Psychological Review
1991, Vol. 98, No. 4, 529-553

0033-295X/91/$3.00

Recognition Failure and the Composite Memory Trace in CHARM

Janet Metcalfe
Dartmouth College

The relation between recognition and recall, and especially the orderly *recognition-failure function* relating recognition and the recognizability of recallable words, was investigated using a composite holographic associative recall–recognition memory model (CHARM). Ten series of computer simulations are presented. Analysis of CHARM and comparisons to other models indicate that the recognition-failure function depends on (a) both recognition and recall being similar (convolution–correlation) processes such that an interpretable representation is retrieved in both tasks and (b) the information underlying both recall and recognition being stored in the same composite memory trace. It is of considerable interest that constructs central to the distributed nature of CHARM are responsible for the model's adherence to the recognition-failure function.

There are few paradigms within the study of human memory in which the data are so orderly that they could be considered to be determined by a psychological law. The recognition-failure paradigm, however, provides one such instance (Gardiner, 1988, 1989; Gardiner, Kaminska, Java, Clarke, & Mayer, 1990; Jones 1983; Jones & Gardiner, 1990; Nilsson, Dinniwell, & Tulving, 1987; Nilsson, Law, & Tulving, 1988). This paradigm is one in which subjects are presented with a list of item pairs to study in such a way that when later given one item as the cue, they will be able to recall the other. In some cases, the members of the pair are simple words, such as "glass-VASE"; in other cases, the cue may be a phrase or sentence:—"A firm and friendly touch—HANDSHAKE" or "He made scientists work harder—ALFRED NOBEL." Before being asked to recall the target items (here presented in capitals), subjects are first requested to recognize them. The question of interest is whether people who are able to recall a particular item as the target to a specific cue are also necessarily able to recognize that target as having been presented in the list. Our intuitions suggest this to be the case and that there should be a strong or perhaps even an absolute dependence relation between recall of an item and recognition of that item—recallable items should also be recognizable. Recognition failure of recallable items should not occur, or at least should be a rare event.

This prediction was made more explicit in models that postulated that recall was a two-stage process in which subjects first generated response possibilities and then recognized the correct alternative from among those generated (Bahrick, 1979; Kintsch, 1978; Martin, 1975). The strong dependence expected intuitively and as a result of such models of recall has failed to appear in experiments that investigated this issue. Gardiner and Nilsson (1990) reviewed the results of 42 published articles yielding a total of 272 different observations. The empirically found relation between recognition and recall is one of near, but not total, independence. There is a systematic dependence relation between these two tasks, but nothing like the dependence that would be expected if episodic recognition were a necessary subprocess of recall. The recognition failure of recallable items is a consistent rather than a rare event that occurs under a wide variety of experimental situations. As Tulving (1983) has noted,

> We know that recognition failure occurs as readily in experiments in which recognition is higher than recall as it does in experiments in which recall is higher than recognition (e.g., Flexser and Tulving, 1978; Wiseman and Tulving, 1976). It occurs in situations in which no practice lists are given to subjects before they learn the critical list (e.g., Begg, 1979; Bowyer and Humphreys, 1979). It occurs as readily in an immediate test as in one given a week after the learning (Tulving and O. C. Watkins, 1977). It occurs with different kinds of to-be-remembered word pairs presented at study: "weak" cues and targets, as in our original experiments, "strong" cues and targets (e.g., Vining and Nelson, 1979), and unrelated words (e.g., Begg, 1979; Rabinowitz et al., 1977). It does not matter whether the recognition-test distractors are semantically related or unrelated to targets (e.g., Begg, 1979; Bowyer and Humphreys, 1979; Postman, 1975; Rabinowitz et al., 1977). Indeed, it is not necessary to have any distractor items at all in the recognition test for recallable words to remain unrecognized (e.g., Begg, 1979; Wallace, 1978). Low-frequency words with few semantic senses (Reder et al., 1974) and words that have only a single meaning in the dictionary (Tulving and O. C. Watkins, 1977) fail to be recognized even though they can be recalled indistinguishably from high-frequency words with many meanings. Whether subjects come to the task without any preconceptions or fully aware of what is happening, and whether they have had a great deal of practice on the recognition-failure paradigm is immaterial (e.g., Rabinowitz et al., 1977; Wiseman and Tulving, 1975). (pp. 286–287)

The invariance can be described by the equation proposed by Wiseman and Tulving (1975) to illustrate the relation between recognition and the recognizability of recallable words:

This research was facilitated by National Institute of Mental Health Grant R29 MH48066-01. I thank J. J. Bharucha, Arthur Flexser, Carol Fowler, John Gardiner, Douglas Hintzman, David Horton, George Mandler, Bennet B. Murdock, Paul Muter, Lars-Goran Nilsson, Endel Tulving, George Wolford, and two anonymous reviewers for their comments. (The fact that these people were gracious enough to scrutinize and criticize this work does not imply that they agree with the theoretical position proposed herein—though perhaps some do.)

Correspondence concerning this article should be addressed to Janet Metcalfe, Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755.

$$p(\mathrm{Rn}|\mathrm{Rc}) = p(\mathrm{Rn}) + .5[p(\mathrm{Rn}) - p(\mathrm{Rn})^2]. \quad (1)$$

Although this equation provides the general form of the relation between recognition and recognition of recallable words, the value of .5 may be a slight overestimation of the dependence. Flexser (1981) noted that item and subject dependencies need to be factored out of the data and devised a method of homogenization that helps to do so. The difference between raw and homogenized data is illustrated later. The basic point, though, is that the model should produce a correlation on or slightly below the Wiseman and Tulving (1975) function illustrated by Equation 1.

As is shown in Figure 1, which illustrates the probabilities of recognition and recall from a variety of experiments, there appears to be no straightforward relation between recognition and recall that allows one to predict the conditional relation between recall and recognition failure that is given by the Tulving-Wiseman function. The data from the experiments in Figure 1 may be reanalyzed according to the conditional probability of recognition given recall, and this probability may be compared with the simple recognition probability. The results of such an analysis are shown in Figure 2.

The highly stable nature of this relation indicates that it is telling us something important about the nature of human memory. The fact that intuitive and plausible ideas about the workings of memory, such as those of generate-and-edit theory (Bahrick, 1979), are contradicted by it makes explanation of the recognition-failure function an especially challenging problem. It is also a problem of sufficient complexity that use of explicit models may be illuminating. To discuss with confidence the implications and predictions of various memory models, it is, of course, necessary to look at *implemented* models. It would be rash to claim that a model that has not yet been formulated for this paradigm can or cannot, in principle, account for the results. In unimplemented models, the boundary conditions, ancillary assumptions that allow the model produce the data, and, indeed, the basic conceptualization of the paradigm by the author of that model, are not specified. So one simply cannot know whether unformulated models will work or not, and speculation is not justified. Of the models that have been sketched out, some can accommodate some recognition failure of recallable words but do not necessarily predict that the data fall on the Tulving-Wiseman function—a much more exacting requirement than the mere prediction of any deviation from complete dependence. Kintsch's (1978) and Gillund and Shiffrin's (1984) models are probably in this class, though neither have been applied to these data in detail. Some models, although they offer accounts of the function itself, fail to offer any plausible story for the exceptions (e.g., Begg, 1979; Flexser & Tulving, 1978; Jones, 1978). Some make predictions, such as that recognition failure should not occur with words that have only a single meaning (Gillund & Shiffrin, 1984; Reder, Anderson, & Bjork, 1974), that are not borne out by the data (Tulving & Watkins, 1977). Some can produce the function but can equally easily predict results that are widely discrepant from the function, such as those that never occur in the data. For example, Hintzman (1987) has applied the multiple-trace model, MINERVA, with results shown in Figure 3. Under some parameter combinations, the model can produce results that resemble the Tulving-Wiseman function (shown in Figure 3, Panel D). However, it can also readily produce independence (shown in Figure 3, Panels A and B), negative dependence (shown in Figure 3, Pan-

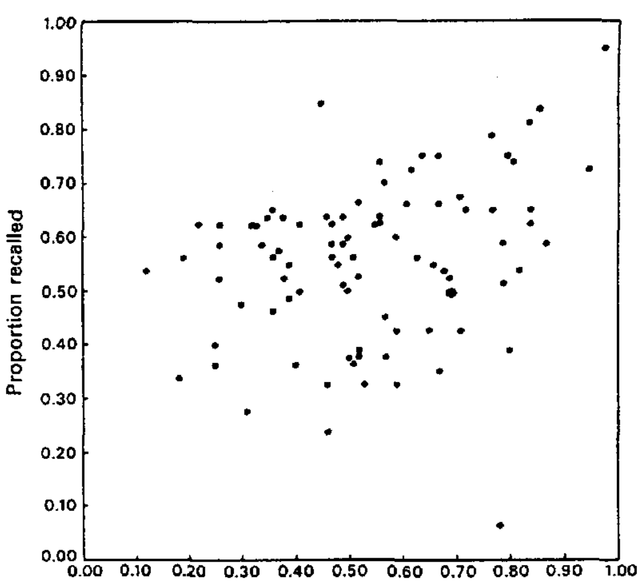


Figure 1. The probability of recognition as compared with recall found in the experimental data. (Each data point is based on a single experiment or experimental condition. From *Elements of episodic memory* (pp. 282–283) by E. Tulving, 1983, New York: Oxford University Press. Copyright 1983 by Oxford University Press. Adapted by permission.)

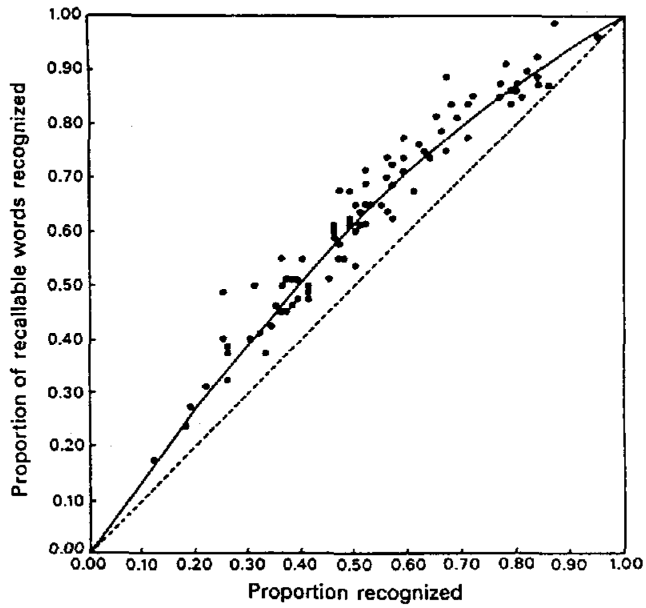


Figure 2. Proportion of recallable words recognized in the experimental data shown in Figure 1, from human subjects. (Each data point is based on a single experiment or experimental condition. From *Elements of episodic memory* (pp. 282–283) by E. Tulving, 1983, New York: Oxford University Press. Copyright 1983 by Oxford University Press. Adapted by permission.)

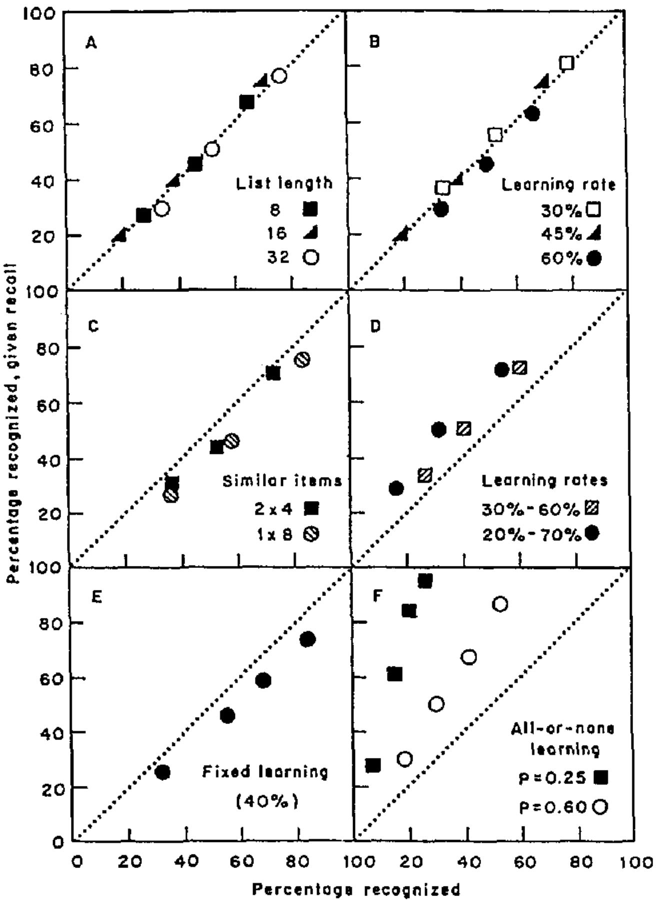


Figure 3. Proportion of recallable events recognized as predicted from several simulations of Hintzman's (1987) MINERVA 2 model. (From "Recognition and recall in MINERVA 2: Analysis of the recognition-failure paradigm' " by D. L. Hintzman, 1987, In P. Morris, *Modelling cognition,* New York: John Wiley & Sons, Inc. Copyright 1987 by John Wiley & Sons, Inc. Adapted by permission.)

els C and E), and vastly overdependent linear functions (shown in Figure 4, Panel F)—patterns that are not produced by human subjects. Such liberty of prediction is inconsistent with the highly constrained and lawful behavior of the human data and suggests that the model is missing the core of the problem. Because of its theoretical significance, I implemented a plain-vanilla version of Murdock's (1982) TODAM model. It does not produce the function. Because the TODAM model and the composite holographic associative recall–recognition memory (CHARM) model are closely related in a number of respects, the reasons for the difference in predictions between the two in this paradigm are of interest and are discussed later. The fact that results other than the empirically observed recognition-failure function and the exceptions to it are produced under plausible assumptions about the nature of human memory indicates that the relation is not artifactual in the sense that it is a necessary outcome of any conceptualization of memory or that it is a statistical quirk. If it were, then presumably all models would produce it. However, as Tulving (1983) put it, "the search for a theory that explains not only recognition failure but also exceptions to it remains wide open" (p. 290).

There is a method of formalizing the recall and recognition processes in a distributed model, CHARM, that produces results falling very close to the recognition-failure function. This model may thus be used as a tool to better understand the reasons for the empirical law and also to provide hypotheses about the structure and operations used in human memory that this relation between recognition and recall implies. By analyzing the model, which is obviously simpler and more tractable than human memory itself, I attempt to extract the critical constructs that are responsible for the relation and to chart some of the implications in terms of exceptions to the law. I also examine which constructs are necessary for the phenomenon by investigating other methods of instantiating recognition (as in the TODAM model) or by varying storage assumptions in the CHARM model, to see whether the effect still obtains. As is later illustrated, the recognition-failure function depends on two central assumptions in the model: (a) The information underlying recall and recognition is stored in the same composite trace and (b) the mechanisms underlying recall and recognition are the same and involve representational retrieval. If either of these assumptions in the model is violated, so long as the items themselves are unrelated, or statistically independent, complete independence between recognition and the recognizability of recallable items results. With these assumptions in place, however, the dependence relation, such as is found empirically, is produced by the model.

Outline of the Model

The CHARM model incorporates as a central construct the idea that the results of many associations or events are stored by being superimposed in a composite memory trace. Because of this superposition, the elements necessarily interact. Superposition is responsible for several of the most interesting predictions of the model (see Metcalfe & Murdock, 1981; Metcalfe Eich, 1982, 1985; Metcalfe, 1990; and Metcalfe & Bjork, 1991, for some examples that apply to the eyewitness testimony paradigm). A number of other related parallel distributed processing models also use the idea of superposition of events on a single surface (e.g., Anderson, Silverstein, Ritz, & Jones, 1977; Lewandowski & Murdock, 1989; McClelland & Rumelhart, 1986; Murdock, 1982, 1989; Rumelhart & McClelland, 1986). Distributed models are motivated by the idea that such superimposed or composite storage is necessary because of the neural structure that underlies cognition (Anderson, 1970; Kohonen, Oja, & Lehtio, 1981). Only a few of the psychological implications of such a concept of memory storage have been charted, however. As is later shown, this composite method of storage in which associations underlying both recall and recognition necessarily interact is critical for the CHARM model's ability to generate the recognition-failure function. Thus, the recognition-failure predictions of this model stem directly from the notion of composite storage in distributed neural memory models.

The model is called *holographic* because it depends on an associative encoding operation of convolution and a retrieval operation of correlation. Gabor (who invented the hologram) analyzed a number of models and determined that these formal operations define a model as being in the class of holographic models. Certain aspects of the popular version of the holographic metaphor, such as the idea that a part of the film (trace) is sufficient to reinstate a whole image or that multiple images can be superimposed on the same film (trace), apply quite nicely. Certain others (such as involvement with laser beams—there are other kinds of holograms in any case) do not apply. Van Gelder (1989) has pointed out that holographic associations provide the most extreme and radical example of a truly interactive distributed encoding method. As with the idea of a composite trace, the idea that items are interactively encoded when they are associated and thus may alter one another is an important notion about the nature of human memory.

The idea that both recall and recognition depend on associations formed by the operation of convolution and hence that both processes involve explicit retrieval of events from memory are investigated later in this article. Recall depends primarily on interitem associations—the convolution of the cue with the target. Recognition depends primarily on autoassociations—the convolution of the cue with itself and the target with itself. When a probe is presented for recognition, it is correlated with the trace, just as in recall the cue is correlated with the trace. An item that is interpretable as such is retrieved in both tasks. In recognition the retrieved item provides the basis for an occurrence decision.

The idea that recognition, like recall, involves the retrieval of an item rather than only the assessment of a strength or a familiarity value has been proposed by Mandler (1980), Tulving and Thomson (1973), Tulving (1983), and many others. Similarly, I proposed that recognition was based on autoconvolutions (Metcalfe Eich, 1985), which allow retrieval of an item from the composite memory trace. I chose this kind of recognition process because it permitted the holographic model to exhibit sensitivity in recognition to the congruence or similarity between the cue and the target, such as is shown in the empirical data. If recognition is not formalized in this way, this particular kind of similarity effect will not be produced. When it is formalized as a retrieval process, however, factors important in recall, such as similarity and organizational effects, will also impact on the recognition-memory results. I did not initially propose the convolution-based recognition model to account for data within the recognition-failure paradigm. Thus, this application provides a test case for that formulation.

There are other ways to think about the recognition process that do not necessarily involve the explicit retrieval of an interpretable item from memory. For example, one might enact a global match to the trace or traces and assess only the strength of that match, as is often assumed in signal-detection analyses of recognition memory. Many researchers have suggested that there is a familiarity or a fluency component in recognition that does not involve specific retrieval. For example, Atkinson and Juola (1973) have proposed that recognition may be based on both a fast familiarity judgment and a slower retrieval process. Similarly, Mandler (1980) has argued for both familiarity-based and retrieval-based recognition. Although I certainly do not wish to deny the possibility of some kind of familiarity-based recognition, in this article I explore some implications of a mechanism that retrieves an item, much like the retrieval mechanism in recall retrieves an item, which provides the basis for recognition-memory judgments.

Representation

Items in the model are conveniently represented as vectors with values randomly distributed around zero with some variance. They may vary in their similarity to one another, as prescribed by the experimental situation. The model allows for more specific delimitation of the exact makeup of particular items if the experimental situation or the nature of the items themselves warrants it. As is shown in some of the simulations presented here, items may vary in the number of features they contain, and also in their similarity, by feature overlap, to other items. Hence, the random vector notion is a default assumption for random stimuli but is made more specific according to the experimental situation being modeled. The idea that memory events may be represented as vectors has been a fruitful one in the study of human memory, allowing researchers better understanding of similarity effects, interference effects, and a number of other phenomena.

Association Formation

Two items, $\mathbf{F} = [f_{-(n-1)/2}, \ldots, f_{-1}, f_0, f_1, \ldots, f_{(n-1)/2}]$ and $\mathbf{G} = [g_{-(n-1)/2}, \ldots, g_{-1}, g_0, g_1, \ldots, g_{(n-1)/2}]$, are associated in the CHARM model by the operation of convolution, denoted with an asterisk ($*$), and defined as

$$(\mathbf{F} * \mathbf{G})_m = \mathbf{T}_m = \sum_{(i,j) \in S(m)} f_i g_j, \tag{2}$$

where $S(m) = [(i, j) | -(n-1)/2 \leq (n-1)/2$, and $i + j = m]$. The subscript m denotes the mth element in the vector formed by convolution. Interitem convolution is said to underlie recall, whereas autoconvolution primarily underlies recognition. Because both of these forms of processing are added into the same trace, they will interact with one another. One may tease them apart, however, and store interitem associations in one trace and autoassociations in another. These separate trace models will still be capable of recall and recognition, respectively. In addition, the same items in this dual system version of the model can be used. The results of doing so for the recognition-failure function are investigated shortly and contrasted to results obtained with the single composite trace for both recall and recognition.

Storage

The results of successive convolutions (be they autoassociations or interitem associations) are added into a single composite memory trace, which is one of the core ideas of the model. The trace $\mathbf{T}$ is made up of associations as follows:

$$\mathbf{T} = \mathring{\alpha}\mathbf{A}*\mathbf{A} + \beta\mathbf{A}*\mathbf{B} + \mathring{\alpha}\mathbf{B}*\mathbf{B} + \mathring{\alpha}\mathbf{C}*\mathbf{C} + \beta\mathbf{C}*\mathbf{D} + \mathring{\alpha}\mathbf{D}*\mathbf{D} + \ldots + \text{preexisting noise}. \tag{3}$$

The weightings for the autoassociations ($\mathring{\alpha}$) and for the interitem associations (β) may vary. The implications of this variation, for recognition failure, are charted shortly. The trace may also be assumed to start out with some noise, perhaps from previous memories, rather than as a blank slate or a zero vector, though in some applications of the model this does not change the basic results. This addition of noise is investigated in simulations.

Retrieval

The retrieval operation is correlation. Retrieval generates a new vector $\mathbf{R}_m$ from the elements of the cue and trace vectors by cross-correlating them. Accordingly,

$$(\mathbf{Q} \# \mathbf{T})_m = \mathbf{R}_m = \sum_{(i,j)\in S(m)} q_i t_j, \tag{4}$$

where $\mathbf{Q}$ is the cue vector with elements q_i, $\mathbf{T}$ is the trace with element t_j, and $S(m)$ is the domain of paired elements over which the correlation is attempted, that is, $S(m) = [(i, j)| -(n-1)/2 \leq i, j < (n-1)/2$, and $i - j = m]$. The result of retrieval is a new vector reflecting what the subject generates from episodic memory. The nature of this retrieved item underlies all studies in this model.

Relation Between Convolution and Correlation

Any given association $\mathbf{A}*\mathbf{B}$, consisting of unrelated items, potentially contains the information from both of the associated items. When $\mathbf{B}$ is used as the retrieval cue $[\mathbf{B} \# (\mathbf{A}*\mathbf{B})]$, the result is $\mathbf{A}$ + error. When $\mathbf{A}$ is used as the cue, the result is $\mathbf{B}$ + error. Under the autoassociation condition, that is where $\mathbf{A}*\mathbf{A}$, when $\mathbf{A}$ is itself used as a cue, two signal terms (both of which are $\mathbf{A}$, in this case) are retrieved. So $\mathbf{A} \# (\mathbf{A}*\mathbf{A}) = 2\mathbf{A} +$ error. The strength with which a particular item is retrieved as a signal component under the operation of correlation depends on the similarity between the retrieval cue, $\mathbf{Q}$, and the item that was associated with the item under consideration. In general, one may say that:

$$\begin{aligned}\mathbf{Q} \# [(\mathbf{A}*\mathbf{B}) + (\mathbf{C}*\mathbf{D}) + (\mathbf{E}*\mathbf{F}) + \ldots] \\ = S_{QA}\mathbf{B} + S_{QB}\mathbf{A} + \text{error}_{AB} + S_{QC}\mathbf{D} + S_{QD}\mathbf{C} \\ + \text{error}_{CD} + S_{QE}\mathbf{F} + S_{QF}\mathbf{E} + \text{error}_{EF}.\end{aligned} \tag{5}$$

Consider an example, to which I return later. Suppose that $\mathbf{A}$ is associated with $\mathbf{B}$, that $\mathbf{A}$ is also associated with itself, and that $\mathbf{B}$ is associated with itself. This is the scheme that is proposed to underlie both recall and recognition. The trace (ignoring weightings of the interitem and autoassociations, for the moment) is

$$\mathbf{T} = \mathbf{A}*\mathbf{A} + \mathbf{A}*\mathbf{B} + \mathbf{B}*\mathbf{B}.$$

If $\mathbf{A}$ is used as a retrieval cue, and hence is correlated with the trace, the retrieved item is

$$\begin{aligned}\mathbf{R} = \mathbf{A} \# (\mathbf{T}) = \mathbf{A} \# [(\mathbf{A}*\mathbf{A}) + (\mathbf{A}*\mathbf{B}) + (\mathbf{B}*\mathbf{B})] \\ = S_{AA}\mathbf{A} + S_{AA}\mathbf{A} + \text{error}_{AA} + S_{AA}\mathbf{B} + S_{AB}\mathbf{A} \\ + \text{error}_{AB} + S_{AB}\mathbf{B} + S_{AB}\mathbf{B} + \text{error}_{BB}.\end{aligned}$$

Assuming that the similarity between an item and itself is 1 and that $\mathbf{A}$ and $\mathbf{B}$ are unrelated (i.e., their similarity is zero), one gets:

$$\mathbf{R} = \mathbf{A} + \mathbf{A} + \mathbf{B} + \text{error}_{AA} + \text{error}_{AB} + \text{error}_{BB}. \tag{6}$$

Thus, $\mathbf{A}$ retrieves itself, but it also simultaneously retrieves the item with which it was associated, namely, $\mathbf{B}$. To compensate for the fact that the autoassociation produces double the signal of the interitem association, I will usually weight the autoassociations by .5 in the simulations that follow. In two of the simulations that follow, however, the values of the weighting parameters on the interitem associations and autoassociations are systematically varied.

Recall

Recall is based on retrieval, that is, on the vector resulting from correlation of the cue with the trace. However, because the output vector is typically noisy or distorted, and because in simulations of the model it is necessary to say what is recalled and how frequently it is recalled, a decision process is also necessary. The decision process is formulated as follows: The retrieved vector is matched to every item in a lexicon of possible outcomes (which excludes the cue—see Metcalfe Eich, 1982, for a discussion and an experiment focused on this exclusion), and in the simplest case, the item yielding the highest dot product will be the item recalled. However, this dot product must exceed a lower threshold. If the retrieved signal is just too noisy to be interpreted, recall will not occur. This threshold on recall controls the intrusion rate. Nothing in the model prohibits several items from being recalled from a single output vector, as for example in the **A-B/A-D** paradigm. However, in the simulations discussed later, only the best match is chosen as the recalled item.

Recognition

Recognition is also based on retrieval. In this case, however, the resulting vector is matched only against the probe itself. If the item currently being used as the probe was autoassociated and entered into the composite trace, then the vector retrieved by the probe will show a positive dot product (or resonance value) with the probe. If it was not encoded (and is unrelated to everything that was encoded), then the match between the item retrieved by the probe and the probe vector has an expected value of zero. A *yes* decision is given if this match between the retrieved item and the probe item exceeds a particular criterion.

Having outlined the basics of the model, I now turn to several applications of this scheme to experimental situations. Following several such applications and discussions of the experimental findings, I return to a discussion of the reasons for the results produced by the model. The overall strategy was to investigate the relation between recognition and the recognition of recallable items under a wide variety of situations and levels of both recall and recognition in an effort to see whether lawlike behavior results with the model as it does with the human data.

Applications to Recognition Failure

Criterion for Recognition

Kintsch (1978) pointed out that the setting of the criterion for recognition may be of theoretical importance in understanding the reasons for recognition failure. He noted that under a special set of circumstances—namely, when the criterion in the recognition task is very high and the criterion in the recall task is low—generate-and-edit theory may be able to account for the finding that subjects are sometimes unable to recognize items that they can recall. It is not clear, however, that generate-and-edit theory would produce the relation between recognition and recognition failure over the entire range found in the data. Because of the possible importance of the criterion in interpreting the recognition-failure findings, it is important to review the empirical results that have been presented on this issue and also to investigate the predictions of the model under variations in criterion.

Nilsson et al. (1988) conducted a cued-recall/recognition experiment on unique names of people or on geographical names. Examples were "He was the first of a long line but the only one on horseback—GEORGE WASHINGTON" or "A well-known building for music in VIENNA." The to-be-recalled targets were the words in capital letters. At the time of recall, 1 week later, subjects were presented with the descriptive sentence and a blank for the target; at the time of the recognition test, they were given the target words themselves, embedded in a list of similar items. Criterion shifts in recognition were investigated by using confidence ratings given by subjects. Subjects rated their responses on a scale from −3 to +3. On the basis of the confidence ratings the data were divided into criterial categories that were lenient, intermediate, or strict. In the lenient cases, items were treated as being recognized if the rating was 1 or greater; in the intermediate category, if the rating was 2 or greater; and in the strict category, only if the rating was 3. These kinds of confidence ratings have been analyzed by Murdock (1974) as mapping into shifts in the subject's criterion, and the results of such an analysis corresponded very nicely to what one would expect under a signal-detection analysis of recognition in which the criterion changes. Hence, this appears to be both a traditional and well-accepted method of investigating criterion shifts.

The data were also segmented according to subjects' ratings of the meaningfulness of the sentences. Meaningfulness was assessed by subjective judgments of how well the descriptive sentence fit with the target or of whether the sentence made sense in terms of the target. Some notion of cue–target congruity seems to apply, though perhaps the meaningfulness of the items themselves was also being measured.

The probabilities of recognition and recall for the person names are presented in Figure 4. The geographical names present a similar picture. As can be seen from these data, the probabilities of recognition varied considerably with recognition criterion; as would be expected, fewer high criterion than low criterion hits were found. The corresponding recall, of course, does not vary as a function of recognition criterion. Both recognition and recall were affected by the congruity (meaningfulness) between the cue and the target, but the levels of recall were affected somewhat more. Highly meaningful events were remembered better than were less meaningful events.

Figure 5 shows the proportion of recallable words recognized as a function of recognition (left panel). The right panels shows a reanalysis of these data, using a method of homogenization devised by Flexser (1981) that factors out spurious correlations resulting from item-selection effects and subject differences.

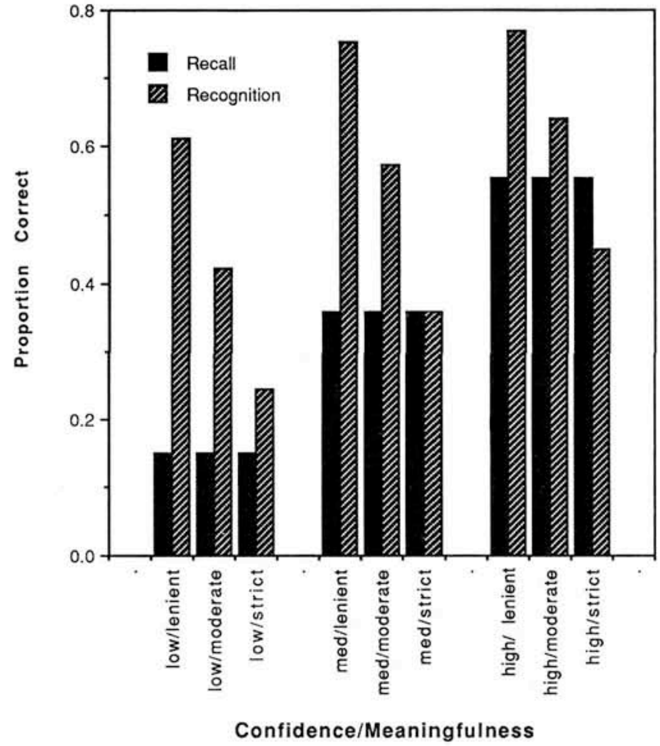


Figure 4. The probabilities of recognition and recall as a function of criterion changes in recognition. (Data are for the person names in Nilsson, Law and Tulving, 1988. From "Recognition failure of recallable unique names: Evidence for an empirical law of memory and learning" by L. G. Nilsson, J. Law, and E. Tulving, 1988, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* pp. 270, 273. Copyright 1988 by the American Psychological Association, Inc. Adapted by permission.)

Despite a wide range in the level of recognition obtained by differences in criterion, the data fell very close to the Tulving-Wiseman function shown in Figure 2.

Additional data on recognition criterion differences were presented by Begg (1979), who varied the number of distractors in the recognition test. He argued that with more distractor items the criterion for recognition would be higher. Begg's data for both conditions, measuring criterion in a manner different from that of Nilsson et al. (1988), also fell very close to the Tulving-Wiseman function.

Simulation Series 1

Method. To simulate the predictions of the model under conditions of a criterion shift in recognition, a $2 \times 5 \times 2 \times 3$ design was set up with the model, where the factors were (a) Type of Test, (b) Recognition Criterion, (c) Number of Features in the Item Vectors, and (d) Convolutions Within Each List. Each of these conditions was run through 1,000 replications (or 1,000 lists), so that the points (collapsed across pairs) presented in Figures 6 and 7 are each based on 3,000 simulated observations.

Each run of each simulation was set up as follows: A lexicon of 70 items was constructed by randomly selecting a value for each of the 31 or 63 features (depending on the level of this factor) of each item from a

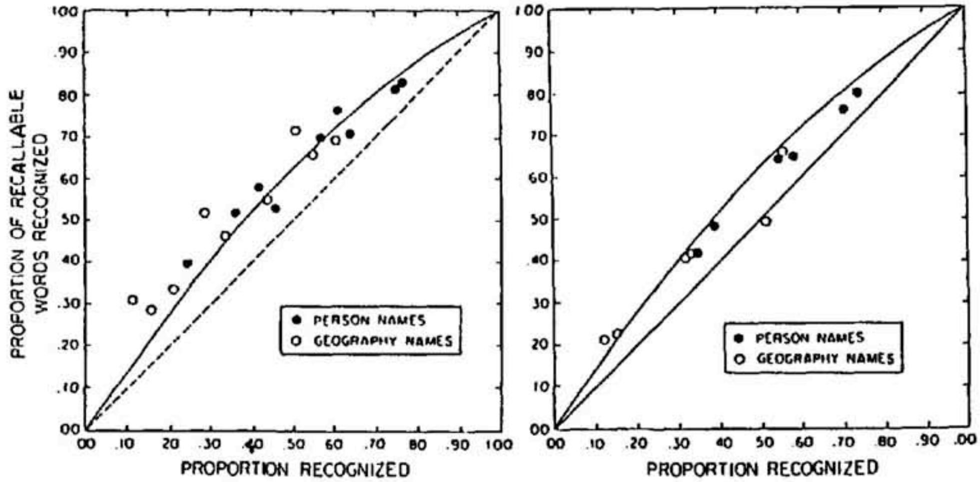


Figure 5. Proportion of recallable words recognized as a function of recognition for the person and geography names in Nilsson, Law and Tulving's (1988) experiments. (The left panel shows the raw data. The right panel shows Nilsson et al.'s data after having been homogenized by means of the Flexser [1981] procedure, which controls for item and subject selection effects. From "Recognition failure of recallable unique names: Evidence for an empirical law of memory and learning" by L. G. Nilsson, J. Law, and E. Tulving, 1988, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* pp. 271, 273. Copyright 1988 by the American Psychological Association, Inc. Adapted by permission.)

truncated unit normal distribution. Then each item was renormalized so that the self dot products were 1. The composite trace was set up as follows, where an asterisk denotes convolution as before and *lex x* refers to lexical item x:

$$\mathbf{T} = .5(\text{lex } 1 * \text{lex } 1) + .5(\text{lex } 2 * \text{lex } 2) + (\text{lex } 1 * \text{lex } 2) + .5(\text{lex } 3 * \text{lex } 3) + .5(\text{lex } 4 * \text{lex } 4) + (\text{lex } 3 * \text{lex } 4) + .5(\text{lex } 5 * \text{lex } 5) + .5(\text{lex } 6 * \text{lex } 6) + (\text{lex } 5 * \text{lex } 6) + \text{random noise}.$$

To recall, each of Cue Items 1, 3, and 5 were correlated with the trace, and the vector that was retrieved was matched to every item in the lexicon except the cue itself. The item that had the highest resonance score, above a threshold of 0.0, was said to be the item that was recalled. If that item was 2, 4, and 6, for Cues 1, 3, and 5, respectively, then recall was said to have been correct.

To recognize, the Items 2, 4, and 6 were correlated with the trace. The vector resulting from the correlation process was then matched by taking the dot product between the retrieved item and the probe itself. The criterion for a match was varied to be 0.0, 0.2, 0.4, 0.6, 0.8, or 1.0. If the degree of match exceeded the criterion, then the probe was said to be recognized (i.e., a hit).

A 2 × 2 contingency table was tabulated that kept track of the number of times items were recalled and recognized, recalled but not recognized, recognized but not recalled, and neither recognized nor recalled. From this contingency table the overall probability of recognition, the probability of recall, and the probability that recallable items were recognized was computed.

Results. Figure 6 shows the simple probabilities of recognition and recall as a function of criterion shifts in the recognition process. It can be seen that the criterion affected recognition in the expected way. If the criterion was low, recognition was high, whereas if the criterion was high, recognition was low. Also, as expected, there was no difference in the level of recall as a function of the recognition criterion. The small differences shown in the simulation results are a result of the fact that different simulations were used for each criterion level, and hence some random fluctuation occurs. Increasing the number of features in the representations of the items had a beneficial effect on recall. Number of features had less effect on recognition hit rate.

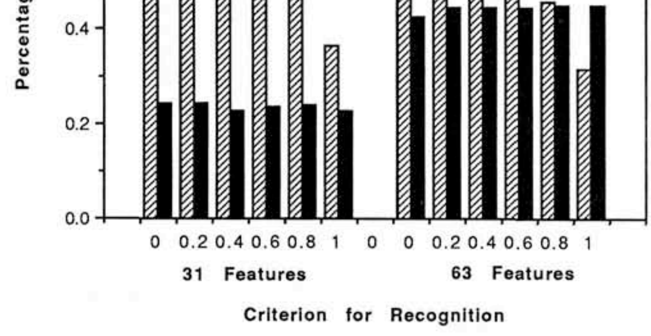


Figure 6. The probabilities of recall and recognition as a function of recognition criterion changes and the number of features in the item vectors. (From Simulation Series 1.)

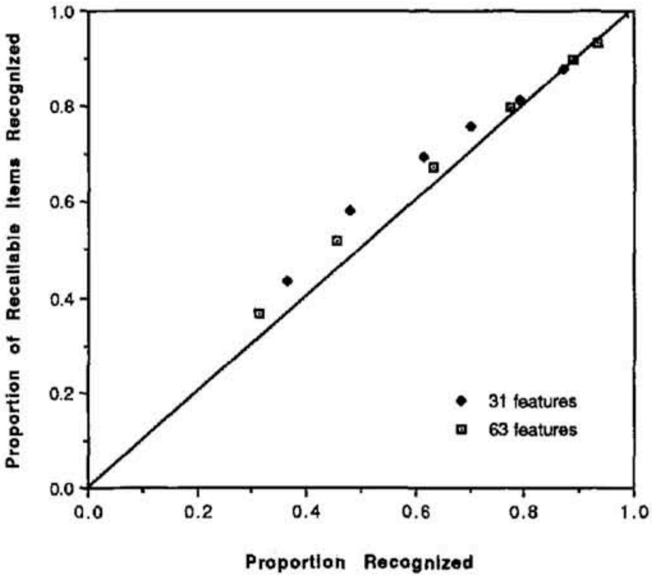


Figure 7. Proportion of recallable words recognized, as a function of recognition. (Variations are attributable to shifts in recognition criterion. From Simulation Series 1.)

Figure 7 shows the contingency relation between recognition and the recognizability of recallable items. The results compare favorably with those of Nilsson et al. (1988). Because there were no systematic item- or subject-selection effects in the model (each trial was independent, and there were no systematic differences in runs, aside from the manipulated number-of-features variable), the most appropriate comparison is with the data presented in the right panel of Figure 4 in which such factors have been removed from the data with human subjects.

Differences in the Recall Criterion

Although the CHARM model differs from classic generate-and-edit models in several ways, it is similar insofar as in both cases there is a criterion for the acceptable degree of match in both recognition and recall. In generate-and-edit theories, a number of response alternatives are generated as possibilities, and then a recognition check is performed to see which one, if any, was a member of the list under consideration. In the CHARM model, only one item is retrieved by the recall cue, but it may be noisy, systematically distorted, or a blend of several items. Its identity needs to be ascertained, and so it is matched against all of the response possibilities to see if any provides a good enough match to be considered the recalled item. Episodic recognition is not considered to be a substage of recall; rather, both recognition and recall involve retrieval (which is episodic), and then a (similar) decision process is enacted. In recognition, the match is only between the retrieved item and the probe itself, whereas in recall all of the items in the lexicon are response possibilities. In the simulations that follow the criterion for the goodness of match in recall is varied.

Simulation Series 2

Method. Simulation Series 2 was like Simulation Series 1 except that the recall criterion was also varied. This criterion gives the lower goodness of the match between the retrieved item and the best-matching lexical item necessary for recall. In the first simulation, this value had been set at zero. Here it was systematically varied from 0.2 to 1. The data shown below factorially cross recognition criteria ranging from 0.2 to 0.8, in steps of 0.2, with recall criteria ranging from 0.2 to 1.0, in steps of 0.2.

Results. Figure 8 shows the probabilities of recall with variations in the recall criterion. These summary data are collapsed across recognition criterion, and so each datum is based on 12,000 simulated observations. The criterion for recall, of course, does not affect recognition. The probabilities of recall vary systematically, with a higher probability of recall (and of intrusions, not shown) for low-criterion values and a lower probability of recall for higher criterion values. As before, recall is better when 63 rather than 31 features are coded for each item vector. The difference in recall as a function of criterion level is particularly obvious in the simulations with 63 features in the item vectors. The simulations with only 31 features are much more variable, and so the target item is the best match (regardless of criterion) less frequently.

The contingency data for each of the treatment combinations (now, not collapsed over recognition criteria) are presented in Figure 9, with each panel in the figure providing the results from separate variation in the recall criterion. The parameter within each panel is the recognition criterion. As can be seen from Figure 9, with few exceptions the simulated data fall on the recognition-failure function established empirically.

One might be able to vary the recall criterion empirically by giving guessing instructions or differential payoffs for correct recalls or penalties for intrusions. The prediction of the model is that this should alter the levels of recall but not the adherence to the function. As previously noted, it is also possible to vary the recognition criterion. Again the model predicts, and the data show, that this should alter the hit rate in recognition but not the adherence of the contingency data to the Tulving-Wiseman function. Figure 10 provides a summary of the recogni-

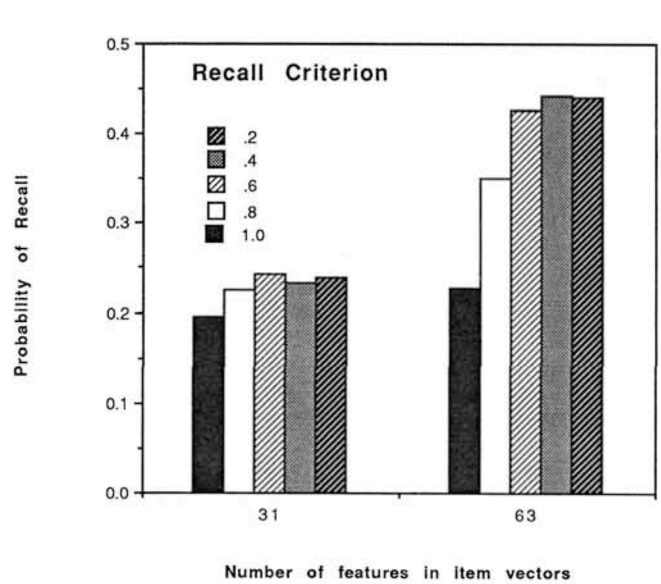


Figure 8. The probability of recall as a function of changes in the recall criterion and the number of features. (From Simulation Series 2.)

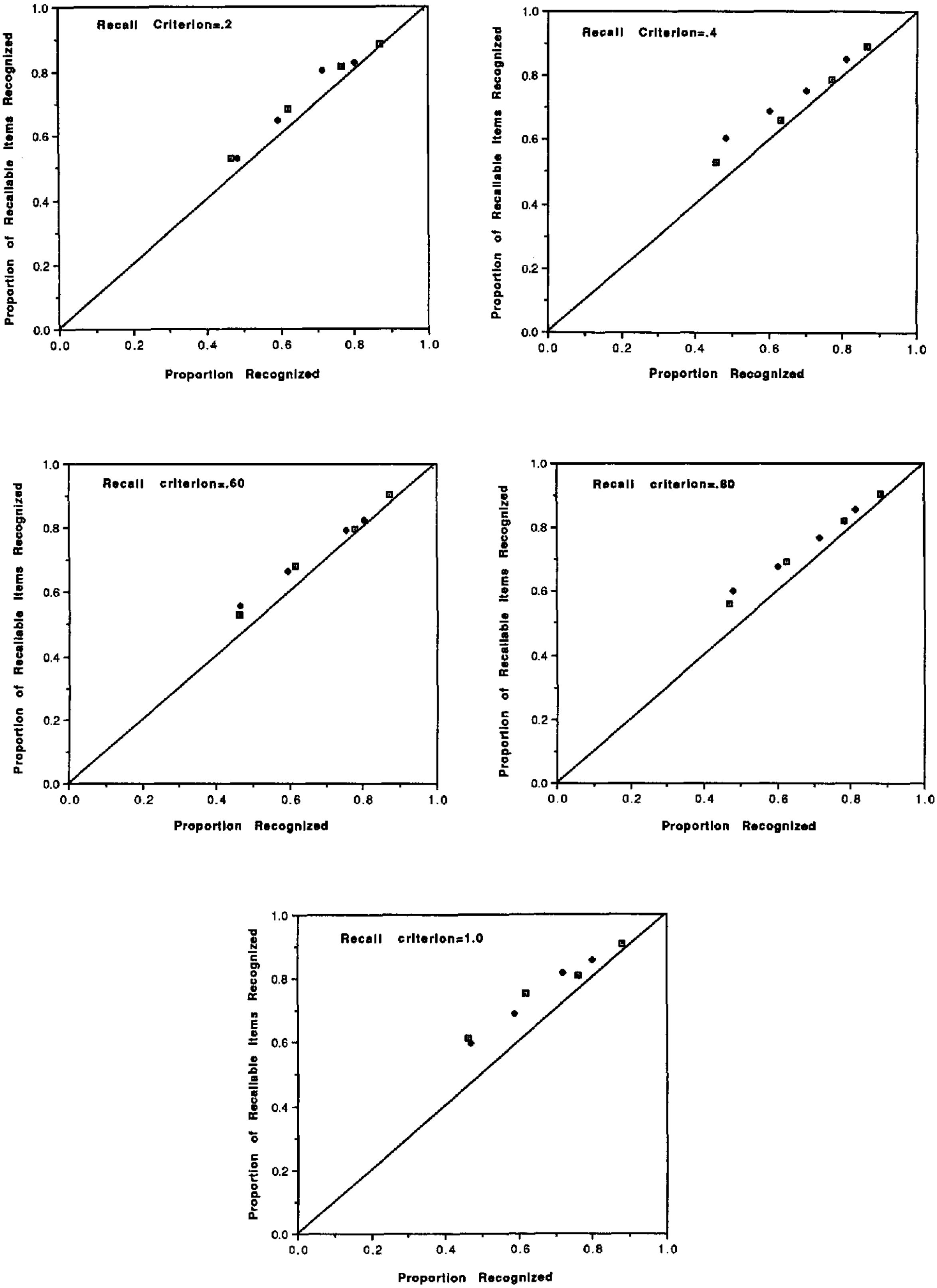


Figure 9. The effect of recall criterion on the relation between recognition and the recognizability of recallable items. (Recall criterion increases from the top to the bottom panel. From Simulation Series 2.)

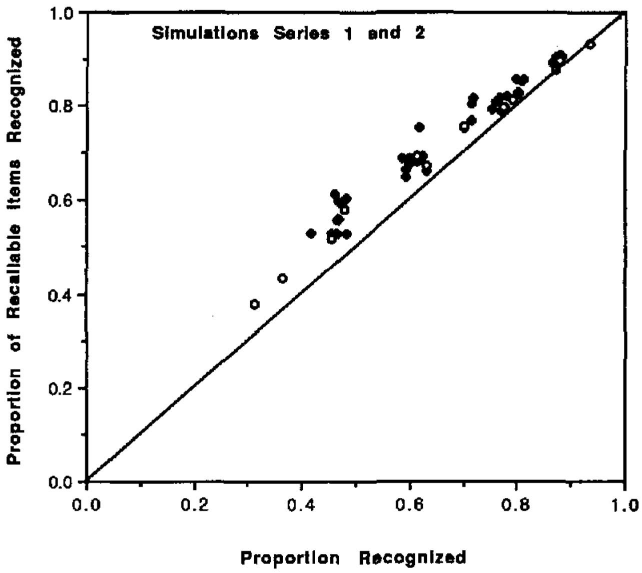


Figure 10. A summary of the proportion of recallable items recognized as a function of recognition, when both the recall and the recognition criteria are varied. (From Simulation Series 1 and 2.)

tion-failure results simulated by changes in the recognition and recall criteria from the first 2 simulation series.

Variations in the Weighting of the Recognition Autoassociations

The level of recall may be higher than (e.g., Wiseman & Tulving, 1976) or lower than (e.g., Postman, 1975) the level of recognition, and still the relation between recognition and the recognizability of recallable words obtains. Rabinowitz, Mandler, and Barselou (1977), too, have provided experimental conditions in which recall exceeds the recognition hit rate, with contingency results close to the Tulving-Wiseman function. Although there are other ways to vary the levels of recognition and recall in the model, changing the weighting of the interitem associations as compared to the autoassociations is perhaps the most straightforward. Before simply allowing these weightings on the interitem associations and on the autoassociations to be free to vary, however, one would like some experimental evidence indicating that the coding on the information responsible for recognition and on that underlying recall may be separable and may be under subject control, as these weighting parameters imply. It could be that anything that affects recall necessarily affects recognition in the same way. If such a result were consistent, it would suggest that an undecomposable single mnemonic representation underlay both processes. Such an undifferentiated view of the mnemonic information underlying recall and recognition contrasts with the view proposed here. In the formulation of the model under investigation, different kinds of associations—interitem associations and autoassociations—are primarily responsible for recall and recognition, respectively. Focused attention on the characteristics of the individual items should increase the weightings on the autoassociations. Integrative processing, or especially attention to the relations between the items, should increase the weightings on the interitem associations. In short, the two tasks should be dissociable, at least insofar as improvements in one task should not necessarily also result in improvements in the other.

A study by Carey and Lockhart (1973) is important for the dissociation between the overall levels of recognition and recall. In Carey and Lockhart's experiment, subjects were given a number of lists in which recall was required, as compared with lists in which recognition was required. On the critical list, subjects were either shifted to the alternate test or not shifted. Carey and Lockhart found that the expectation of a recall task hurt recognition performance, although there was no main effect of the expectation on the recall performance. There was a difference as a function of task expectation on which parts (categories) of the list were recalled. These results suggest that there is some subject control of encoding that is task specific and that the recall and recognition tasks have different requirements. The results are not perfectly straightforward for the present purposes, however, because free recall rather than cued recall was used. Wicker (1970) showed differences in recall as a function of the materials used, specifically that picture–word differences were attributable to stimulus recognition rather than to the associative component of paired-associate recall. Although a difference in level of recall for word and picture stimuli showed up in the simple recall data, when these data were conditionalized on stimulus recognition, the materials manipulation showed no effect. Bower (1970) showed a difference between recognition and paired-associate recall of a related nature. He found that the use of interactive imagery had no impact on stimulus recognition although it had a large effect on associative recall. These findings are generally consistent with the idea that the interitem associations and the autoassociations may be treated differently.

In the model, although the autoassociations and the interitem associations interact (a property that is discussed in more detail shortly), the former primarily underlie recognition performance, whereas the latter underlie recall. Thus, one may affect recognition by altering the weightings on the autoassociations. In Simulation Series 3, the weightings on the autoassociations were systematically varied. In Simulation Series 4, the weightings on the interitem associations were varied.

Simulation Series 3

Method. A lexicon of 50 unrelated items, each consisting of 63 features, as in the previous simulations, was set up. The items were convolved and entered into the composite memory trace according to the scheme given in the previous simulations, except that the weighting on the autoassociations was systematically varied rather than being fixed at 0.5, as had been the case previously. These weightings varied from 0.0 to 1.0, in steps of 0.1. For each of these simulations, the weighting on the interitem associations, on which recall depends, was held constant at 1. The criterion for recall was held constant at 0.0 (and hence recall depended only on the item being the best match). The criterion for recognition was also held constant for these simulations, at 0.8. The simulations were run through 200 replications, and there were three pairs in each list. Hence, each data point shown in Figures 11 and 12 is based on 600 observations.

Results. Figure 11 shows the simple probabilities of recall and recognition as a function of weighting changes on the au-

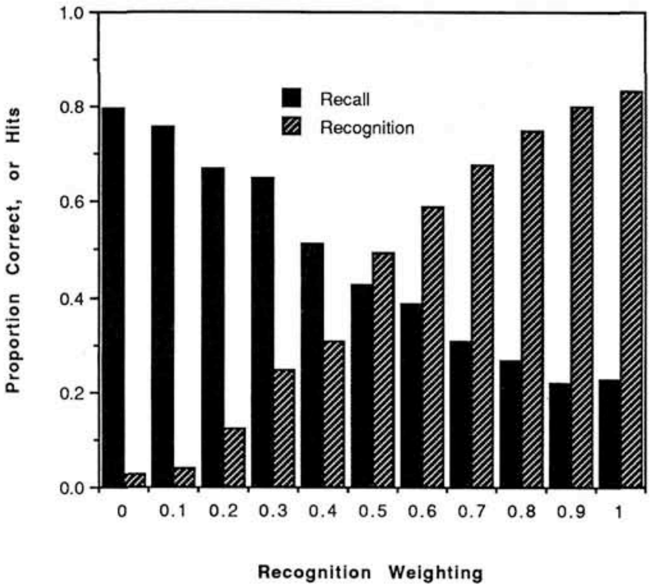


Figure 11. The probabilities of recall and recognition as a function of changes in the weighting of the autoassociations primarily underlying recognition in the model. (From Simulation Series 3.)

toassociation. One might, at first, expect that weighting changes on the autoassociations would only affect recognition —with a higher weighting giving rise to a higher level of recognition. In fact, a higher weighting does result in a higher level of recognition, but it also results in a lower level of recall. This trade-off occurs because the information for recognition and recall is stored in the same trace, and thus increasing the strength of one type of information, provided it is unrelated to the other type, as it is in this simulation, increases the interference on the other. This interference effect is particularly pronounced in recall because it may cause the wrong lexical item to be selected.

Figure 12 shows the recognition-failure functions that resulted from this series of simulations. The contingency data adhere quite nicely to the function, despite large differences in the level of recognition. Whether recognition is better than recall or vice versa in the model does not affect the relation between recognition and the recognition of recallable items.

Variation in the Weighting of the Recall Interitem Associations

Simulation Series 4

For the sake of completeness, I ran the analogous simulations to the preceding series with the weightings on the interitem associations responsible for recall varying (and the recognition autoassociative weightings being held constant at 0.5). This simulation series was set up exactly like the preceding one, except that the weightings on the interitem associations were varied from 0 to 1 in steps of 0.1. The recognition criterion was fixed at 0.8, as before. The recall criterion was fixed at 0.0 so that the best-matching item above that criterion was taken to be what was recalled. I ran this first through 200 replications and then through 2,000 replications.

The probabilities of recall and recognition are presented in Figure 13. The level of recall varied with weighting, as expected: The higher the weighting on the interitem association, the higher the level of recall. One might have expected, given the preceding series, that the increase in recall would result in a

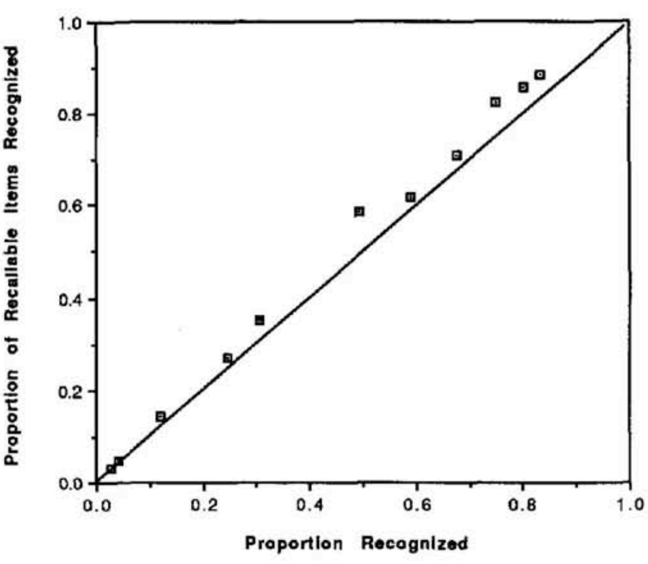


Figure 12. Proportion of recallable items recognized as a function of recognition, when the weightings on the autoassociations vary. (From Simulation Series 3.)

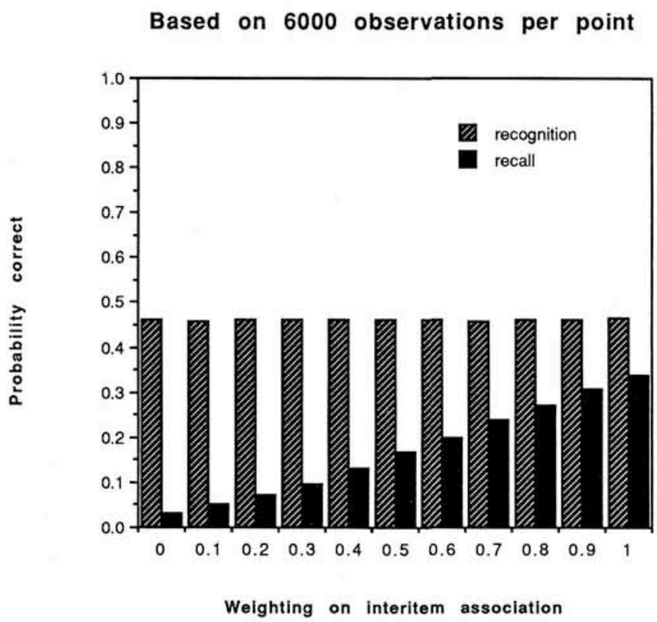


Figure 13. The probabilities of recall and recognition as a function of changes in the weightings of the interitem associations primarily underlying recall performance. (From Simulation Series 4.)

decrease in recognition. Figure 13 shows that the weighting of the interitem associations had little effect on the recognition hit rate, however. The increased weighting of the interitem associations causes a small increase in the variance of the entire trace, but it has no effect on the mean resonance between the retrieved item and the probe. The mean value of the resonance of the item retrieved to the probe was 0.76 in the case of a 0.0 weighting on the interitem associations and 0.76 in the case of a weighting of 1.0, but the *standard deviations* of these resonances were 0.42 and 0.60, respectively. The increase in variance has a much greater effect on recall than on recognition. In recall, all of the alternatives in the lexicon are available as responses. Given an increase in variance, the chance that a wrong lexical item will be selected increases considerably. In recognition, though, only the probe is considered, and thus increases in variance have fewer opportunities to exert themselves.

I included some lures in this simulation because I expected to show a small increase in the false-alarm rate given the increase in weighting of the interitem associations. In the entire first simulation series there were 1,100 lures presented for recognition. However, given the same criterion of 0.8 that was used for recognition hits, the false-alarm rate was zero. With the second simulation, replicating 2,000 times, similar results (slightly more reliable) obtained. The false-alarm rate was still zero though this time there were 11,000 opportunities. This and a variety of other simulations I have conducted on related issues indicate that recognition is less sensitive than recall to alteration in the weightings of other associations in the trace.

The contingency relation between recognition and the recognition of recallable items produced by this simulation is shown in Figure 14. Because there was little variability in the level of recognition, the scores are not spread over the entire range. Nevertheless, these contingency data cluster where one would expect given the empirical recognition-failure function.

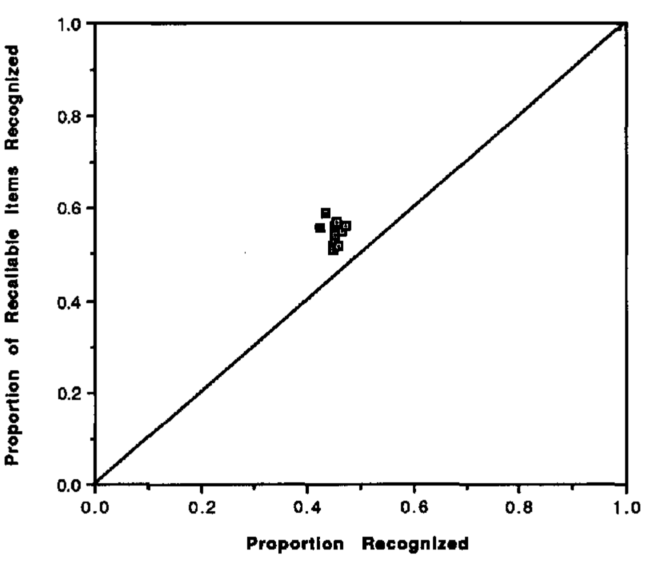


Figure 14. Proportion of recallable items recognized as a function of recognition, when the weighting on the interitem associations varies. (From Simulation Series 4.)

The Effect of Noise

A number of psychological phenomena are reasonable candidates for modeling with the addition of noise. For example, the passage of time might be represented as the addition of noise into the composite trace. Tulving and Watkins (1977), in a recognition failure experiment with unique targets (following Reder et al., 1974) such as "you can sometimes eat it but never sit on it—CACTUS, hairy on the outside but delicious on the inside—COCONUT, a very busy biological matchmaker—ENZYME," tested half of such a list immediately and the other half at a delay of 1 week. The immediate levels of recall and recognition were very high, and both of these levels dropped considerably at a delay. Both sets of data were consistent with the Tulving-Wiseman function, though the immediate test data were near ceiling and so are difficult to interpret. Many other experiments have used an immediate test and have shown data falling on the function, so the question is really whether the interference that accrues over time produces deviations from the function. The data indicate that it does not. Donnelly (1988) tested 4 weeks after study. These results, too, conformed to the recognition-failure function. Nilsson et al. (1988; shown in Figure 5) tested 1 week after study with results on the function. Finally, Muter (1978) used cue–target pairs that were learned preexperimentally. He also found data that fell on the function. It appears that delay in testing does not alter the adherence of the data to the Tulving-Wiseman function.

A second situation that might be modeled by the addition of noise in the composite memory trace is the effect of aging on memory performance. There are, of course, other hypotheses about the nature of the impairment in memory that are often found with aging, so I wish to make no strong claim that the addition of noise is the only way in which one might model this factor. It does, however, seem like a plausible default candidate for modeling how memory may be impaired. Rabinowitz (1984) tested aging patients as compared with normal adults in the recognition-failure paradigm. There were differences in recognition and especially recall performance. Nevertheless, the contingency data showing the relation between recognition and the recognizability of recallable items fell on the function for both subject populations. These experimental results converge on the conclusion that although the addition of noise should affect performance, and particularly recall performance, it should not result in marked deviations from the recognition-failure function.

Simulation Series 5

Method. This simulation was set up in a manner similar to the previous simulations except that, in addition to the three interitem convolutions underlying recall and the six autoconvolutions underlying recognition, an additional irrelevant pair was convolved as a noise item and added into the composite trace. The weighting on this noise term was varied to be 0, 1, 2, or 3. The design of the simulation was a $2 \times 2 \times 2 \times 4 \times 3$ factorial design in which the factors were Task (recall or recognition), Recognition Criterion (0. or 1), Number of Features in the Item Vectors (31 or 63), Noise Level (0, 1, 2, or 3), and Pairs (3). With 1,000 replications of each list, collapsed over pairs, each datum is based on 3,000 observations.

Results. The probabilities of recall and recognition are pre-

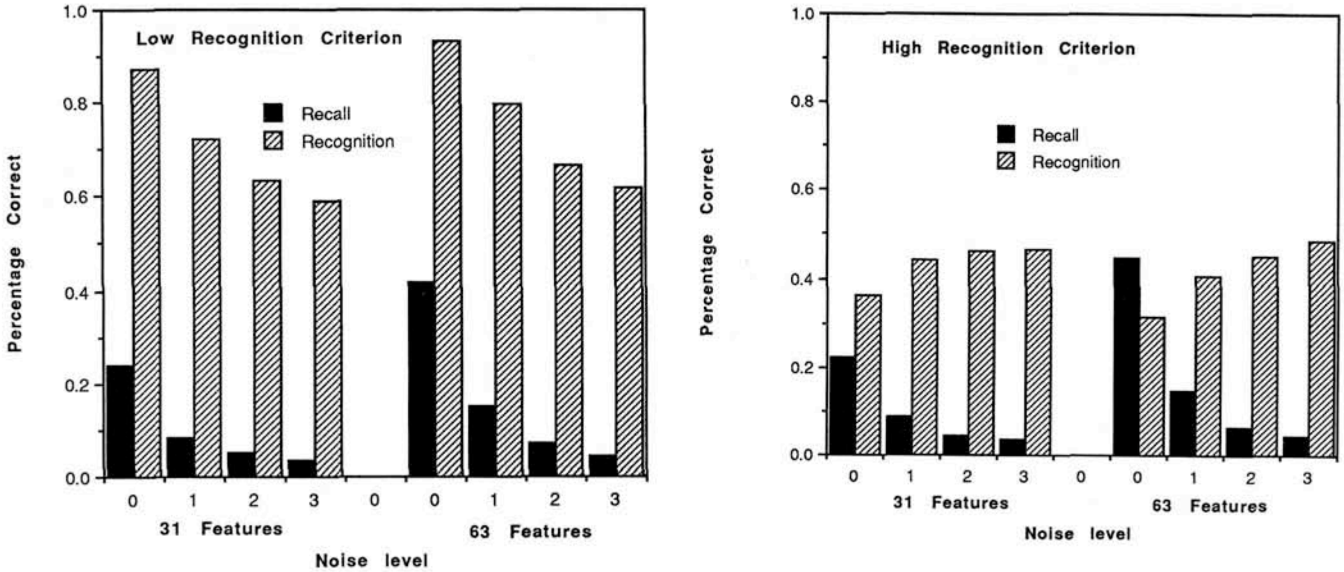


Figure 15. The effect of noise on the probabilities of recognition and recall. (From Simulation Series 5.)

sented in Figure 15. As can be seen from this figure, the probabilities of recall decreased with an increasing amount of noise. The effects of noise on recognition were not as simple as were the effects on recall. With a low recognition criterion there was a lower recognition hit rate with high noise than with low noise. With a high criterion, however, the recognition hit rate increased with increasing noise. The reason for both of these effects is that although the *mean* resonance score of the retrieved item to the probe was the same in all of the noise conditions, the *variance* of the resonance scores increased with increasing noise. Thus, there were more observations at a high noise level than at a low one that were below the low criterion: An increase in noise resulted in a decrease in hit rate with the low criterion. There were also more observations with a value exceeding the high criterion at a high rather than a low noise level: An increase in noise resulted in an increase in hit rate. In an additional simulation, I chose a single recognition criterion near the mean of the resonance scores. The recognition hit rate in this case was fairly stable, despite differences in the amount of noise.

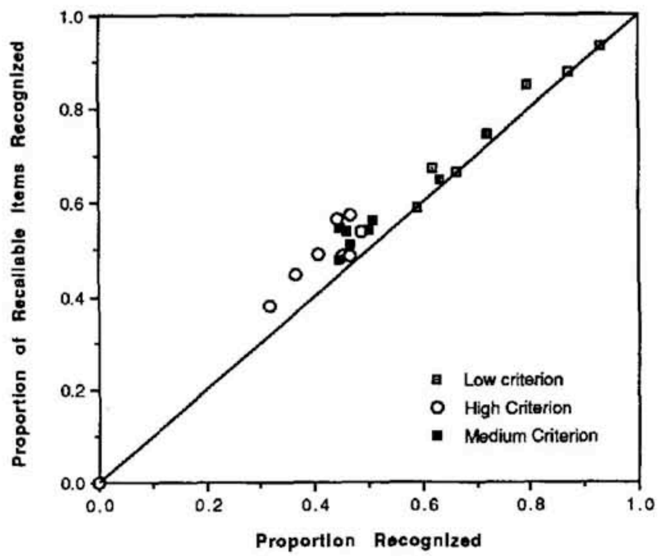


Figure 16. Proportion of recallable words recognized as a function of recognition, when the amount of extraneous noise varies. (From Simulation Series 5.)

The recognition failure results from the simulations with high, medium, and low criteria and varying levels of noise are shown in Figure 16. As might be expected with such low probabilities of recall on which the conditional data are based, the results are noisier than in the previous simulations. Nevertheless, they conform reasonably well to the empirical function. As seems to be true in the data, the amount of extraneous noise in the model does not substantively alter the relation between recognition and the recognizability of recallable items.

List Length

List length is a standard memory variable that can be modeled easily by varying the number of pairs in the model. The effect of list length in recognition and recall has been investigated in a study reported by Horton, Pavlick, and Julian (1989a, 1989b). Their finding is straightforward with respect to the recognition-failure function: The data were on it.

Simulation Series 6

In this series the number pairs in the trace were varied from 2 to 6, along with the recognition criterion, which ranged from .2 to .8. The number of features in the vectors was 15 or 63; and the simulation was replicated 500 times. As can be seen from the probabilities for the 63 feature vectors (shown in Figure 17), increasing the number of pairs decreased the recall levels but

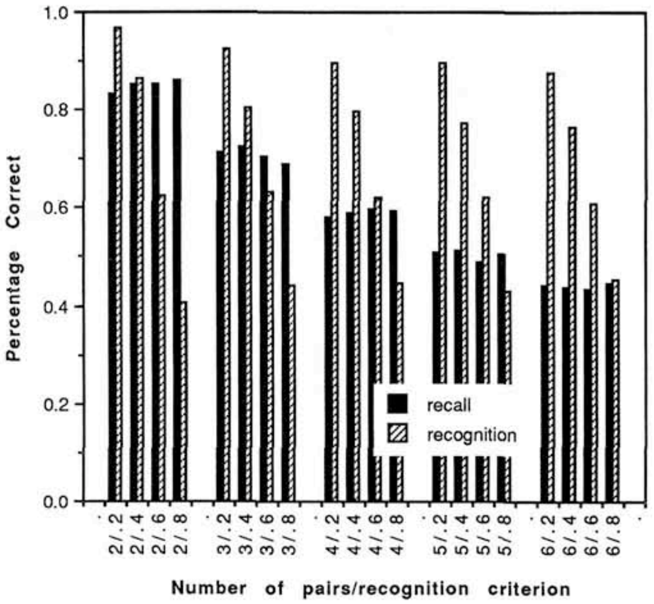


Figure 17. The effect of list length and recognition criterion on the probabilities of recognition and recall. Item vectors consisted of 63 features. (From Simulation Series 6.)

did little, overall, to the recognition hit rate. As was the case with the previous two simulations, increasing the number of pairs increased the variance of the resonance between the retrieved item and the probe but had no effect on the mean resonance. The contingency data (shown in Figure 18) are close to the empirical function.

Probabilistic Encoding of Features

The idea that people may not encode all possible features but will over time, sample features of the item is a very powerful construct that has enjoyed considerable success in explaining much data about human memory. The idea goes back to Estes's (1955) stimulus sampling theory. It was modified and elaborated by Bower (1972) and by Martin (1968) as encoding variability theory and had considerable explanatory power in that context. More recently, Murdock (1989) has used another variant of this probabilistic encoding idea to explain some of the effects of different presentation rates, and Bjork and Bjork (in press) have used it to help explain the effects of retrieval on later performance. The core idea is a highly plausible one. Although there may exist for items like words or sentence fragments an idealized replete potential mental representation, under most normal circumstances the events of memory will not be fully encoded. Rather, only some subset of the features that could be encoded for a given item will actually be sampled. If this characterization of the encoding process is correct, then it is important to see whether the recognition-failure function, as generated by the model, is robust with feature sampling. In all of the previous simulations, all of the features that were assumed to represent particular items were convolved and entered into the composite trace. In these simulations, the proportions of features that were encoded for particular items were systematically varied.

This manipulation in the model is akin, perhaps, to encoding time manipulations in experiments (see Murdock, 1989). It could also relate to an age variable where older subjects might not be capable of encoding as many features of the items in a given amount of time as younger subjects. As mentioned earlier, Rabinowitz (1984) showed that both young and old age groups produced results that conformed to the function.

Simulation Series 7

Method. A lexicon of 70 items was constructed by randomly assigning feature values to each item, as in the previous simulations. In the four major conditions of the simulations, these vectors started out with 15, 31, 63, or 95 features. Before the items were convolved and stored in the composite trace, however, random subsets of the features for each item were selected to be the sampled features. The proportion of features sampled at encoding ranged from 0 to 1 in steps of 0.2. The actual features that were sampled were randomly selected, and a different random selection was taken each time an item was coded. So if Lexical Item 1 was associated with Lexical Item 2 and also with itself, there were three tokens of Item 1 (all based on the Canonical Item 1) processed and entered into the trace. The sample of features for Item 1 was randomly different for each of these three tokens. At the time of retrieval, the canonical item vector was used as the retrieval cue, on the assumption that subjects had unlimited time to encode the cue. This cue was then correlated with the trace to produce a retrieved item that went through the usual decision process for recognition or recall. The recognition criterion was varied to be 0.2, 0.4, 0.6, or 0.8. Each list contained six pairs and was replicated 100 times, giving data based on 600 observations per point.

Results. The results showed wide variations in the simple probabilities. Performance increased with both the number of features in the canonical vectors and the proportion of features sampled. These results are shown in Panels A–D in Figure 19.

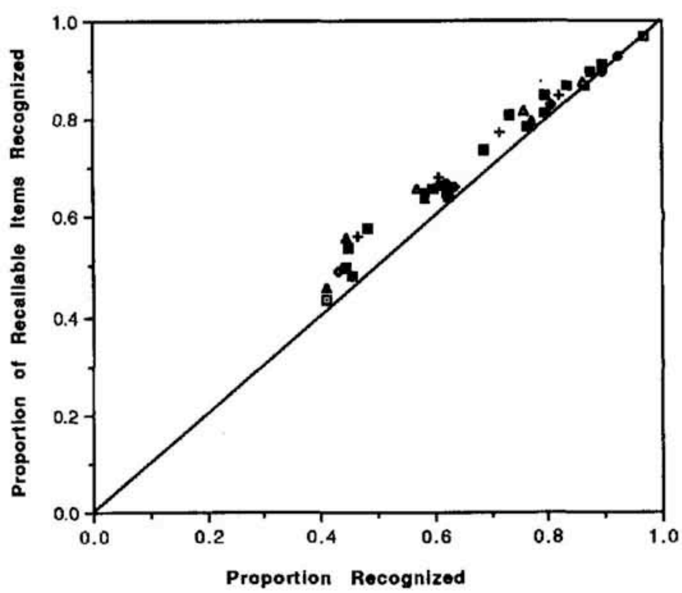


Figure 18. Proportion of recallable words recognized as a function of recognition, when the list length varied. (From Simulation Series 6.)

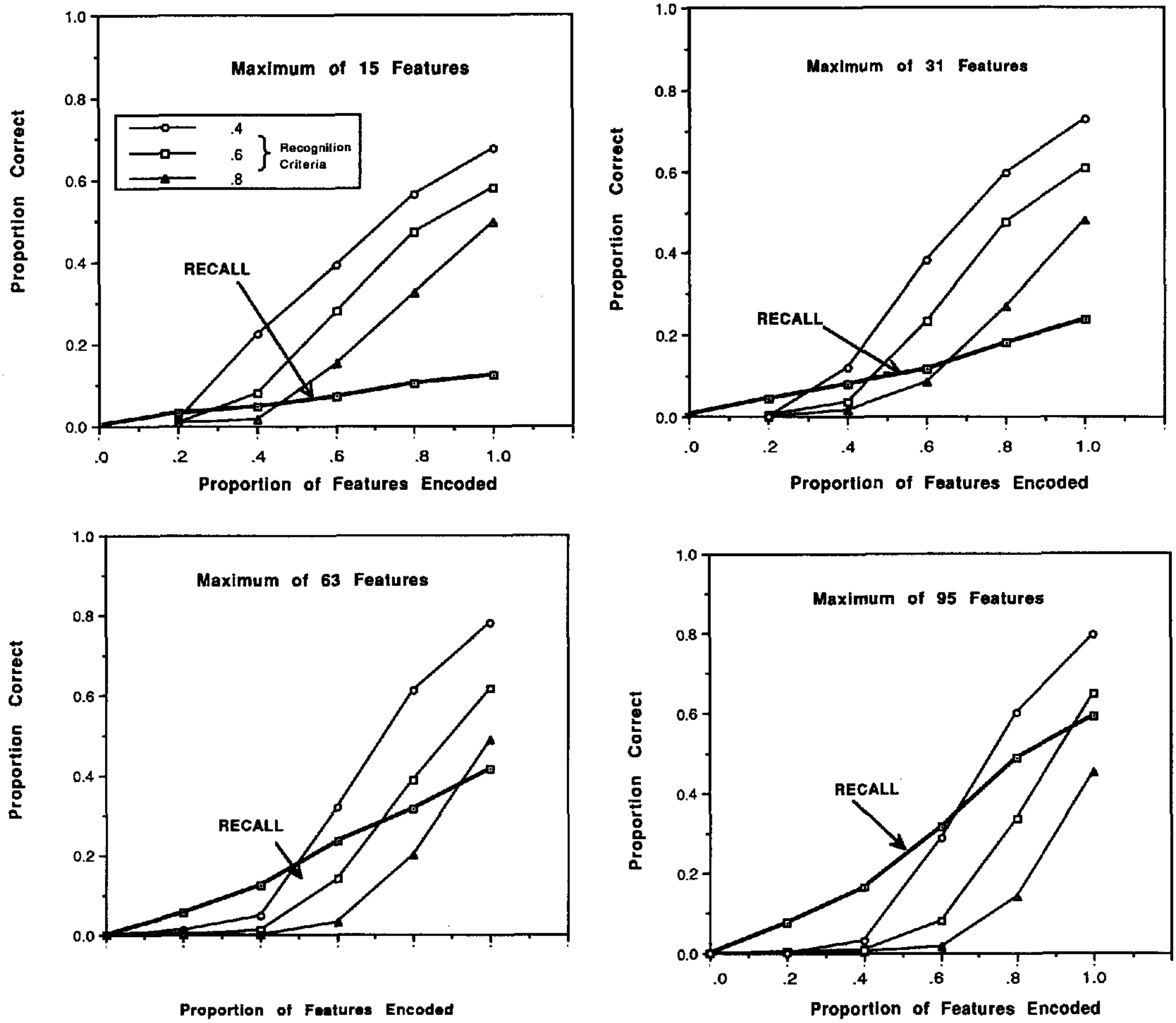


Figure 19. The probabilities of recall and recognition as a function of the number of features in the item vectors and the proportion of those features that are sampled. (From Simulation Series 7.)

The contingency data, showing a dependency relation between recognition and the recognizability of recallable items, are presented in Figure 20. There was also one point, within the 15-feature set of simulations, that fell slightly below the line of dependence. I have no explanation for it except that the model depends on random assignment of features, and occasionally the randomization can produce an aberrant result. The probabilistic encoding manipulation itself had no effect on the contingency relation. Regardless of what proportion of features was selected to be the sample encoded, the relation between recognition and the recognition of recallable items showed a small amount of dependency, approximating the Tulving-Wiseman function.

Interestingly, the number of features in the canonical vectors did have an effect on the dependency relation. Fewer features resulted in more dependence between recognition and the recognizability of recallable items than did more features. This observation is taken up in the next section of the article.

Exception to the Function: Impoverished Stimuli Rotely Encoded

Simulation Series 8

As was suggested in the previous simulations, the model in its intact form reveals a condition that produces an exception to the function in the direction of too much dependence. Dependency increases as the number of features in the item vectors decreases. This trend can be seen by examining the overall effect of the number of features from Simulation Series 8

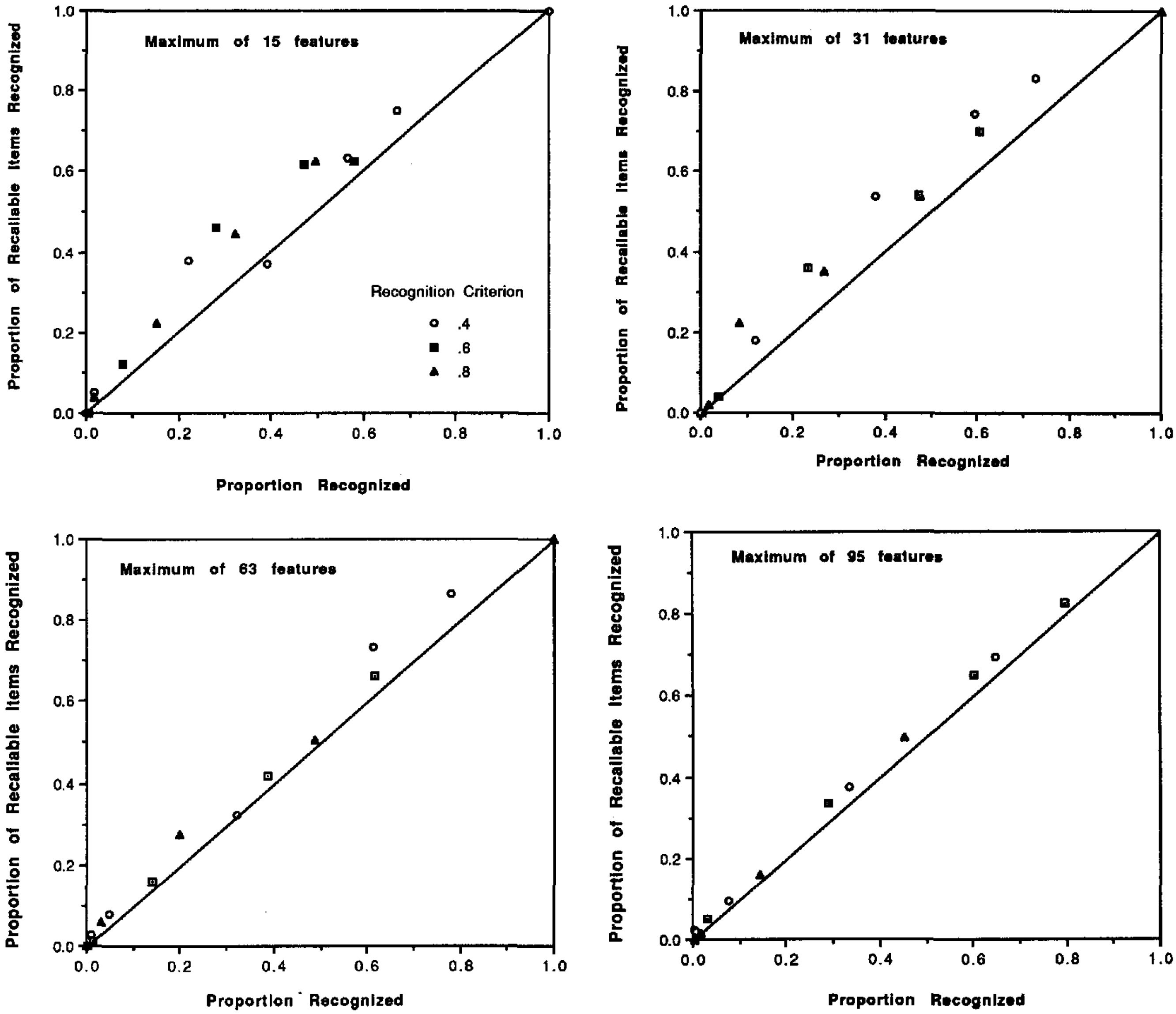


Figure 20. The proportion of recallable items that were recognized as a function of recognition when the number of features in the canonical representation varied (shown in separate panels) and the proportion of features sampled varied. (From Simulation Series 7.)

(shown in the right panel of Figure 21). When the items were constructed of very few features—seven features total—the contingency data showed considerably more dependence than is predicted by the Tulving-Wiseman function. This result is shown in the left panel of Figure 21.

Experimental Data

Four experiments show dependence of this sort. Begg (1979) showed that rote as compared with meaningful encoding affected the degree of dependence—meaningful encoding produced data on the function; rote encoding produced overdependence. Very recently, Bryant (in press) has provided similar results. When subjects encoded shallowly, more overdependence was shown than when they encoded deeply. They interpreted these results as indicating that differences in the integration between the cue and target were critical, but one might also interpret these findings in terms of the number of features encoded in the various conditions. Neely and Payne (1983) found results within a few percentage points of the function for related pairs, for unrelated pairs, and for famous names, but they found overdependence for unknown names. It would seem that unknown names may be meaningless in the sense of having very few features, whereas known names may be encoded by more features. Most compelling, though, is the work of Gardiner and Tulving (1980), which showed that impoverished stimuli, such as digits, paired with words (e.g., "27–BEING; 74–OPEN") produced overdependence. When subjects were given detailed in-

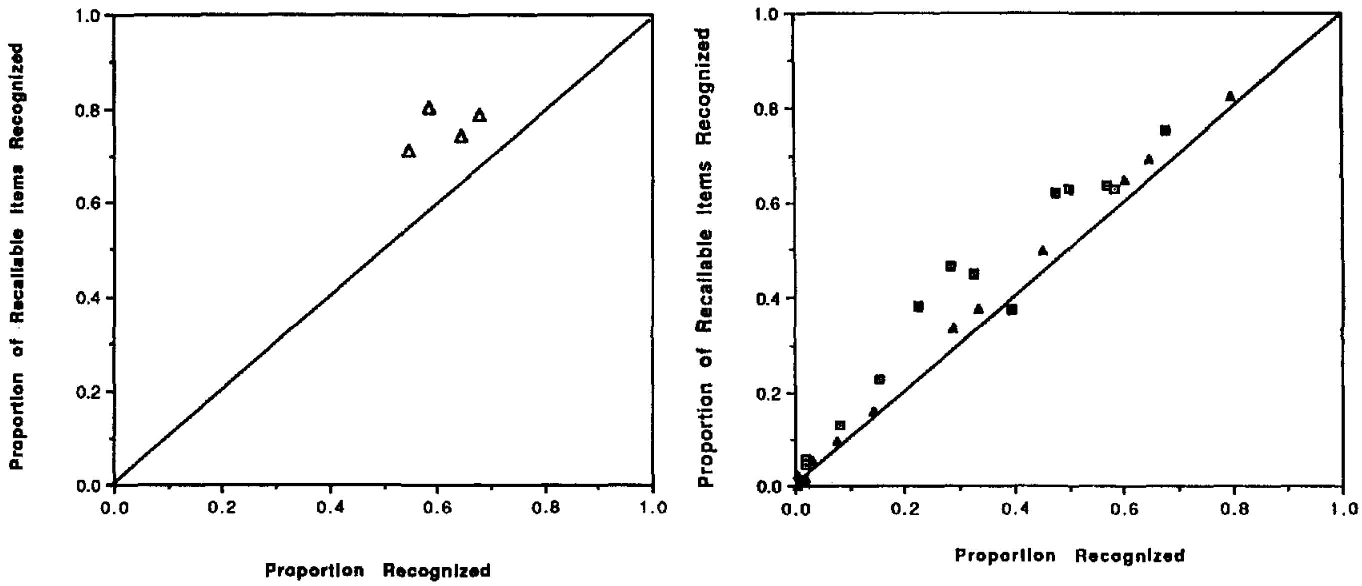


Figure 21. The left-hand panel shows the proportion of recallable items that were recognized as a function of recognition when the vectors consist of only 7 features. (The right-hand panel shows the systematic increase in dependency with a decrease in the number of features. Triangles show the results for 95 features, squares for 15 features. Summarized from Simulation Series 8.)

structions about how to convert these stimuli to meaningful events—by, for instance, thinking about them as dates for important life events—then (but not with rote encoding) data reverting to the function were produced. The example they gave of elaborative encoding was for the pair "63–HOW." Along with other mnemonic suggestions, they told subjects that

> on encountering such a pair in the study list, they might think of 63 as the year 1963. They might then remember that 1963 was the year in which John F. Kennedy was assassinated, and if they did so, they would then undoubtedly remember *how* he was assassinated (Gardiner & Tulving, 1980, p. 199)

A similar, but less extreme result—overdependence with rote encoding but data on the Tulving and Wiseman function with elaborative encoding—was found with abstract word pairs such as "Honor–ANXIETY." Gardiner and Tulving's and Begg's data showing overdependence with these impoverished stimuli are shown in Figure 22.

There is converging evidence that impoverished items like digits (when not given special elaboration) may differ from other stimuli in the number of features encoded. This evidence comes from a review of the experimental literature on the Rock replacement paradigm by Restle (1965). In that paradigm, data indicating all-or-none learning are consistently observed in paired-associate learning for impoverished stimuli like digit–letter pairs. Presumably, if there are very few features, the stimuli do not support partial learning. With more complex materials, including both nonsense syllables and word pairs, the data in the Rock paradigm indicate incremental learning (as would be possible if there were many features in the representations of the more meaningful materials). It appears that the complexity of the materials may map in a natural way into a theoretical construct like number of features. I postpone the question of why the model predicts more dependence when the items have fewer features until I have discussed why the model produces dependence between recognition and the recognizability of recallable items at all.

Exception to the Function: Similarity

The nature of the materials in the recognition-failure paradigm has caused considerable controversy. In many experiments, the weakly related pairs originally used by Tulving and Thomson (1973, e.g., "whiskey–WATER, glass–HARD, bath–NEED") have been used as the to-be-remembered cue–target pairs. However, in some cases the items have been highly related to one another and in other cases they have been unrelated word pairs. In a sequence of experiments, Rabinowitz et al., (1977) varied the similarity of the cue–target pairs from unrelated (Experiment 6) to low relatedness (in the remainder of their experiments). In all cases, however, the data were close to the function.

Neely and Payne (1983) also varied the relatedness of word pairs. Percentage correct was enhanced by relatedness in both recognition and recall—a near ubiquitous finding in the literature. The effect of relatedness was larger on recall (20.6%) than on recognition hit rate (1.8%). Neely and Payne reported overdependence with the related pairs and conformity to the recognition-failure function with the unrelated pairs. Although significantly different from the function, their results with the related pairs were off by only about 1%. So it would seem that if there is some overdependence due to the relatedness of the cue and target, it might be hard to detect.

Muter (1978), in an experiment that relied on subjects' preex-

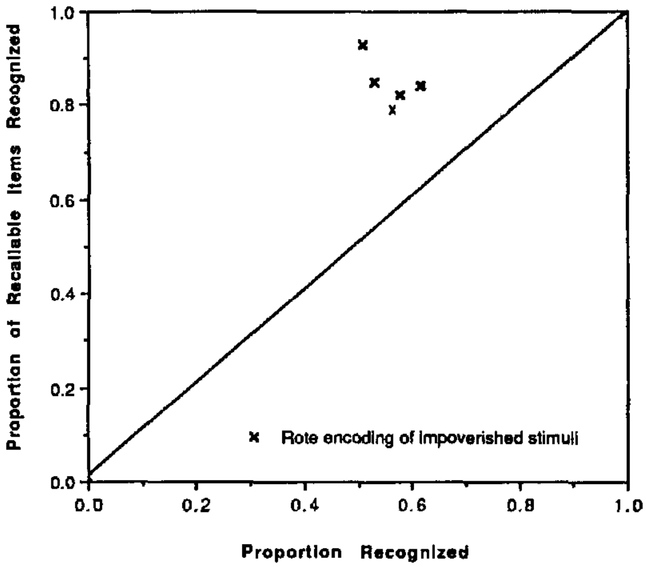


Figure 22. Data from Gardiner and Tulving (1980) and Begg (1979) showing overdependence with impoverished stimuli rotely encoded.

perimental knowledge, used cue–target pairs, such as "Maker of the U.S. flag: Betsy ROSS," or "Author of the *Last of the Mohicans:* James Fenimore COOPER." Presumably, as long as people knew who the famous person was, the cue and target were fairly redundant or similar, though because there could be other descriptions of, say, COOPER (such as "Cognitive psychologist working on the issue of visual mental representations: Lynn") that would also be appropriate, the cue and target cannot be said to be totally redundant in Muter's experiment. Subjects were asked to recognize from a list of common last names those of people who were famous or to recall the name given the uncapitalized part of the sentences as a cue. Muter found data close to the function.

Muter (1984) replicated this result with common names, but when the materials were rare names with their associated cues, such as "Siberian peasant and mystic who exerted great influence on Empress Alexandra of Russia: Grigory RASPUTIN; Danish philosopher, founder of existentialism: Søren KIERKEGAARD; (1854–1891), French novelist (*Madame Bovary*): Gustave FLAUBERT," a high degree of dependence (i.e., almost no recognition failure) was found. It seems that the uniqueness of the names may not be the critical factor, especially because Tulving and Watkins (1977) also used unique (but rather familiar and meaningful) words such as HANDSHAKE) and Nilsson et al. (1988) used unique proper names, and both groups found data on the function. There are two factors other than uniqueness per se that may be important in producing the overdependence. First, the rare names may have been unknown to many of the students participating in the experiment and hence were rather meaningless. If they did know the people, the information in the cue may have exhausted that knowledge for many of the students on many of the names. One may speculate that these items might have relatively few features. The common names may have had more semantic features and thus have been more meaningful. Second, the cue and the target were *highly redundant.* The cue "English economist best known for the theory that population tends to outrun food supply: Thomas" seems to be coextensive with the target MALTHUS. This experiment may thus provide an instance of very high similarity between the cue and target coupled, perhaps, with few features in the representations. It appears from these data that very high cue–target similarity produced considerable overdependence—a clear exception to the function.

Some other exceptions to the function in the direction of overdependence have been found with extremely high cue–target similarity. Nilsson and Shaps (1980, 1981) used homogeneous lists in which all of the pairs of words were highly similar to one another—category name–category exemplar pairs, such as "flower–ROSE, four-legged animal–HORSE." Nilsson and Shaps also found much more dependence between recognition and the recognizability of recallable items than would be expected on the basis of the Tulving-Wiseman function. Nilsson et al. (1987) conducted follow-up work on this overdependence result. Their experiment differed from the earlier one insofar as only some of the pairs in the list bore the relation category name–category exemplar. Hence, subjects could not engage in a recall strategy of generating the first category exemplar that came to mind. Although such a strategy would have been feasible in Nilsson and Shaps' (1980, 1981) earlier experiments, it would not consistently work in the later Nilsson et al. (1987) study. In this nonhomogeneous-list experiment, designed to eliminate such a guessing strategy, the highly similar pairs produced data that reverted to the Tulving-Wiseman function.

Finally, the most recent and straightforward exploration of the effects of similarity on adherence to recognition failure was made by Jones and Gardiner (1990). They asked subjects to study word pairs that were either semantically related, associated with a normative probability of 1%, or identical. They varied whether subjects were or were not encouraged to guess. It is not clear, from the point of view of this model, what should happen in the guessing-encouraged condition, so it will not be considered further. In the guessing-discouraged condition, however, only the identical (or maximum-similarity) condition showed a marked deviation from the Tulving-Wiseman function. As with the previously cited studies, the direction of this deviation was toward overdependence.

The effects of similarity are complex. Some studies in which similarity has been varied produce data that conform fairly closely to the function. There are indications, however, that very high cue–target similarity, especially with otherwise relatively meaningless materials, may result in more dependence than is usually found between recognition and the recognizability of recallable items.

Simulation Series 9

Method. To investigate the effect of similarity on the recognition-failure function in the model, three simulations varying the similarity

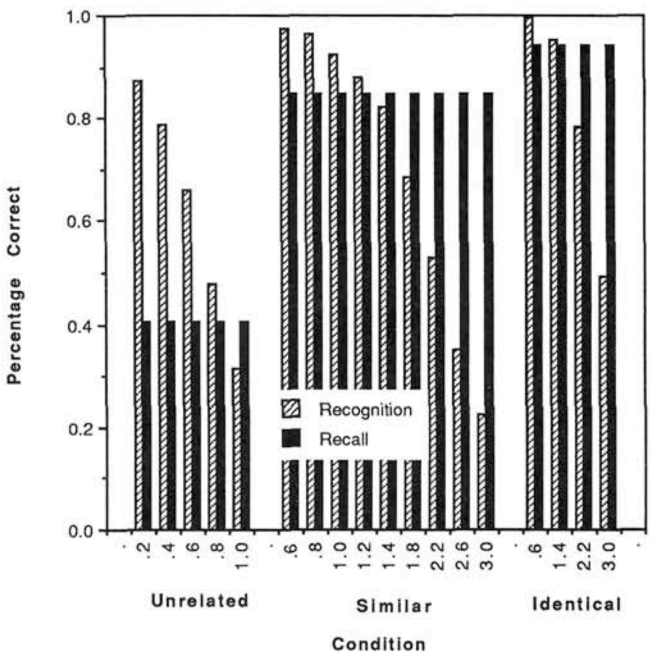


Figure 23. The probabilities of recall and recognition as a function of increasing cue–target similarity. (The recognition criterion is varied. From Simulation Series 9.)

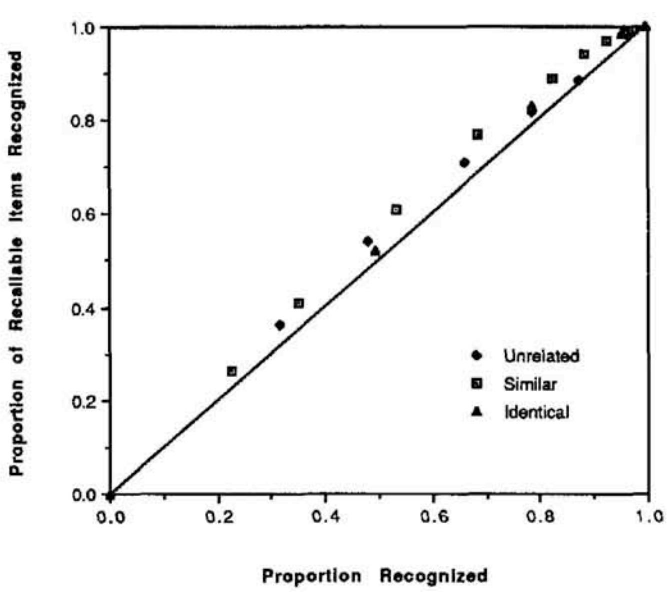
of the cue–target pairs were conducted. The items in the lexicon were initially constructed as random vectors consisting of 63 features. In the unrelated-pairs simulations they were left unchanged. In the similar-pairs simulations, there was a 0.6 chance that any given feature value in the first item in each pair would be replaced by the value of that feature in the second item. Items were unrelated between pairs, however. In the identical-pairs simulations, all of the features were identical in the cue and target but unrelated between pairs. For each simulation run, three pairs of items and their corresponding autoconvolutions were added into the composite trace, with a weighting of 0.5 on the autoassociations and 1.0 on the interitem associations. The criterion for recognition was varied from 0.2 to 0.8 in the unrelated case and 0.6 to 3.0 in the similar and identical cases.

Here, as in most other work on the model, the cue itself was eliminated from consideration as a response possibility in recall. If it were not eliminated, then the effects of similarity are not straightforward. Cue intrusions result and the simple probabilities are nonmonotonic with an increase in similarity. Some of these effects are discussed in Metcalfe Eich (1982, 1985). However, allowing that the cue itself can be eliminated, as it can be in most experiments, the effects are straightforward.

Results. The results of within-pair similarity on the probabilities of recall and recognition are shown in Figure 23. Both recall and recognition were improved by the increase in the within-pair similarity. As can be seen from the left panel of Figure 24, all three levels of cue–target similarity produced a slight dependence between recognition and the recognizability of recallable items. From these initial results I thought that similarity did not matter for the recognition-failure function, as seems to be the case from several of the experiments reported earlier. However, this conclusion was premature. In fact, cue–target similarity did have an effect on the recognition-failure predictions of the model.

High Similarity With Few Features

Having found that when the items have few features the extent of dependence increases, I decided to redo the simulations on similarity, using few features. This increased the chances of being able to detect a difference. The right panel of Figure 24



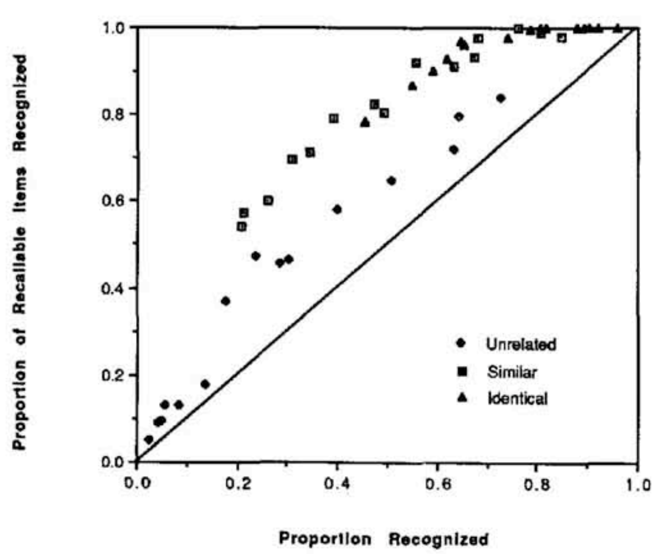


Figure 24. The effects of similarity on the proportion of recallable items recognized as a function of recognition. (The left panel shows the results of similarity with 63-element vectors. The right panel shows the effects with 15-feature vectors. From Simulation Series 9.)

gives the relation between recognition and the recognizability of recallable items as a function of similarity when the item vectors consisted of only 15 features. There was an effect of similarity in the model, though when many features are coded it is difficult to detect it.

Similarity between the cue and target produced a straightforward correlation between the cue in recall and the probe in recognition. As such, the resulting correlation between the two tests is unsurprising (and is, in fact, predicted by a number of other models, such as that of Flexser & Tulving, 1978). What is more surprising is the finding that there is a dependency relation in both the model and the data even when the cue and target are unrelated. The reason for this relation is discussed shortly.

What Causes the Recognition-Failure Function?

One strategy that may be used in understanding why the model produces the dependency function (and by inference, why people produce these results) is to violate central assumptions and see whether the result still holds. If it does, then that assumption was not critical; if not, then there is something interesting to be learned. Two critical assumptions are investigated here: (a) that recognition and recall both involve retrieval of, qualitatively, the same sort of representations from memory and (b) that the information underlying both recall and recognition be stored in the same composite memory trace.

Representational Retrieval in Both Recognition and Recall

In CHARM, recognition is formulated as a true retrieval process of the sort used in recall. An item, which is in the form of the initially encoded event—that is, represented as a vector, is retrieved in recognition as in recall. The fact that such a vector is generated as the basis on which a decision about previous experience is made is what I mean by retrieval of a representation. This is not the only way, nor, indeed, the most common way, to think about recognition memory. It is possible to make a recognition decision without retrieving representational information. Indeed, most models of recognition memory rely on mere strength decisions rather than on the reinstatement of representations. Mathematically, one can distinguish whether a representation is retrieved or a strength value is assessed because the former involves the reinstatement of a vector, whereas the latter entails only the production of a scalar. Here I look at whether representational retrieval in recognition as well as in recall is critical for CHARM's prediction of the Tulving-Wiseman recognition-failure function. In all previous simulations reported here, recognition was formulated as representational retrieval. In the simulations that follow, which are, in fact simulations of the TODAM model rather than the CHARM model, recognition is formalized instead as a global match process involving only the assessment of a scalar strength value.

In TODAM, as in CHARM, items are represented as vectors, and for recall, pairs of items are convolved with one another. For recognition, simple vectors (untransformed) are added into the composite memory trace. Instead of retrieving an item at time of recognition and then deciding whether that item matches the probe (as in CHARM), the recognition-decision process in TODAM consists of taking the dot product between the probe and the composite memory trace. A scalar is returned and no retrieval (in the sense that a memorial event is reinstated) takes place. As Murdock (1982, 1985, 1989), has documented in a variety of articles, this model does a rather nice job of accounting for much of the data within human recognition memory. The question is not whether this is a successful model (it is) but rather whether this formalization of the recognition process, in combination with the convolution–correlation recall process, will automatically give rise to the recognition-failure function of Tulving and Wiseman (1975). The equation for TODAM that is comparable to Equation 3 for CHARM is

$$\mathbf{M} = \mathbf{A}*\mathbf{B} + \mathbf{A} + \mathbf{B}.$$

Retrieval for recall is

$$\mathbf{A} \# \mathbf{M}.$$

Retrieval for recognition is

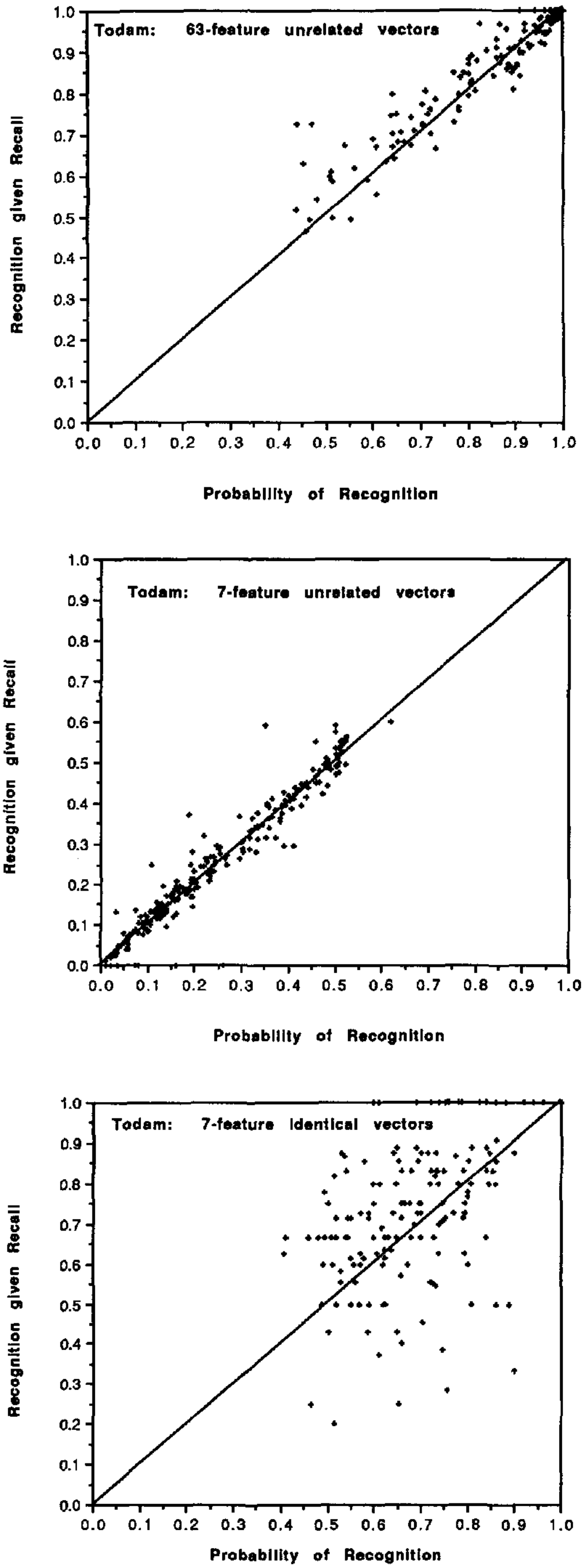
$$\mathbf{B} \cdot \mathbf{M}.$$

Although much work has been done on TODAM investigating weighting parameters—adding noise externally and studying decision processes, in the simulations that follow I investigated only this simple version of the model, which provides the most straightforward comparison to the basic schema used for recognition and recall in CHARM. Hence, this exercise should not be taken to be an exhaustive exploration of TODAM.

TODAM Simulations

The first simulation series was conducted by constructing a lexicon of 70 item vectors in a manner like that used in the previous simulations. The vectors in this initial series each had 63 features, like those in most of the aforementioned simulations of CHARM. Interitem associations were formed by convolving the first lexical item with the second, the third with the fourth, and so on. Instead of autoconvolving the item vectors for recognition (as was done in CHARM), the simple lexical vectors were added into the trace. The simulations were conducted as a factorial design, where the factors were Number of Pairs (5, ranging from 1 to 5), Recognition Threshold Necessary for a Yes Response (11, ranging from 0 to 1 in steps of 0.1), and Lower Boundary Necessary for Recall (3, ranging from 0 to 1 in steps of 0.5). This resulted in a total of 165 simulated conditions, each of which was replicated 50 times. To retrieve for recall, the first item in each pair was correlated with the composite memory trace, and the lexical item that provided the best match to the retrieved item was said to be what was recalled. If it was the second item in the pair, recall was said to be correct. To recognize, the probe (second item in each pair) was dotted with the trace. If the degree of match exceeded threshold (which varied as just mentioned) then a hit was said to have occurred. Contingency tables were constructed, as previously reported for CHARM, and the TODAM-predicted relation between recognition and recall was ascertained from them.

The results are shown in Figure 25, top panel. As can be seen, independence between recognition and the recognizability of recallable items was the result. Two other simulation series were



conducted with TODAM to further investigate changes that in CHARM had produced exceptions to the function in the direction of overdependence.

The first case of overdependence in CHARM resulted when very few features were used in the vectors. Accordingly, I ran the 165 simulated conditions again with a different initial randomization but with only 7 features in the item vectors. The results are shown in the center panel of Figure 25. As can be seen, the data were more dispersed in this series than in the series in which 63 features had been used, but independence still resulted. About half the points in this simulation series (as in the preceding one) were in negative dependence regions of the space—regions in which human data almost never appear.

Finally, the similarity between the cue and the target was a factor in CHARM that caused overdependence. Thus, I wanted to see if high cue–target similarity would result in a dependence relation in TODAM. I immediately took this to the limit of having the cues and targets be identical to one another—maximum similarity. I also used only seven features in this simulations series, because this produced the maximum dependence relation in CHARM. As can be seen from the bottom panel of Figure 25, this manipulation did increase the level of recognition. The hit rate tends to be higher than .65 rather than .40, as was found in the previous two TODAM simulations. The reason this manipulation increased the hit rate is that when both the A terms and the **B** terms are identical, it is the same as adding the A terms in twice. Thus, the dot product between the probe A and these two A terms is higher than between the probe and only a single A term, and the probability of exceeding a given threshold is accordingly increased.

The similarity manipulation did not result in dependence between recognition and the recognizability of recallable items. They were still quite independent, and still quite unlike the pattern shown by CHARM or shown by human subjects. It appears, from these simulation results, that the operations for encoding and retrieval need be the same in recall and in recognition and need involve the retrieval of a representation in recognition as well as in recall for the recognition-failure dependence to emerge.

Information Storage in a Single Composite Trace

The assumption investigated in this simulation was composite storage of the associations that underlie both recognition and recall in the same memory trace. The idea of a composite memory trace is one of the central or core assumptions in distributed models. In these simulations, I relax this assumption

Figure 25. Simulation results from Murdock's (1982) TODAM model for recognition and recall. (This model does not use representational retrieval in recognition. The top panel shows the results with 63-feature vectors; the center panel, the results with 7-feature vectors. The bottom panel uses 7-feature vectors and maximal similarity, i.e., identity, between the A and B term in each pair. From "A theory for the storage and retrieval of item and associative information," by B. B. Murdock, 1982, *Psychological Review, 89,* 609–626. Copyright 1982 by the American Psychological Association, Inc. Adapted by permission.)

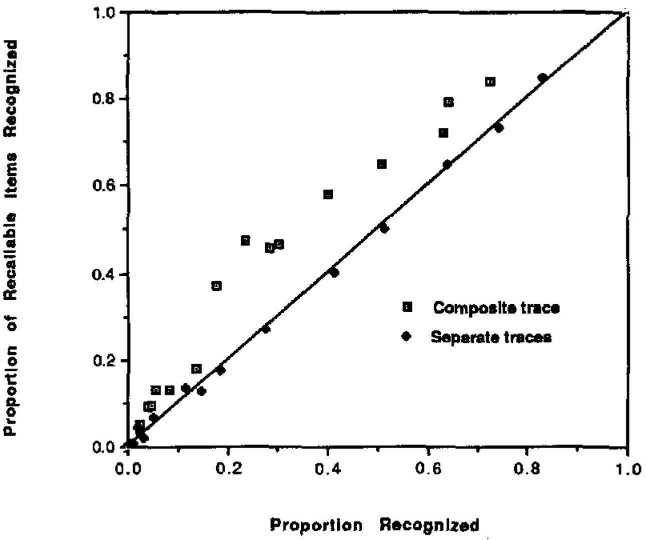


Figure 26. The proportion of recallable items that were recognized as a function of recognition (a) with a single composite trace and (b) with separate traces for recognition and recall. (From Simulation Series 10.)

and investigate the repercussions. All other assumptions were maintained. Thus, in the separate trace conditions the interitem convolutions underlying recall were stored in one trace, whereas the autoassociations underlying recognition were stored in another. In the single composite trace conditions all associations were stored in the same composite trace. The same items, decision processes, number of features, operations of convolution and correlation, and so on were used for both the single composite trace and separate trace simulations. Because the greatest dependency had been found previously with few features in the vectors, I used only 15 features here. The items were unrelated. Recognition criterion was varied, and each data point is based on 3,000 observations.

As is shown in Figure 26, with the single composite trace for both tasks the familiar recognition-failure dependence relation resulted. However, when the information underlying recognition was stored in a trace separate from that underlying recall the result was complete independence.

Discussion

The results of retrieval under the usual composite-trace assumption in the model, whereby the associations from both the interitem associations (primarily responsible for recall) and the autoassociations (primarily responsible for recognition) are stored together in the same trace, are given by the following equations, for a single pair of items. When **A** and **B** are the to-be-remembered pair, the trace is

$$T_{\text{composite}} = \mathbf{A}*\mathbf{B} + .5(\mathbf{A}*\mathbf{A}) + .5(\mathbf{B}*\mathbf{B}).$$

When the recognition of the target **B** is tested by probing with **B**, the result retrieved from the trace is

$$\begin{aligned} R_{\text{recognition composite}} &= \mathbf{B}\#[(\mathbf{A}*\mathbf{B}) + .5(\mathbf{B}*\mathbf{B}) + .5(\mathbf{A}*\mathbf{A})] \\ &= \mathbf{B}\#(\mathbf{A}*\mathbf{B}) + .5\mathbf{B}\#(\mathbf{B}*\mathbf{B}) + .5\mathbf{B}\#(\mathbf{A}*\mathbf{A}) \\ &= \mathbf{A} + .5(2\mathbf{B}) + 0 \\ &= \mathbf{A} + \mathbf{B} + 0. \qquad \text{[from equation 6]} \end{aligned}$$

So, the target **B** is retrieved, but there is an additional signal term **A** that is also retrieved.

When recall is tested, the cue is **A** (rather than **B**, as in recognition). The result of retrieval is

$$\begin{aligned} R_{\text{recall composite}} &= \mathbf{A}\#[(\mathbf{A}*\mathbf{B}) + .5(\mathbf{A}*\mathbf{A}) + .5(\mathbf{B}*\mathbf{B})] \\ &= \mathbf{A}\#(\mathbf{A}*\mathbf{B}) + .5\mathbf{A}\#(\mathbf{A}*\mathbf{A}) + .5\mathbf{A}\#(\mathbf{B}*\mathbf{B}) \\ &= \mathbf{B} + .5(2\mathbf{A}) + 0 \\ &= \mathbf{B} + \mathbf{A} + 0. \qquad \text{[from equation 6]} \end{aligned}$$

Once again the target item **B** is retrieved, but an additional signal term **A** is also retrieved. It is this additional signal term that results from composite storage of both the interitem associations and the autoassociation in the same trace that is at the heart of the recognition-failure correlation shown in the model.

To illustrate the importance of this term consider the equations for storage and retrieval that result under the separate trace assumption, in which independence obtained, as the simulations showed. For recognition, the trace (for a single event) is

$$T_{\text{recognition, separate}} = .5(\mathbf{A}*\mathbf{A}) + .5(\mathbf{B}*\mathbf{B}).$$

When the probe for recognition is given, as before, it is correlated with the trace:

$$\begin{aligned} R_{\text{recognition separate}} &= \mathbf{B}\#[.5(\mathbf{A}*\mathbf{A}) + .5(\mathbf{B}*\mathbf{B})] \\ &= .5\mathbf{B}\#(\mathbf{A}*\mathbf{A}) + .5\mathbf{B}\#(\mathbf{B}*\mathbf{B}) \\ &= 0 + .5(2\mathbf{B}) \\ &= \mathbf{B}. \end{aligned}$$

B is produced, but no **A** term is simultaneously produced. The separate trace for recall is

$$T_{\text{recall, separate}} = (\mathbf{A}*\mathbf{B}).$$

When the cue for recall, **A**, is correlated with the trace, the following is the result:

$$\begin{aligned} R_{\text{recall, separate}} &= \mathbf{A}\#(\mathbf{A}*\mathbf{B}) \\ &= \mathbf{B}. \end{aligned}$$

Again, the to-be-recalled item, **B**, is produced, but with no **A** term. This is the critical difference between the result of retrieval in the composite versus the separate storage conditions.

In the single composite trace situation, if the items are unrelated, this (spurious) **A** term has an expected correlation with the target **B** of zero. Why, then, should it produce a positive correlation between recognition and the recognizability of recallable words? Although the expected correlation between the correct item **B** and the **A** term is zero, the actual correlation varies randomly around zero. If the correlation between **B** and **A** is above zero, then the presence of **A** will bolster the resonance of **B**. If there is a negative correlation between **A** and **B**, then the presence of A will decrease the resonance of the target B. On any particular test, however, regardless of whether the actual value is above or below zero for any particular A, it will have the *same* effect in both recall and in recognition (because it is the identical term that is retrieved in both cases). Hence, it causes a correlation between recall and recognition.

Why Is There More Dependency With Fewer Features?

As noted earlier, even with unrelated items, as long as the information underlying both recognition and recall is stored in the same composite trace, the nontarget A term corresponding to the cue in recall is produced along with the target **B** term. It is this A term that accounts for the dependency between recognition and recall. In Simulation Series 8, it was found that the fewer features there are in the item vectors, the greater is this dependency. Overdependency with few features corresponds to the data from three studies, showing that impoverished stimuli rotely encoded produced overdependence. The reason for the relation between the number of features and the extent of dependency has yet to be explained.

As noted earlier, the expected correlation between the A items and the target **B** items is zero. However, the amount of *deviation* from the expected value of zero will depend on the sample size. Sample size, in this case, is the number of features in the vectors. With a small sample size, the deviation from the expected value will be greater; whereas with a large sample size a closer approximation to the expected value of zero correlation will be found.

A familiar example of such an effect of sample size is the hospital problem given by Tversky and Kahneman (1982). In this problem, subjects are asked whether a small hospital having few births per day or a large hospital having many births per day would have more days on which the proportion of boys is greater than 60%. The answer, known to all psychologists but not to most other people, is that the small hospital can expect more such days because there is greater variability from the expected value with a smaller sample. Similarly, this situation, although the expected correlation between **A** and **B** is zero, the divergences from this mean on any particular occasion are larger with a small number of randomly assigned features contributing to the correlation than with a large number. Because any divergence, whether positive or negative, has the same impact on recall and recognition (helping or hurting them both) and causes the correlation between these two, one would expect and find that when these divergences are greater the dependency between the two memory tasks is also greater.[1]

Conclusion

The holographic model of human episodic memory presented here, though obviously oversimplified, produces results that correspond well to the experimental data in the recognition-failure paradigm. Despite large changes in the levels of recognition and recall that result through various manipulations in the model the simulated data, like the human data, showed dependence corresponding to the Tulving-Wiseman function. This dependence occurred even when the items were a priori unrelated to one another as long as the information underlying both recall and recognition was stored in a single composite trace, and so long as both recognition and recall involved representational retrieval. Under these conditions, not only did the cue in recall retrieve the target item from memory, but it also simultaneously retrieved itself. This cue retrieval in recall is nonoptimal—a distortion or mistake that is produced unavoidably from the composite method of storage. Similarly, the probe in recognition retrieved not only the appropriate target term from memory, but also the item that had been associated with it. Interestingly, the superimposed retrieval of the nontarget half of the association in both recognition and recall accounted not only for the generally found Tulving-Wiseman recognition-failure function but also for the exceptions to it.

The results of this research are most surprising and exciting because they stem directly from constructs in the distributed model that are at its core. As a psychologist interested in understanding the workings of human memory, one might ask, with some justification, why one should bother with all the complicated micro mechanisms of neural models. Do they generate new predictions, elucidate poorly understood phenomena, or lead to a new way of thinking about human memory? Although the data are not all in, of course, these distributed models do seem to have consequences for our understanding of human memory. In contrast to symbolic models and computer metaphors that view memory representations as static, in distributed models the representations themselves are continually changing and altering one another. These models, then, allow us an analytical method for thinking about human cognition as dynamic at the level of the representations themselves. They allow us not only to ask the traditional questions of whether a certain item is recalled and how quickly but also to look for transformational changes in content. We are led to ask what is retrieved? Distortions and alterations in the content of memory events can be explored by comparing model predictions with

---

[1] A reviewer recommended that the model be fitted to the data. This was not done for two reasons. First, it is not clear whether the appropriate data are the raw data, as shown in Figure 1 of this article and as conform reasonably well to the Tulving-Wiseman equation, or the homogenized data (as shown in the right panel of Figure 5). The method of homogenization purges item- and subject-selection effects. This seems appropriate because neither of these factors are of theoretical interest. If the homogenized data are the appropriate comparison, then a model should show consistent dependency, but slightly less dependency than the Tulving-Wiseman equation predicts. The simulation results that were produced (when the number of features used in the model was 63 or 35) showed consistent dependency of magnitude a hair less than that given by the Tulving-Wiseman equation. So if one were willing to argue that the corrected data are the "correct" data one could claim that the model fits well. Even so, as was shown, the degree of dependency varies as a function of the number of features in the vectors. Thus, even if the initial simulations (in which this number was chosen arbitrarily) did not fit, it would be trivial, and tedious, to redo the simulations with a better chosen feature number. The contribution of the model is not in its precise fit to the data, which in this context is meaningless, but rather in the invariance over a wide variety of conditions in which invariance is produced, the overdependence under conditions in which overdependence is produced, the explanation of the causes underlying such results, and the new predictions that it provides.

human data. These distortions sometimes appear in strange garb—here, in the recognition-failure function and its exceptions.

Van Gelder (1989) has noted that since the late 1970s,

> the broad concept of distributed representation has been granted an important place in the developing discipline of cognitive science. It forms a landmark around which there clusters a variety of issues and questions, matters which can only be ignored to the general detriment of one's understanding of the whole area. Yet surprisingly, there has never been any attempt to understand in a comprehensive way the nature and properties of distribution. (p. 7)

When it comes to understanding the implications of the properties of distributed representations as they might play themselves out in human data or task dependencies, we may sometimes be surprised. At the outset of this research, I did not think that the model would produce the recognition-failure function. I was agnostic about the reasons for the function but certainly did not think it had much or anything to do with composite distributed-memory storage. When simulations showed that the holographic model did produce the function, I thought that all mechanistic models would do so (and so was surprised to find that they did not). Further analysis of the model revealed that the distributed properties were in fact important in producing the relation, but this was not obvious at first, nor were these properties in any sense designed to have this effect. The result was unexpected. It seems likely that the "understanding in a comprehensive way" that Van Gelder sought, particularly when referring to properties that may have implications for (and, it is hoped, enhance) the understanding of human cognition, is liable to be an extended adventure.

References

Anderson, J. A. (1970). Two models for memory organization using interactive traces. *Mathematical Biosciences, 8,* 137–160.

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.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV.* San Diego, CA: Academic Press.

Bahrick, H. P. (1979). Two-phase model for prompted recall. *Psychological Review, 77,* 215–22.

Begg, I. (1979). Trace loss and the recognition failure of recallable words. *Memory & Cognition, 7,* 113–123.

Bjork, R. A., & Bjork, E. L. (in press). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes.* (Vol. 2), Hillsdale, NJ: Erlbaum.

Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior, 9,* 529–533.

Bower, G. H. (1972). Stimulus sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in learning and memory* (pp. 85–123). New York: Wiley.

Bowyer, P. A., & Humphreys, M. S. (1979). Effect of a recognition test on a subsequent recall test. *Journal of Experimental of Psychology: Human Learning and Memory, 5,* 348–359.

Bryant, D. J. (in press). Exceptions to recognition failure as a function of the encoded association between cue and target. *Memory & Cognition.*

Carey, S. T., & Lockhart, R. S. (1973). Encoding differences in recognition and recall. *Memory & Cognition, 1,* 297–300.

Donnelly, R. (1988). Priming effects in successive episodic tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 256–265.

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

Flexser, A. J. (1981). Homogenizing the 2 × 2 contingency table: A method for removing dependencies due to subject and item differences. *Psychological Review, 85,* 153–171.

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

Gabor, D. (1948). A new microscopic principle. *Nature, 161,* 777–778.

Gardiner, J. M. (1988). Recognition failure and free-recall failures: Implications for the relation between recall and recognition. *Memory & Cognition, 16,* 446–451.

Gardiner, J. M. (1989). The Tulving-Wiseman law and exceptions to it. In A. F. Bennett & K. M. McConkey (Eds.), *Cognition in individual and social contexts* (pp. 253–260). Amsterdam: North-Holland.

Gardiner, J. M., Kaminska, Z., Java, R. I., Clarke, E. F., & Mayer, P. (1990). The Tulving-Wiseman law and the recognition of recallable music. *Memory & Cognition, 18,* 632–637.

Gardiner, J. M., & Nilsson, L.-G. (1990). *Relation between recognition and recall: The Tulving-Wiseman law.* Manuscript submitted for publication.

Gardiner, J. M., & Tulving, E. (1980). Exceptions to recognition failure of recallable words. *Journal of Verbal Learning and Verbal Behavior, 19,* 194–209.

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

Hintzman, D. L. (1987). Recognition and recall in MINERVA 2: Analysis of the "recognition-failure" paradigm. In P. Morris (Ed.), *Modelling cognition* (pp. 215–229). New York: Wiley.

Horton, D. L., Pavlick, T. J., & Julian, M. W. (1989a, November). *More about the relationship between recognition and recall.* Paper presented at the Annual Meeting of the Psychonomic Society, Atlanta, GA.

Horton, D. L., Pavlick, T. J., & Julian, M. W. (1989b, June). *The relationship between recall and recognition.* Paper presented at the Annual Meeting of the American Psychological Society, Alexandria, VA.

Jones, G. V. (1978). Recognition failure and dual mechanisms in recall. *Psychological Review, 85,* 464–469.

Jones, G. V. (1983). Structure of the recall process. In D. E. Broadbent (Ed.), *Functional aspects of human memory* (pp. 135–147). London: Royal Society.

Jones, G. V., & Gardiner, J. M. (1990). Recognition failure when recognition targets and recall cues are identical. *Bulletin of the Psychonomic Society, 28,* 105–108.

Kintsch, W. (1978). More on recognition failure of recallable words: Implications for generation-recognition models. *Psychological Review, 85,* 270–281.

Kohonen, T., Oja, E., & Lehtio, P. (1981). Storage and processing of information in distributed associative memory systems. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 105–143). Hillsdale, NJ: Erlbaum.

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

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

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review, 75,* 421–441.

Martin, E. (1975). Generation-recognition theory and the encoding specificity principle. *Psychological Review, 82,* 150–153.
McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing* (Vol. 1). Cambridge, MA: MIT Press.
Metcalfe, J. (1990). Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General, 119,* 145–160.
Metcalfe, J., & Bjork, R. A. (1991). Composite models never (well, hardly ever) compromise: Reply to Schooler and Tanaka (1991). *Journal of Experimental Psychology: General, 120,* 203–210.
Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 20,* 161–189.
Metcalfe Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.
Metcalfe Eich, J. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review, 92,* 1–38.
Murdock, B. B. (1974). *Human memory: Theory and data.* 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. (1985). Convolution and matrix systems: A reply to Pike. *Psychological Review, 92,* 130–132.
Murdock, B. B. (1989). Learning in a distributed memory model. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium in cognition* (pp. 69–105). Hillsdale, Erlbaum.
Muter, P. (1978). Recognition failure of recallable words in semantic memory. *Memory & Cognition, 6,* 9–12.
Muter, P. (1984). Recognition and recall of words with a single meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 198–202.
Neely, J. H., & Payne, D. G. (1983). A direct comparison of recognition failure rates for recallable names in episodic and semantic memory tests. *Memory & Cognition, 15,* 389–396.
Nilsson, L.-G., Dinniwell, M., & Tulving, E. (1987). Recognition failure of categorized words. *Memory & Cognition, 15,* 389–396.
Nilsson, L.-G., Law, J., & Tulving, E. (1988). Recognition failure of recallable unique names: Evidence for an empirical law of memory and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 266–277.
Nilsson, L.-G., & Shaps, L. P. (1980). A functional view of memory. In F. Klix & J. Hoffman (Eds.), *Cognition and memory: Interdisciplinary research of human memory activities,* (pp. 40–46). Berlin, Federal Republic of Germany: Deutscher Verlag der Wissenschaften.
Nilsson, L.-G., & Shaps, L. P. (1981). A reconstructive processing interpretation of the recognition failure phenomenon. *Acta Psychologica, 47,* 25–37.
Postman, L. (1975). Tests of the generality of the principle of encoding specificity. *Memory & Cognition, 3,* 663–672.
Rabinowitz, J. C. (1984). Aging and recognition failure. *Journal of Gerontology, 39,* 65–71.
Rabinowitz, J. C., Mandler, G., & Barselou, L. W. (1977). Recognition failure: Another case of retrieval failure. *Journal of Verbal Learning and Verbal Behavior, 16,* 739–663.
Reder, L. M., Anderson, J. R., & Bjork, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology, 102,* 648–656.
Restle, F. (1965). Significance of all-or-none learning. *Psychological Bulletin, 64,* 313–325.
Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing* (Vol. 2). Cambridge, MA: MIT Press.
Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.
Tulving, E., & Watkins, O. C. (1977). Recognition failure of words with a single meaning. *Memory & Cognition, 5,* 513–522.
Tversky, A., & Kahneman, D. (1982). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 3–22). Cambridge, England: Cambridge University Press.
Van Gelder, T. (1989). *Distributed representation.* Unpublished doctoral dissertation, University of Pittsburgh.
Vining, S. K., & Nelson, T. O. (1979). Some constraints on the generality and interpretation of the recognition failure of recallable words. *American Journal of Psychology, 92,* 257–276.
Wallace, W. P. (1978). Recognition failure of recallable words and recognizable words. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 441–452.
Wicker, F. W. (1970). On the locus of picture–word differences in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 9,* 52–57.
Wiseman, S., & Tulving, E. (1975). A test of confusion theory of encoding specificity. *Journal of Verbal Learning and Verbal Behavior, 14,* 370–381.
Wiseman, S., & Tulving, E. (1976). Encoding specificity: Relation between recall superiority and recognition failure. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 349–361.

Received March 21, 1990
Revision received December 27, 1990
Accepted March 22, 1991 ■

Zahn-Waxler Appointed New Editor, 1993–1998

The Publications and Communications Board of the American Psychological Association announces the appointment of Carolyn Zahn-Waxler as editor of *Developmental Psychology.* Zahn-Waxler is associated with the National Institute of Mental Health. As of January 1, 1992, manuscripts should be directed to

Carolyn Zahn-Waxler
4305 Dresden Street
Kensington, Maryland 20895

Manuscript submission patterns make the precise date of completion of the 1992 volume uncertain. The current editor will receive and consider manuscripts through December 1991. Should the 1992 volume be completed before that date, manuscripts will be redirected to the incoming editor for consideration in the 1993 volume.