-2\kappa)w^2$; $D' = \kappa + (1-2\kappa)v^2$.

Note. L = list length.

$(\alpha\rho^2)_{kt}$ means the product of α and ρ^2 as a function of lag. For the PC component, one needs the summation of $\alpha\rho^2$ over all study and test positions to date; for the PI and PA components, one needs α_{kt}; and, for the CI component, one needs $(\alpha\rho)_{kt}$. The latter two are averaged over study position, so I denote them as α_{Lt} and $(\alpha\rho)_{Lt}$. Explicit expressions for these three variable terms, $s1$–$s3$, and the various t terms are shown in Table A1.

As a numerical example, if $L = 6$, $\alpha = .9$, $\rho = .8$, $\gamma = .2$, $\kappa = 0$, $p_s = .4$, $p_r = .707$ ($a = .1$, time $= .1$), $p_m = .7$, $\sigma = .7$, $\tau = .3$, $N_a = N_b = 1$, and $N_c = 4$, at $t = 1$ (first output position) $s1 = .727$, $s2 = .0537$, $s3 = .1321$, $t2I = .0555$, $t3I = .1968$, $t4I = .0380$, $t2R = .0537$, and $t3R = .1868$. These particular parameter values were chosen only to illustrate the computation. With these values, $d_I'(t) = .1858$ and $d_A'(t) = .0397$.

Received April 14, 1995
Revision received May 8, 1996
Accepted October 15, 1996 ■