

Recognition of multiple-item probes

STEVEN E. CLARK and RICHARD M. SHIFFRIN
Indiana University, Bloomington, Indiana

In the present study, we contrasted models in which *familiarity* or *strength* is the sole basis for recognition judgments (e.g., Gillund & Shiffrin, 1984) with models incorporating retrieval of specific information in a recall-like process (e.g., Humphreys, 1976, 1978; Humphreys & Bain, 1984). We also examined the possibility that an item's "strength" is determined in part by the match between the verbal encoding contexts at study and at test. In two experiments, study items were presented in triplets. In Experiment 1 recognition was tested with all possible combinations of one-, two-, or three-item targets and distractors; in addition, three different decision criteria were employed in different lists. Experiment 2 included cued recognition among other test conditions. Recognition performance did not increase as more cues were added to the probe, a finding that is inconsistent with Humphreys's model and other models of that type. Both studies were well fit by a version of the Search of Associative Memory model for recognition (Gillund & Shiffrin, 1984).

Most theories explain recognition memory in terms of the strength or familiarity of the test probe, where familiarity is the result of a fast parallel comparison of the test probe to memory. The necessity for such a fast matching process is suggested by subjects' ability to reject some distractors quickly and confidently (Atkinson & Juola, 1974; Fischler & Juola, 1971; Herrmann, Frisina, & Conti, 1978; Murdock, 1974; Murdock & Anderson, 1975; see also Glucksberg & McCloskey, 1981). A negative response that must await termination of serial search for the test item would certainly take too long.

Partly for these reasons, essentially all current recognition models include a familiarity process. However, we might ask whether such a process is sufficient. Humphreys (1976, 1978; Humphreys & Bain, 1984) argued that familiarity alone is insufficient to predict data from studies in which items are presented and tested in groups.

Consider first a simple familiarity model in which the familiarity of an item tested as part of a group is independent of the familiarity of the other items in the group. Such a model would have difficulty with the oft-replicated finding that recognition is improved when words are tested in the presence of other words with which they were studied. For example, assume two words (AB) are studied and followed by one of two kinds of test probe: an *intact* pair (AB), which consists of two words that were studied together, or a *rearranged* pair (AB'), which consists of two words from the list but from different pairs (Humphreys, 1976, 1978). If decisions were based solely on the independent familiarity of each item, discrimination would be impossible since both items in both test cases are equally familiar by virtue of having been studied. Although discriminating between intact and rearranged

probes is difficult compared with other kinds of recognition tests, performance is nonetheless reliably above chance. In addition, if subjects are asked to distinguish AB and AB' probes from probes that contain nonlist items, hit rates are higher for the intact AB probes than for the rearranged AB' probes. We shall refer to these two kinds of results collectively as the *intact advantage*.

Clearly, a model that bases recognition on the independent familiarity of individual items within the multiple-item probe cannot account for such findings. Three approaches to recognition have evolved to deal with this problem.

Probing Memory with Joint Cues

One approach maintains the familiarity component, but relaxes the independence assumption. In this approach, the different cues, including the test items themselves, are assembled *interactively* into a single joint cue, and that single cue is used to probe memory. Various models make this assumption, and, although they differ conceptually, they all combine test items to probe memory in such a way as to emphasize the connections between paired items. Matrix models (Anderson, Silverstein, Ritz, & Jones, 1977; Pike, 1984), Hintzman's (1984) vector model MINERVA, convolution-correlation models (Murdock, 1982), and the Search of Associative Memory (SAM) model (Gillund & Shiffrin, 1984) are examples. Although conceptually different, these models are all capable of predicting the intact advantage.

The Change-of-Meaning Hypothesis

The second approach also retains the notion of familiarity judgments and in a sense allows the familiarity of different "items" to be independent. However, the functional items that are used to probe memory are not necessarily assumed to match the nominal items that are presented for test. Such an approach assumes that a test

This research was supported by NIMH grant 12717 and an Indiana University Waterman research grant to Richard Shiffrin. Reprint requests should be directed to Richard M. Shiffrin, Department of Psychology, Indiana University, Bloomington, IN 47405. Steven Clark is now at University of California, Riverside.

item is first encoded in semantic memory, and that that particular encoding is used to probe episodic memory. In this view, the encoding process, not the actual computation of familiarity, is affected by other items in the test group. For example, the *ham* of *radio-ham* is encoded as a functionally different word than the *ham* of *baked-ham*. Thus the functional words that are used as probes to the familiarity process are different. Such an argument was put forth by Light and Carter-Sobell (1970).

Each of the first two proposals is related to Tulving and Thomson's (1973) *encoding specificity principle*, which assumes that "what is stored about the occurrence of a word . . . is information about the specific encoding of that word in that context in that situation" (p. 359). Only retrieval cues that utilize or retrieve that particular context will be effective. The change-of-meaning hypothesis can be thought of as a special case of encoding specificity. Meaning is only one dimension of context that must match.

Familiarity Plus Search

The third approach assumes that familiarity judgments are supplemented by some additional mechanism, perhaps the recall of specific information about a particular stored item. The models of Atkinson and Juola (1973), Humphreys (1976, 1978), and Mandler (1980) are of this type. Mandler and Humphreys, in particular, have applied this approach to multiple-item testing. Both models propose a recall-like process but differ concerning the conditions under which that process operates (see Humphreys & Bain, 1984, for a comparison of the two models). For our present purposes, we focus on the model by Humphreys.

In Humphreys's model verbal context mediates the intact advantage, not by influencing the encoding of words at study and test, but instead by providing an additional retrieval cue. When AB pairs are presented, two distinct kinds of information are stored: *item* information indicates an item's previous occurrence, and *relational* information indicates the co-occurrence of items. Thus, item information is stored for both A and B, and relational information is stored for AB. Relational information provides an additional retrieval cue only when intact pairs are tested. For example, a subject can recognize the B item by retrieving item information for B or by retrieving item information for A and using A as a recall cue to retrieve AB relational information. However, if B is tested alone, this alternative retrieval route is not available.

These three proposals alter the independent strength model in different ways: The joint-cue model modifies the computation of familiarity, so that the familiarity of a given item is based on an interactive combination of all items in the probe. The change-of-meaning hypothesis does not assume modification of familiarity computation, but instead assumes that the encoding of items at study and at test is modified by verbal context. The familiarity-plus-recall model supplements the familiarity computation with a recall-like component.

The literature that bears on these issues is not very helpful in distinguishing the different model types. Most studies have presented and tested at most two-item groups (e.g., Humphreys, 1976, 1978; Light & Carter-Sobell, 1970; Mandler, Rabinowitz, & Simon, 1981; Thomson, 1972) and have used only a few test types, enabling only a few measures of sensitivity in each study. Because the instantiations of the model types we have mentioned generally have more parameters than the studies have sensitivity measures, reliable tests are difficult to find. Our first study, therefore, was designed to provide a much richer data base. Three-item word groups were studied, and recognition was tested with all possible combinations of old and distractor items in groups of one, two, and three words. In addition, three different decision rules were used for different lists.

EXPERIMENT 1

Method

Subjects. The subjects were 128 introductory psychology students participating as partial fulfillment of a course requirement.

Procedure. Words on each study list were presented in groups of three (triplets) and tested in triplets, pairs, and single items, in all possible probe combinations, creating 13 possible test types shown in Table 1. The table is to be read as follows: Let ABC represent three words presented together on the study list. For test probes, prime symbols indicate that test items are from different triplets. X, Y, and Z represent new items not studied. Note that these designations represent types, not actual instances; no item appeared more than once on any test list. Thus ABC represents an intact triplet exactly as presented on the study list; ABC' represents two items from the same triplet with the third item, C', from a different triplet. The AB'C" probe consisted of three items from different triplets. In none of the test probes was the ordering of items changed. For example, an ABC study triplet would not be tested with a BCA probe.

The subjects made "old"-"new" decisions on each test trial. Clearly, there is more than one way to dichotomize these 13 test types into old and new categories. For each test list, the subjects were instructed to use one of three decision rules. In the *intact* decision rule, the subject was instructed to say "old" only if all words in the test probe were studied together. These are the ABC, AB, and A probes as shown in Table 1. In the *all-old* decision rule, the subject was to say "old" when all words in the probe were from the study list but were not studied together. In the *any-old* rule,

Table 1
Materials for Experiment 1

Study List (ABC)			Test List			
DOOR	RADIO	WALL	DOOR	RADIO	WALL	(ABC)
HORSE	BOOK	PHONE	DOOR	RADIO	PHONE	(ABC')
TABLE	PENCIL	SHELF	DOOR	BOOK	SHELF	(AB'C'')
			DOOR	RADIO	COAT	(ABX)
			DOOR	BOOK	COAT	(AB'X)
			DOOR	KNIFE	COAT	(AXY)
			APPLE	KNIFE	COAT	(XYZ)
			DOOR	RADIO		(AB)
			DOOR	BOOK		(AB')
			DOOR	COAT		(AX)
			KNIFE	COAT		(XY)
			DOOR			(A)
			COAT			(X)

Note—Three of 21 studied triplets and 13 possible test type examples from those three studied triplets.

the subject was to say "old" if any word in the test probe was from the list, even if other words in the probe were new. The three decision rules crossed with 13 test types combined for a total of 39 test conditions.

For each test list to consist of roughly half old and half new trials, the different test types were tested with different frequencies both within and across test lists. In the intact decision rule, only the ABC, AB, and A tests were old; these test probes were presented three times each, and the other 10 test types were presented once each, for a total of 19 test trials. The situation was the opposite for the any-old decision rule. Only the XYZ, XY, and X tests were new, so they were presented three times each, and the other 10 types were tested once, again for a total of 19 trials. For the all-old rule, 6 of the test types were old and 7 new, so each type was tested once; an additional six test types were made by randomly repeating three old and three new test types, again for a total of 19 test trials.

Six study lists were presented, each consisting of 21 triplets, presented for 5 sec each. Each list was followed by 30 sec of mental arithmetic, after which the subjects were told which of the three decision rules to use throughout that test phase only. A prompt remained visible throughout the test phase to remind them of the rule. The subjects did not know which decision rule they would use at the time they were studying the list, so presumably all lists were studied in the same way.

The subjects were told that all items would be presented in triplets and to code them together interactively. They were told that each test phase would consist of several different test probes and were informed in detail with examples as to the nature of the different test types and decision rules. The subjects were run in groups of up to 4, and the experiment took about 45 min.

The selection of words for study and test lists, the order in which words were presented, and the sequence in which decision rules were used were all randomized differently for each subject. The sequence in which decision rules were used was random within two blocks of three test phases.

Materials and Apparatus. All words were high-frequency nouns from Kučera and Francis's (1967) and Thorndike and Lorge's (1944) norms. Each noun had a rating of at least 50 occurrences per million in both norms. Stimulus presentation and response collection were controlled by a DEC PDP-11/34 computer.

Results

The probability of responding "old" for each test type is given in Figures 1, 2, and 3 for the intact, all-old, and any-old decision rules, respectively.¹ Error bars indicate 99% confidence intervals around each data point. Points to the left of the vertical line in each panel indicate old test trials; points to the right indicate new test trials. If performance were perfect, points to the left would all equal 1.0 and points to the right would all be zero, making the difference between points crossing the vertical 1.0. Although the actual function is more gradual, for all decision conditions and numbers of test items, the largest difference is that which crosses the vertical.

Three two-way analyses of variance (ANOVAs) were performed for one-, two-, and three-item tests, each crossing test type with decision condition. For tests of two or three items, main effects of decision rule and test type and the interaction were significant ($p < .0001$) in all cases. For single-item tests, however, the main effect for test type and the interaction were significant, but not the main effect for decision condition [$F(2,254) = 1.29, p > .19$]. One would not expect the decision rule to make

a difference, since for single items all decision rules are equivalent. The reason for the interaction is not clear. Aside from this minor anomaly, however, the data are quite systematic.

Four additional tests were performed to evaluate the intact advantage in the all-old and any-old decision conditions. Separate tests were performed for the ABC, ABC', and AB'C" tests and for the AB and AB' tests. Both tests were significant in the all-old condition [$F(2,381) = 4.61, t(254) = 2.30, ps < .025$], but not in the any-old condition [$F(2,381) = 1.14, p > .20; t(254) = 1.67, p > .15$].

Discussion

Detailed discussion of the data is deferred until after presentation of relevant theories; however, certain aspects of the data are worth noting.

First, the subjects were quite good at discriminating intact from rearranged test probes. This discrimination was better for three-item than for two-item probes. $ABC - AB'C = 0.34$ and $ABC - AB'C'' = 0.43$ are both greater than $AB - AB' = 0.25$. Also, the three-item data suggest that the intact advantage drops quickly as other list items are mixed into the test probe. The difference between ABC and ABC' is three times larger than that between ABC' and AB'C".

Also, intactness facilitated recognition when the task did not require an intact-rearranged distinction. This trend was shown in both the all-old and the any-old decision rules, but was not significant in the any-old rule. As will be shown later, these data are important in distinguishing between models.

THE SAM MODEL

To compare different approaches to multiple-item recognition, it can be useful to work within a common theoretical framework. We therefore begin by applying the recognition theory proposed by Gillund and Shiffrin (1984). We present below a version of the model that fit the data adequately and then discuss variants of the model that did not.

It is assumed that each studied item is represented in memory as an image. When test items are presented, the subject probes memory with cues consisting of *context* and the *test items* themselves. These cues produce activation of the images in memory. The sum of the activations of all of the images represents *familiarity*. The familiarity is compared to a criterion value in order to make an "old"- "new" decision. This system is represented quantitatively as:

$$F(Q_1, Q_2, Q_3, \dots, Q_M) = \sum_{j=1}^N \prod_{i=1}^M S(Q_i, I_j)^{w_i} \quad (1)$$

The term on the left represents the familiarity arising when cues $Q_1, Q_2 \dots Q_M$ are used to probe memory. Normally one of these cues is context and the other $M - 1$ cues

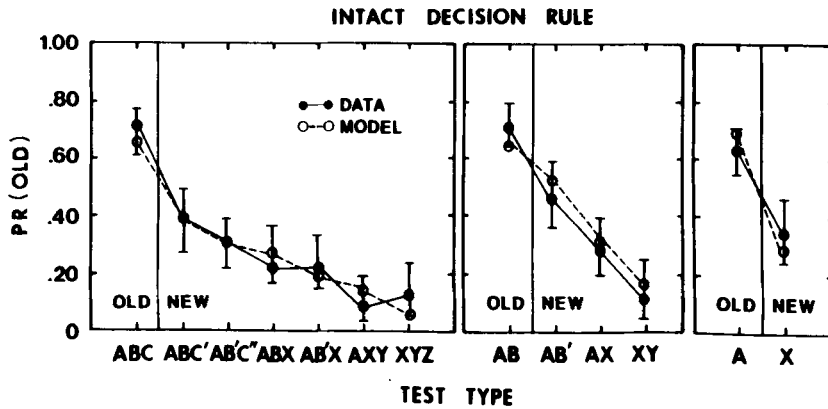


Figure 1. Mean probability of responding "old" for each test type for the intact condition. Data are represented by solid lines, and fit of the SAM model is represented by dotted lines. Error bars indicate 99% confidence intervals about each mean. The parameters used to obtain the fit are as follows: $b = .11$, $c = .098$, $d = .029$, $m_1 = 1.18$, and $m_2 = 1.66$. The criteria were 2.996, 2.972, and 2.972 for three-, two-, and single-item tests, respectively.

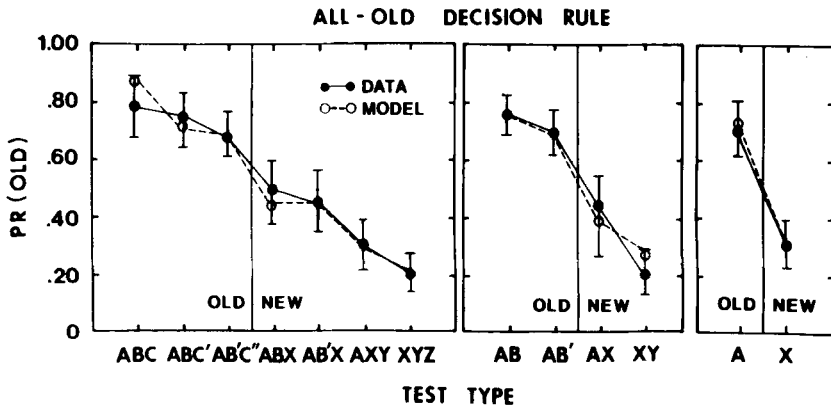


Figure 2. Mean probability of responding "old" for each test type for the all-old condition. The criteria were 2.90, 2.916, and 2.956 for three-, two-, and single-item tests, respectively.

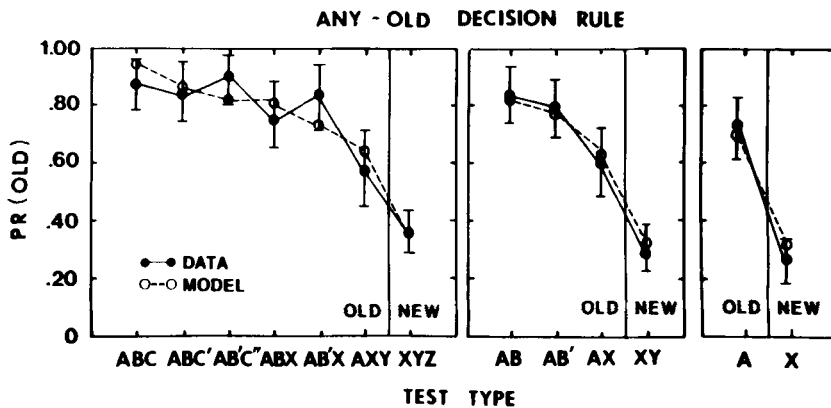


Figure 3. Mean probability of responding "old" for each test type for the any-old condition. The criteria were 3.004, 2.988, and 2.964 for three-, two-, and single-item tests, respectively.

are the items tested. The expression $S(Q_i, I_j)$ represents the *retrieval strength* between the cue Q_i and a particular image I_j . This strength will be higher the longer Q_i was rehearsed with I_j during study of the list. The exponent w_i is an attention weight that may take on a value between 0 and 1 and represents the attention given to Q_j . It is assumed that retrieval has a limited capacity for cue utilization, quantitatively represented by:

$$\sum_{i=1}^M w_i = 1.0 \quad (2)$$

The product of the M weighted retrieval strengths gives the activation for a given image; the sum of these gives the familiarity value.

To use this system, one must understand the rules governing the retrieval strengths, and the weights. The values of the average retrieval strengths vary systematically as follows: The strength between the context cue and an image is given by a , the *context strength*. The strength between a test item cue and an image depends on rehearsal during study. For test items from the list, the cue will be connected to its own image in memory with a *self strength*, c . The cue will be connected to images with which it was rehearsed with an *interitem strength*, b . Finally, an item cue will have some *residual connection* to images with which it was not rehearsed, given by d .

The distributions of these strength values are approximated with a three-point distribution. Given an expected strength value of x , the actual value may be x , $1.5x$, or $.5x$, each with equal probability. (Other distributions with other variance assumptions work about as well.)

Finally, we assume that the encoding of a cue (before it is used to probe episodic memory) is affected by the other items with which it is presented. To the extent that the coding of an item at test matches its study encoding, the strength between the test item and its own stored image as well as the strength between the test item and the image of another item within the same study triplet ought to be higher. These assumptions follow closely those suggested by Gillund and Shiffrin (1984; although this suggestion was not implemented in the simulations in that paper).

In the present study, we assume that the encoding of a test item A will more closely match its encoding at study if other test items were studied with A . We assume that the values of b and c for a given test item A are increased by a match value m_1 when one other item in the test group was studied with A , and by a value of m_2 if two items in the test group were studied with A . To be precise, let $b_0, c_0, b_1, b_2, c_1,$ and c_2 be rehearsal strengths, where the subscript refers to the number of other test items that were in the same study triad as the item in question. Then we let $b_0 = b, c_0 = c, b_1 = m_1b, c_1 = m_1c, b_2 = m_2b, c_2 = m_2c$, where m_1 and m_2 are match parameters such that $m_2 > m_1 > 1.0$.

For completion of the model, the subjects' decision strategies in the various decision conditions must be specified. Decisions in the intact condition concern an aspect of the probe cue as a whole; thus we assume that test items are combined into a single joint probe of memory. If the familiarity of this joint probe exceeds criterion, an "old" response is made. Decisions in the all-old and any-old rule, however, require sensitivity to individual words; in these cases we assume each item is used individually to probe memory. In the all-old rule, if the familiarity of *each* individual item exceeds criterion, the probe is called "old." In the any-old rule, if the familiarity of *any* of the items exceeds criterion, the probe is called "old."

For all tests, the subject probes with context and item cue(s). Context is always given a weight of 0.5, and the remaining 0.5 is evenly distributed among the item cue(s). This preferential weighting for the context cue is assumed in order to focus retrieval and restrict the memory set to only the items presented on the list.

For each of the three decision conditions and for each of the three sizes of test groups, the subject is assumed to choose a separate criterion. Hence, the model has nine parameters representing the various criteria, and also the parameters $b, c, d, m_1,$ and m_2 to be estimated.² The predictions are not affected by the value of a (as noted previously by Gillund & Shiffrin, 1984).

Predictions of the model can be obtained either by Monte Carlo simulation (as in Gillund & Shiffrin, 1984) or by analytic derivations of the means and variances of the familiarity distributions, followed by calculations of $pr(\text{old})$ for each test type. Simulation proved to be unduly time-consuming for the purposes of model fitting; thus, predictions were obtained analytically. However, a positive correlation between familiarity values for individual items in the all-old and any-old decision conditions prevented complete analytic solutions. Instead a close approximation was constructed. Parameters from the best fit from this analytic approximation (based on least squares) were then used to obtain a fit by simulation. The predictions from this fit of the model, which is close to a best fit, are shown for the intact, all-old, and any-old rules in Figures 1, 2, and 3, respectively.

Discussion of the Model

Three factors that would produce an intact advantage were discussed in the introduction. The present model incorporates two of these. First, the multiplicative rule used to calculate the joint familiarity of several cues results in the prediction of greater familiarity for intact probes than for rearranged probes in the intact decision condition. This applies only when the intact decision rule is being used, since items in the all-old and any-old cases are judged individually rather than jointly. Also, encoding-match assumptions predict a larger intact advantage for the intact rule, and alone account for the intact advantage in the all-old decision rule.

For record-keeping purposes, we will denote the above model as Model A. Four additional models (B-E) failed to fit the data. Each is described and the reasons why each fails are given in Table 2.

To determine whether the multiplicative combination of cues alone could handle the results, Model B held $m_1 = m_2 = 1.0$. Single-item probes were used for the all-old and any-old rules, as in Model A. Model B underestimates the intact advantage in the intact rule. Also, without match parameters, single-item probing cannot predict any intact advantage in the all-old decision rule.

Model C replaced match parameters with the assumption that subjects mix probe strategies. To predict intact advantages in the all-old and any-old conditions, we assumed that subjects probed memory both with a joint cue and with individual probes of each test item. For the intact and all-old cases, we assumed that all decisions from both joint and single-item probes had to exceed criterion to lead to an "old" response. Thus the decision strategies are identical in these two cases, differing only in the relative placement of joint and single-item criteria. For the any-old case, either the joint probe or one of the single-item probes had to exceed its respective criterion. Model C predicts the intact advantage in the all-old rule, which Model B could not predict, but still failed to produce a large enough intact advantage in the intact decision condition.

The second factor in the model that produces an intact advantage is the matching of encoding at study and test (instantiated by values of m_1 and m_2 greater than 1.0). Models D and E have this feature. Model D probes only with individual item cues, and Model E probes only with joint cues. Both fail in a similar manner, although for opposite reasons. Because single-item probing does not produce an intact advantage, match parameters must be

set high in Model D to predict the intact advantage in the intact rule. However, these match parameters overpredict the intact advantage in the other two decision conditions. In Model E, on the other hand, joint probing produces a sizable intact advantage for all conditions. Thus, match parameters must be set to low values so as not to overpredict the all-old and any-old data. These parameter values underpredict the intact advantage in the intact decision condition. In both models the large intact advantage in the intact decision condition and the small intact advantage in the other conditions cannot be predicted simultaneously using the same probe strategy.

Thus, the best fit was given by the model with context match parameters somewhat greater than 1.0 and with joint probing used for the intact decision rule and individual probing used for the all-old and any-old conditions.

In summary, the results of the model explorations within the SAM framework suggest that, in the absence of some sort of recall process, a familiarity model of recognition requires an extra process, such as encoding matching (or some other process that would improve the model in similar ways). In light of this conclusion, let us consider a variety of objections to context-match assumptions. In particular, Humphreys (1976, 1978; Humphreys & Bain, 1984) argued that although such models may account for the intact advantage, they are incorrect on other grounds.

First, Humphreys questioned the relationship between meaning and recognition. Light and Carter-Sobell (1970, Experiment 3) presented homographic noun-adjective pairs, such as *baked-ham*, and tested recognition of the noun with the same adjective (intact) or with a new adjective that either matched the same meaning (SM) of the noun (i.e., *sliced-ham*) or biased a different meaning (DM) of the noun (i.e., *radio-ham*). Recognition proba-

Table 2
Summary of Variations of SAM Model Applied to Experiment 1

Model	Memory Probe	Context Match	Outcome
A	Joint probe for intact rule. Single probe for all-old and any-old rules.	Yes	Fit the data.
B	Joint probe for intact rule. Single probe for all-old and any-old rules.	No	Underpredicts intact advantage for intact rule. Predicts no intact advantage for all-old rule.
C	Joint-plus-single probe for all decision con- ditions.	No	Underpredicts intact advantage for intact rule.
D	Single probes for all decision rules.	Yes	Cannot predict magnitude of intact advantage for intact and all-old rules simultaneously.
E	Joint probes for all decision rules.	Yes	Cannot predict magnitude of intact advantage for intact and all-old rules simultaneously.

Note—Only Model A yielded an adequate fit.

bilities were ordered as follows: intact > SM > DM. However, the intact – SM difference was larger than the SM – DM difference, which Humphreys argued is contrary to the change-of-meaning hypothesis: If encoded meaning is the important factor, then recognition should be reduced very little when meaning is preserved and should be reduced more when meaning is changed between study and test. Note, however, that if change of meaning was not operating, then the pattern intact > SM = DM would have obtained, which is also contrary to the data. The observed data therefore seems most compatible with a model in which performance is determined by two components: some performance decrement is due to any change from the intact pair, and an additional decrement is due to the change in meaning. These two factors are given particular instantiations in the encoding-match version of SAM; when applied to the Light and Carter-Sobell study, their intact – SM difference would be accounted for by the multiplicative combination of cues, and their intact – DM difference would be due to this plus encoding mismatch.

Underwood and Humphreys (1979) presented single nouns for study that were tested alone, or with a preceding adjective, or in a sentence. They argued that such changes should “without doubt” change the encoded meaning of the word; nevertheless, recognition was unaffected. However, Winograd and Conn (1971) explained a similar result by assuming that when a word is in isolation, the dominant meaning of the word is encoded; when the test context also selects the dominant meaning, then a close match obtains.

Humphreys’s second criticism was directed at a fundamental prediction of all versions of encoding specificity. Such proposals account for the intact advantage by increasing the strength of each item independently when study context is reinstated at test. Thus, if subjects are asked to judge each item separately, the probability of recognizing at least one item should be higher when those items form an intact test pair than when they form a rearranged pair (even though all items are old). This can be expressed reciprocally: the double-miss rate should be lower for intact pairs. Humphreys (1976, 1978) showed the two double-miss rates to be equal, despite finding the intact advantage for pair recognition. Although this finding may be inconsistent with the encoding-match position, the strength of the result has been diluted somewhat by subsequent findings. Humphreys and Bain (1985) reported that double-miss rates were equal only when subjects were told to rehearse pair members separately. When items were rehearsed interactively, double-miss rates were lower for intact pairs, as the encoding-match hypothesis predicts.

Aspects of our present data are relevant. In the any-old decision rule, the probability of responding “old” is the probability of recognizing at least one item in the group. To be consistent with Humphreys’s original results, our results would have to show that the $pr(\text{old})$ does not increase due to intactness (i.e., $ABC = ABC'$

$= AB'C''$ and $AB = AB'$). In fact, it was observed that $ABC > ABC'$ and $ABC > AB'C''$ and $AB > AB'$, although these trends did not quite reach statistical significance.

In summary, the SAM model with a context-match component remains a viable contender among accounts of multiple-item memory probing: The model is consistent with Light and Carter-Sobell’s (1970) findings and also with the double-miss rates for interactively encoded pairs found by Humphreys. We do not claim that the context-match process is necessarily of large magnitude, but rather that the SAM model with this factor can account for the present data and is consistent with past findings. It should be noted that the model we have proposed is based entirely on familiarity. We next consider an alternative approach based on the model and hypotheses of Humphreys (1976, 1978; Humphreys & Bain, 1984).

HUMPHREYS’S MODEL

Central to Humphreys’s model (1976, 1978) is a distinction between *item* information and *relational* information. Item information indicates a word’s occurrence on the study list; relational information indicates the co-occurrence of items. The basic idea is that for intact test groups, the subject can recognize an item by retrieving item information (for that item) or, if that fails, by retrieving item information from another item in the test group and then using that item information to retrieve relational information. The relational information would indicate the retrieved item’s co-occurrence with the to-be-recognized item, and thus lead to a correct response. For nonintact probes, this alternative retrieval route is not available, thus leading to an intact advantage. The system is made more complicated by the possibility of a variety of guessing strategies and of various kinds of false recognitions. As a result, when extending Humphreys’s model to three-item study and test groups, many variants are possible within the general framework. We examined a number of these, and describe below a version that handles the data well and seems reasonably consistent with the basic approach.

We assume that for any list item in any test probe, the probability of retrieving item information is a . Similarly, the probability of incorrectly retrieving item information for a nonlist item is e . The probability of retrieving relational information given that item information was retrieved from a list item is α , but it is zero if item information is retrieved from a nonlist item or if no item information is retrieved. In addition, we assume that relational information gives *full* information concerning the co-occurrence of all study list items in a given triad. Therefore, in the intact condition, the retrieval of relational information from any one item is sufficient to recognize intact test groups and to reject nonintact test groups. In the intact and any-old conditions, retrieval of relational information always leads to a correct decision. In the all-old condition, retrieval of relational information leads to

a correct "old" response for ABC probes (and helps in a few other special cases as well).

Next, we must specify strategies and guessing probabilities. In the intact condition, a correct "old" response is made if relational information is retrieved; if relational information is not retrieved, subjects may still correctly respond "old" by retrieving one, two, or three items and guessing with probabilities G_1 , G_2 , and G_3 , respectively. False alarms are made if item information is retrieved from one, two, or three nonlist items; if relational information is not retrieved; and if the subject guesses "old" with probabilities G_1 , G_2 , and G_3 .

In the all-old rule, any combination of relational and item information that results in the retrieval of all test items is assumed to result in an "old" response. If one or more items is not retrieved, the following guessing parameters apply: $G_{i,j}$ is the probability of responding "old" if i items are retrieved (via item or relational information) from a j -item test group, in which $i < j$. There are six of these parameters in all.

For the any-old condition, the subject is assumed to respond "new" if no relational information is retrieved and if all items that are retrieved are guessed to be new. These guessing parameters are $H_{i,j}$: the probability of guessing new if i items are retrieved from a j -item test group. This adds six more parameters.

Given these rules, it is possible to specify the 39 equations giving $\text{pr}(\text{old})$ for each of the test conditions in our study. There are three retrieval parameters and 15 guessing parameters. A best fit of this model proved quite adequate; the fit was as good as and very similar to the one shown in Figure 1.

The results of these modeling projects leave us with two contending approaches to multiple-item recognition. The first, incorporated within the SAM framework, utilizes interactions among cues and context-match assumptions. The second, incorporated in an extension of Humphreys's model, assumes that recognition can take place either by direct judgment of an item or through retrieval of a link between some other item and a test item. Both approaches give adequate fits to the data of Experiment 1. We therefore designed Experiment 2 to discriminate between the two approaches.

The key experimental technique that allows the approaches to be distinguished is borrowed from the cued condition used by Gillund and Shiffrin (1984). They presented word pairs for study, and tested using a *cued recognition* procedure in which a list item was presented as a cue either with its corresponding pair member or with a nonlist item. The cue word was labeled as a list item, and subjects made decisions only about the test word presented with it. Performance on the cued test was compared to that on a single-item condition, which was identical to the cued test except that no cue was used. Familiarity-plus-search models, like Humphreys's, would predict that, at least on some trials, subjects could use the cue item to recall the test item if it was studied with the cue, and reject the test item if it was not studied with

the cue. To the extent that this occurs, cued recognition should be better than single-item recognition. The data from Gillund and Shiffrin, however, show a slight decrease in performance for the cued condition. In SAM this is accounted for by the redistribution of attention *away* from the test item to the cue. Since both targets and distractors are accompanied by cue items, the cues are not diagnostic, unless they can be used to recall the item(s) with which they were originally presented. Because there is no recall process in the SAM model, the attention weight given to the cue is largely wasted; moreover, the weight given to the cue is removed from the tested item, thus leading to poorer performance. In SAM cues do help because of the factors that produce the intact advantage (i.e., the multiplicative cue combination rule, and the context-matching factor proposed in this paper). However, any advantage due to weight given to the cue is considerably outweighed by the *removal* of weight from the test item.

The finding that cued recognition is poorer than single-item recognition appears to provide evidence against a wide class of models that incorporate recall processes. It also provides evidence against context-match processes in models that do not contain a countervailing process such as limited-capacity cue-weighting. It is therefore important to establish the stability and generality of this finding. Experiment 2 extended the cued-recognition procedure that Gillund and Shiffrin (1984) used for pairs to include triplets. Also, Gillund and Shiffrin did not include tests requiring discrimination of intact from rearranged pairs, which may have biased subjects against using recall as part of their recognition process. Thus, additional tests that would require intact-rearranged distinctions were included.

It is important to note that Humphreys's model predicts a noticeable improvement in recognition when cues are provided. Using the parameter values estimated from Experiment 1, predictions for single-item and cued recognition with one or two cues are given in Table 3. We have assumed in the cued conditions that any relational information retrieved will result in a correct decision. The model predicts an advantage for cuing that increases as the number of cues is increased. This is generally true of Humphreys's model and probably of other similar models as well. Thus, our second study should discriminate between the contending models.

EXPERIMENT 2

Experiment 2 used a study-test procedure similar to that of Experiment 1. Each study list consisted of 24 triplets

Table 3
Predictions of Humphreys's Model for Cued Recognition Using Best-Fitting Parameter Values From Experiment 1

	Single	One Cue	Two Cues
Pr(Hit)	0.85	0.87	0.89
Pr(False Alarm)	0.38	0.33	0.29

Note—Values of retrieval parameters are $a = .85$, $e = .38$, $\alpha = .15$.

of words, and each list was followed by one of seven test conditions: single-, paired-, and triple-item recognition; cued recognition with one or two cues; and two intact-rearranged tests, one using two and the other using three test items.

Method

Subjects. Fifty-six introductory psychology students participated to fulfill a course requirement.

Materials. Study and test lists consisted of low-frequency words (fewer than 4 occurrences per million) from Thorndike and Lorge's (1944) and Kučera and Francis's (1967) norms.

Procedure. Seven study lists were presented, each followed by a different kind of test. Study lists consisted of 24 triplets, each presented for 3.25 sec. Immediately following list presentation, the subjects performed mental arithmetic for 30 sec, and then were tested with one of seven testing procedures. The different kinds of tests can be grouped into three categories—old-new tests, cued tests, and intact-rearranged tests—and are described as follows:

Old-new tests. Three kinds of old-new tests were given. One test list contained all single items, another contained all pairs, and another contained all triplets. Using the same notation as in Experiment 1, the single-item tests consisted of old (A) or new (X) items, paired-item tests consisted of two old (AB) or two new (XY) items, and triple-item tests consisted of three old (ABC) or three new (XYZ) items.

Cued tests. Cued tests presented a single test item with one or two cues, and the number of cues was manipulated between lists. The cue or cues were always from the list; when there were two cues, they were from the same study triplet. The test item was either an old item from the same study triplet as the cue(s) (AB or ABC, where the cues are underlined) or a new item not from the list (AX or (ABX). In both tests, cue and test items were randomly chosen. The test item was identified by a row of question marks above it.

Intact-rearranged tests. Intact-rearranged tests consisted of two or three items, manipulated between lists. In the two-item case, the subjects were tested with either two old items from the same study triplet (AB) or two old items from different study triplets (AB'). In the three-item case, subjects were tested with three old items from the same triplet (ABC) or three old items from different triplets (AB'C'').

The selection and assignment of words to test conditions was completely random. Note that designations used to describe the tests indicate types including possible variations. For example, what is designated generically as an AB test is a test of *any* two items from a study triplet, encompassing AB, AC, and BC variations. In all one- and two-item tests, blanks are used to indicate nontested items and indicate the position of tested items. Seven sequences of tests were made so that each test type appeared in each of seven positions during the experiment. Subjects were run in groups of up to 4, and the experiment took about 50 min.

Results

Table 4 shows mean hit and false-alarm rates for the seven test conditions. Hit rate minus false-alarm rate for each test condition was computed for each subject and submitted to an analysis of variance [$F(6,330) = 48.67, p < .0001$]. For old-new tests, performance steadily increased with more test items, although differences were small. In the cued conditions, one cue decreased performance relative to performance for single items, replicating previous results. In addition, performance with two cues was worse than with one cue. However, a compari-

Table 4
Fit of SAM to Experiment 2 Hit Rates (HRs) and False-Alarm Rates (FARs)

	Data		$m_1 = m_2 = 1$		$m_2 > m_1 > 1$	
	HR	FAR	HR	FAR	HR	FAR
Single Item	0.70	0.22	0.72	0.21	0.64	0.26
One Cue	0.69	0.24	0.69	0.24	0.67	0.24
Two Cues	0.67	0.26	0.67	0.26	0.67	0.27
Pairs	0.72	0.15	0.72	0.14	0.67	0.20
Triples	0.77	0.13	0.76	0.12	0.77	0.14
Intact-Rearranged (2)	0.64	0.57	0.67	0.55	0.67	0.54
Intact-Rearranged (3)	0.67	0.46	0.64	0.47	0.74	0.41

Note—Parameters used in the fit of the model with $m_1 = m_2 = 1.0$ were $a = .01, b = .024, c = .016, d = .004, w_1 = .34, w_2 = .28$ for both joint and individual probes for paired- and triple-item tests. Parameters for the model with $m_1 = 1.18$ and $m_2 = 1.66$ were $a = .01, b = .018, c = .008, d = .004, w_1 = .41, w_2 = .39$.

son of the uncued single-item test with the two cued conditions was not statistically significant [$F(1,53) = 2.10, p > .10$]. The main point to note is the failure of the cues to improve performance.

Data Fits and Discussion

The data is consistent with previous work by Gillund and Shiffrin (1984): Cued recognition with one cue was slightly poorer than single-item recognition. Recognition with two cues was poorer yet, extending their result. Although the differences were not statistically significant, it is quite clear that cues do not facilitate performance. This result is inconsistent with Humphreys's (1976, 1978) model and similar models that propose a recall-like process to supplement familiarity in recognition (Atkinson & Juola, 1973; Mandler, 1980). However, this result does not rule out other types of recall-assisted recognition models. If recall is occurring, some other factors must operate to offset the gain due to recall. The models under consideration, except for SAM, contain no such factor. Of course, it is possible to argue that subjects do not use relational information (or do not attempt to retrieve it) when cues are provided. If this were so, the cues would not help. However, it is difficult for us to see why subjects would not use relational information since they are presumed to do so in all other conditions, they are induced to do so by the instructions, and their performance would be greatly enhanced by doing so.

Next we applied the SAM model to the data of Experiment 2. Joint probes were assumed for all test conditions in which multiple-item probes were tested. Clearly, joint probes are necessary for intact-rearranged tests. However, for paired- and triple-item tests, it is not obvious whether to probe with individual items or with the joint probe. The old-new tests in the present study (as contrasted with Experiment 1) are redundant with intactness: *all* old test groups were intact. Since it is not obvious a priori whether to assume that a subject should probe with a joint cue or with individual cues, separate fits were made using each assumption (for the paired- and triple-item tests only). It turns out that this assumption makes very little difference, and only the joint-cue predictions

are given. For the cued tests, we used a version of the assumptions made by Gillund and Shiffrin (1984). A weight of .5 was given to context, and a weight of .5 was divided among the two or three items in the probe including the cue. The weight given to the test item was a parameter w_t ($0 < w_t < 0.5$), and the remaining weight $1 - w_t$ was divided evenly among the cue item(s).

Fits of the model with $m_1 = m_2 = 1.0$ and with context-match parameters set at $m_1 = 1.18$ and $m_2 = 1.66$ (in accordance with the best-fitting values of Experiment 1) are shown in Table 4. Context matching is not necessary to fit the data, and in fact the fit is slightly better when the match parameters are set to 1.0. Both fits, however, are quite good.

It is important to note that the SAM model can, with various parameter choices, predict an advantage for the cued condition. For example, the best fit of the model with $m_1 = 1.18$ and $m_2 = 1.66$ predicts that performance with one cue should be slightly better than the single-item case (see Table 4). As mentioned earlier, both encoding-match assumptions and the multiplicative combination of cues function in the model to give an advantage to cuing. The advantage due to encoding matching operates regardless of the weight given to the cue.³ The advantage due to multiplicative combination of probe items is maximum when the weights given to the cue and test items are equal. On the other hand, removing any weight from the test item (to give to the cue) produces a cue disadvantage that increases as the weight removed from the cue increases. With such a balance of factors, it is not surprising that the addition of matching parameters greater than 1.0 reverses the slight cuing disadvantage that is predicted when match parameters equal 1.0.

Why were match parameters greater than 1.0 needed for Experiment 1 but not for Experiment 2? Two possibilities deserve discussion. First, high-frequency words were used in Experiment 1 and low-frequency words were used in Experiment 2. Given this, the good fit with match parameters set to 1.0 is consistent with the change-of-meaning hypothesis. Encoding-match effects are seen only to the extent that it is possible to encode the same item differently at study and at test. Reder, Anderson, and Bjork (1974) argued that since low-frequency words have fewer meanings, it is less likely that encodings will differ between study and test.

A paradigm termed *recognition of recallable words* has been used to address the issue of encoding invariance and number of meanings. To the degree that recalled items are not recognized, it has been argued that the encoding of test items in the recognition task has changed from the study list encoding. Muter (1984) presented surnames of famous individuals for which most people have only one reference, such as *Kierkegaard* and *Rachmaninoff*. In the 147 cases in which these words were recalled, only twice were they not recognized. Such a result is consistent with the change-of-meaning hypothesis proposed by Light and Carter-Sobell (1970) and strongly suggests that encoding-match effects should be smaller for low-frequency words.

A second hypothesis to explain the lower match parameters for Experiment 2 is based on shorter presentation times in that study. Encoding variability effects require that stored encodings are sufficiently context specific to allow matches and mismatches of encodings at test. It may take more time than was allowed in Experiment 2 for subjects to create such specific encodings. This could reduce apparent encoding-match effects in a number of ways. For example, it may be that initially all meanings of a word are activated (Marcel, 1980), but it takes time to store the one meaning biased by context. Conversely, it may be that only the primary meaning is stored initially, and it takes time to store a secondary meaning biased by context.

GENERAL DISCUSSION

The SAM model was fit to a wide range of tasks involving the presentation and testing of multiple items. In all, SAM was fit to 39 test conditions in Experiment 1 and 7 in Experiment 2, *without assuming any search processes* in recognition. All recognition decisions in the version of SAM described here are based entirely on a familiarity measure obtained by probing memory with weighted cues that jointly produce global activation of memory.

An important feature of the SAM model is the multiplicative (interactive) way in which multiple cues are combined. This feature, which provides a way to predict an intact advantage, allowed the data of Experiment 2 to be fit without adding encoding match assumptions. However, without match parameters greater than 1.0, the model underpredicted the intact advantage found in Experiment 1. The encoding-match assumption states that when an item is tested with other contextual items with which it was learned, the test item will be encoded in much the same way as it was when it was presented for study. In the model, this was implemented by increasing cue-to-image strengths when items were tested in their original verbal context.

Alternatives to Encoding Match

The basic SAM model without an encoding-match process could not satisfactorily handle the results from Experiment 1. We now consider alternative ways in which the SAM model might be extended to handle the data without invoking an encoding-match process. Humphreys, Pike, Bain, and Tehan (personal communication, July 1986) suggested that higher order units are formed in which all items studied together are represented holistically. Their hypothesis might be instantiated within the SAM model by assuming storage of single images representing multiple-item groups. That is, a single image containing all items presented within a given study triad might be stored in addition to storage of the images of individual items in that triad. Of course, the cognitive literature is filled with hypotheses suggesting memory storage of higher order units, particularly in experiments

dealing with meaningfully related materials, such as sentences, paragraphs, and longer texts (Bower, Black, & Turner, 1979; Kintsch, 1974; Kintsch & van Dijk, 1978; Thorndyke, 1977). It seems reasonable that higher order units may also be stored to some extent when unrelated items are presented together. Such proposals are discussed by Arnold (1976) and Mandler et al. (1981). Lee and Estes (1981) have proposed a multilevel model for short-term memory of unrelated items (see also Cunningham, Healy, & Williams, 1984).

Storage of higher order units in addition to associations at the item level may account for the results from the present experiments. The strength between a higher order unit and its own intact test probe would be greater than that between the unit and any other test probe, thus producing a greater intact advantage than would otherwise be seen. However, distinguishing between models positing higher order units and models positing encoding-match processes would not be easy, and we do not pursue the possibility further in the present paper.

A different way to extend the SAM model to handle the present results would be through the assumption that search processes augment a recognition decision based on familiarity. If recall of particular images or particular triads occurs, then a greater intact advantage would be expected due to the occasional recall of information that would override an incorrect familiarity decision. Augmenting SAM in this way would greatly increase the complexity of the recognition model, and distinguishing such an extended model from that presented here would not be easy. This hypothesis is also left for future research.

It is crucial to note that our findings in Experiment 2 do not rule out the possibility of search operating in recognition as long as one operates within a model, like SAM, that includes a factor producing a cued disadvantage. That is, the SAM model predicts that cued performance will decrease relative to single-item performance to the degree that weight is removed from the test item and given to the cue. Thus, processes that alone would provide an advantage for cuing are acceptable within SAM to the degree that they offset the cued disadvantage caused by the sharing of attention weights. The results from Experiment 2 argue against familiarity-plus-recall models of recognition, such as Humphreys's, in which there are no factors that produce a disadvantage for cuing. The argument against recall contributing to recognition performance should not be generalized beyond this point.

REFERENCES

- 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.
- ARNOLD, P. G. (1976). *On the nature of associations*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- ATKINSON, R. C., & JUOLA, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press.
- ATKINSON, R. C., & JUOLA, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Learning, memory, & thinking* (Vol. 1). San Francisco: W. H. Freeman.
- BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, **11**, 177-220.
- CUNNINGHAM, T. F., HEALY, A. F., & WILLIAMS, D. M. (1984). Effects of repetition on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 575-597.
- FISCHLER, I., & JUOLA, J. F. (1971). Effects of repeated tests on recognition time for information in long-term memory. *Journal of Experimental Psychology*, **91**, 54-58.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.
- GLUCKSBERG, S., & MCCLOSKEY, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 311-325.
- HERRMANN, D., FRISINA, R. D., & CONTI, G. (1978). Categorization and familiarity in recognition involving a well-memorized list. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 428-440.
- HINTZMAN, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, **16**, 96-101.
- HUMPHREYS, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition*, **4**, 221-232.
- HUMPHREYS, M. S. (1978). Item and relational information: A case for context independent retrieval. *Journal of Verbal Learning & Verbal Behavior*, **17**, 175-187.
- HUMPHREYS, M. S., & BAIN, J. D. (1984). Recognition memory: A cue and information analysis. *Memory & Cognition*, **11**, 583-600.
- HUMPHREYS, M. S., & BAIN, J. D. (1985, May). *Cue and information analyses as alternatives to the information processing approach to human memory*. Invited address, 57th meeting of the Midwest Psychological Association, Chicago.
- KINTSCH, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- KINTSCH, W., & VAN DIJK, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, **85**, 363-394.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LEE, C. L., & ESTES, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 149-169.
- LIGHT, L. L., & CARTER-SOBELL, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning & Verbal Behavior*, **9**, 1-11.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.
- MANDLER, G., RABINOWITZ, J. C., & SIMON, R. A. (1981). Coordinate organization: The holistic representation of word pairs. *American Journal of Psychology*, **94**, 209-222.
- MARCEL, A. J. (1980). Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. In R. S. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, NJ: Erlbaum.
- MURDOCK, B. B., JR. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.
- MURDOCK, B. B., JR. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.
- MURDOCK, B. B., JR., & ANDERSON, R. E. (1975). Encoding, storage and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, NJ: Erlbaum.
- MUTER, P. (1984). Recognition and recall of words with a single mean-

- ing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 198-202.
- PIKE, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review*, **91**, 281-293.
- REDER, L. M., ANDERSON, J. R., & BJORK, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, **107**, 648-656.
- THOMSON, D. M. (1972). Context effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior*, **11**, 497-511.
- THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University Press.
- THORNDYKE, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, **9**, 77-110.
- TULVING, E., & THOMSON, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.
- UNDERWOOD, B. J., & HUMPHREYS, M. S. (1979). Context change and the role of meaning in word recognition. *American Journal of Psychology*, **92**, 577-609.
- WINOGRAD, E., & CONN, C. P. (1971). Evidence from recognition memory for specific encoding of unmodified homographs. *Journal of Verbal Learning & Verbal Behavior*, **10**, 702-706.

NOTES

1. The data are available in table form from Richard M. Shiffrin for researchers interested in additional quantitative analyses or fitting of other models.
2. It should be noted that the nine criteria and the five remaining parameters play quite different roles. A variety of predictions can be derived from the five parameters alone. For example, techniques based on signal detection theory could be used to convert the $pr(\text{old})$ values into estimates of the spacings among the various familiarity distributions (akin to the d' values of signal detection theory). These spacings are predicted by the five basic parameters independently of the criterion values. We have chosen, however, to fit the raw data, rather than some derived data, because the assumptions used in converting the raw data into another form could well be questioned.
3. Cue weighting and context matching may be independent if cue weighting has its effect only after the test item is encoded. However, cue weights may be correlated with encoding processes, if both reflect a common attentional mechanism. If so, one may expect encoding-match effects to be small when cue items are given low weights.

(Manuscript received April 23, 1986;
revision accepted for publication December 2, 1986.)