

Cuing effects and associative information in recognition memory

STEVEN E. CLARK

University of California, Riverside, California

and

RICHARD M. SHIFFRIN

Indiana University, Bloomington, Indiana

Item recognition requires discrimination of studied words from nonstudied words. Associative recognition requires subjects to discriminate studied word groups from recombinations of words from different groups. Cued recognition requires the same old-new discrimination as item recognition, but list items are presented as cues along with the test item. The results from three experiments show (1) little or no effect of cuing for low-frequency words, but (2) positive cuing effects for high-frequency words; (3) increasing levels of overall performance with increases in study time, but (4) unchanging effects of cuing with study time; and (5) stronger positive cuing effects for two cues than for one cue. Five models (Independent Cue Model, Matrix model, MINERVA 2, SAM, and TODAM) were fit to the data of Experiment 1. Each model has trouble with at least one aspect of the results. Theoretical implications and modifications are discussed at length.

Current models of recognition memory can, in principle, account for both item and associative recognition. Item recognition requires subjects to distinguish events that have occurred from events that have not. Associative recognition requires discrimination of events that have co-occurred from events that have not. In the typical paradigm, subjects study lists of word pairs, with pairs denoted AB, CD, EF, GH, etc., and they must distinguish between *intact* (AB) and *rearranged* (AD) test pairs. In this paper, attention is focused on the following models: Gillund and Shiffrin's (1984) search of associative memory (SAM) model, Humphrey's (1978) Independent Cue Model (ICM), Hintzman's (1984) MINERVA 2, Murdock's (1982) TODAM, and Pike's (1984) Matrix model.

The mechanisms that these models posit as underlying item and associative recognition are quite varied; however, it has been shown that they make very similar, and in some cases identical, predictions (Humphreys, Pike, Bain, & Tehan, 1989). We will test these models by applying them to data from a variety of test conditions. In the three experiments reported in this paper, memory is tested with cued recognition in addition to item and associative recognition. Cued recognition requires the same discrimination as item recognition, except that cue words,

which the subject *knows* are from the list, are presented with each test item. Thus, AB and AX denote target and distractor trials, where A indicates the cue item, and B has been studied with A.

A central question in the current research regards the relationship between item, cued, and associative recognition. Cued recognition provides an important test condition because (1) it has characteristics of both item and associative recognition, (2) the effect of adding cues (comparison between cued and item recognition) has been shown to vary across experiments, and (3) it provides an additional constraint on models. Each of these points is discussed briefly below.

1. As noted above, cued recognition requires precisely the same discrimination as does item recognition. However, to the extent that the episodic association between the cue and the test item is utilized, cued recognition may be based on an associative discrimination. At an extreme, cued recognition may be approached in the same manner as associative recognition, where B is called old because A and B have co-occurred, and X is called new because A and X have not co-occurred.

2. The effect of adding cues varies across experiments, the results ranging from strong facilitation (Thomson, 1972, Experiments 1 and 4; Tulving & Thomson, 1971), through little or no facilitation (DaPolito, Barker, & Wi-ant, 1972; Slamecka, 1975; Thomson, 1972, Experiments 2, 3, and 6; Underwood, 1974), to interference (Clark & Shiffrin, 1987; Gillund & Shiffrin, 1984; Gronlund, 1986).

However, only in Gronlund's (1986) study was the interfering effect of adding cues statistically significant. Following study of a list of five-tuples (denoted ABCDE,

This research was supported by Grant 5-538430-1990-7 from the Academic Senate of the University of California to S.E.C. Grant NIMH 12717 to R.M.S. Special thanks to Doug Hintzman, Ben Murdock, Arthur Glenberg, and an anonymous reviewer for extensive comments and helpful discussion, and to Michael Fitzwater for conducting Experiments 2 and 3. Correspondence should be addressed to S. E. Clark, Psychology Department, University of California, Riverside, CA 92521 (e-mail: clark@ucrvms.bitnet).

where each letter represents a word), subjects were better at distinguishing two old words (DE) from two new words (XY) when the words were tested alone than when the two words were tested with the other three words from the sentence (ABCDE vs. ABCXY).

Clark and Shiffrin (1987) presented subjects with unrelated word triples (ABC, DEF, etc.) and showed that recognition performance declined slightly as cues were added at test. Performance was best for single-item recognition (C vs. X), slightly worse for cued recognition with one cue (BC vs. BX), and worse yet with two cues (ABC vs. ABX). However, the differences were small and statistically unreliable. Similarly, Gillund and Shiffrin (1984) showed a (statistically unreliable) cuing deficit with word pairs.

Of the models listed above, only SAM can account for negative effects of adding cues; other models either predict positive cuing or predict that cuing effects will covary with associative recognition performance. One purpose of the current work is to determine the extent to which negative cuing effects occur.

3. Cued recognition may provide an important third data point for model-fitting. A full discussion of model predictions and their underlying mechanisms will be given after Experiment 1. The bottom line of that discussion is the following:

MINERVA 2 predicts a large cuing advantage because the associative information in the AB pair contributes to its familiarity. For TODAM, the ICM, and the Matrix model, the effect of cuing will covary with the level of associative recognition performance. In TODAM and the ICM, this is because the associative information that underlies associative recognition also produces positive cuing. Unlike in MINERVA 2, positive cuing is not obligatory, but it will appear concurrently with high levels of associative recognition performance. The predictions are linked in the Matrix model for similar reasons (however, while item and associative information may vary independently in TODAM and the ICM, they are not separate in the Matrix model).

Negative cuing is predicted only by SAM, and the original version of SAM cannot predict positive cuing. The reason is that SAM assumes that retrieval resources are limited, and these limited resources must be distributed among the multiple cues used to probe memory. Attention given to the cue is taken away from the test item. Clark and Shiffrin (1987) suggested that when the resource cost of utilizing a cue exceeds its effectiveness, use of the cue will produce a decrement in memory performance, relative to when the cue is not present (or not used).

What factors underlie a cue's effectiveness? Certainly, the strength of the cue's association to the test item is one factor. We will test the hypothesis that it takes more time to encode and store associative information than it does to encode and store item-specific information. Thus, longer presentation rates should lead to cuing advantages.

This presentation time hypothesis presumes some degree of independence of item and associative informa-

tion. The models under consideration posit differing relationships between these types of information. Briefly, SAM, TODAM, and the ICM assume that item-specific and associative information are stored separately, and may vary independently. MINERVA 2 and the Matrix model do not explicitly distinguish between item and associative information at the level of the memory trace. Consequently, predictions of these models are more constrained than those for TODAM and SAM.

EXPERIMENT 1

Pairs of words were studied, and single (A vs. X) and cued (AB vs. AX) recognition were tested. Pair recognition, which requires discrimination of intact (AB) pairs from two new words (XY), and associative recognition (AB vs. AD) were also tested. Associative recognition places a constraint on models like TODAM and the Matrix model, which can predict positive or negative cuing effects, but which cannot simultaneously predict negative cuing and high levels of associative recognition performance.

A specific hypothesis tested is that cuing deficits occur when the episodic associative link between A and B is weak, and that longer presentation rates will lead to an increasing cuing advantage. Thus, study pairs were presented at fast or slow rates in different lists. We wanted to provide conditions that would be likely to produce a cuing deficit; thus, based on previous results, low-frequency words were used.

Method

Subjects. Forty-eight subjects from Indiana University participated, either to fulfill an introductory psychology course requirement or for cash.

Procedure and Materials. Eight study lists were presented, each followed by a 30-sec arithmetic distractor task and one of four kinds of recognition test: single-item, cued, pair, or associative. Low-frequency words listed as having fewer than four occurrences per million in Thorndike and Lorge's (1944) and Kučera and Francis's (1967) word frequency norms were used to create study lists and recognition tests.

Each study list consisted of 16 word pairs. Four lists were presented at a fast rate (1.25 sec/pair) and four were presented at a slow rate (5.0 sec/pair).

Following presentation of each list, subjects mentally added 15 single digits (from 4 to 9) presented at a 2-sec/digit rate. Following the arithmetic task, one of four recognition tests was presented: single-item (S), cued (Q), pair (P), or associative (A) recognition.

Single-item recognition (S). On each trial, a study list word (A) or a new word distractor (X) was presented. The subject's task was to say whether the word was old (from the list) or new. Half of the target items corresponded to right-hand members of a study pair, and half to the left-hand members. The targets were presented in their appropriate locations (left, right), and half of the distractors were presented on the left and half on the right. A blank line was presented next to the test item to indicate the word's position. Sixteen old and 16 new items were tested.

Cued recognition (Q). On each test trial, two words were presented—a cue and a test item. The cue was always one member of a studied pair. The test item was either the other member of the pair, or a new item. Let AB denote A as a cue and B as the (old)

test item, and let AX denote A as a cue and X as a (new) test item. The cue item was always the left-hand item, and the test item was always the right-hand item.¹ Eight old and eight new items were tested. A cue from a given study pair was used in a test pair only once. The subjects were fully informed regarding the cue's relation to the test item for target and distractor trials. The subjects were told that the cues would be useful in helping them make their recognition decisions, and they were encouraged to utilize them. In addition, the subjects were explicitly told that on old trials, the cue would be the word originally paired with the test word.

Pair recognition (P). Target trials consisted of two old items studied as a pair (AB), or two new items (XY); 16 old and 16 new pairs were tested.

Associative recognition (A). An intact test consisted of two old items studied as a pair (AB). A rearranged test consisted of two old items studied in different pairs (AD). Positions of test words were never reversed relative to study; for example, BA and DA probes were not tested. Five intact and five rearranged pairs were tested.

For all of the test conditions, words from a given study list pair contributed to only one test trial. For example, given study of AB, CD, and EF, if AD was tested as a rearranged pair in associative recognition, B and C would not be used for any test trial—AB, CD, and CB could not be tested. Similarly, for single-item and cued recognition, A and B items could not both be tested.

Test trials were group-paced so that the next test trial appeared after all subjects entered responses on the current trial. The subjects were run in groups of 1 to 4.

Prior to the collection of data, the subjects were given extensive instructions regarding the nature of the study lists and testing procedures, and they were shown four practice study lists, each of which was followed by one of the different test conditions. Thus, the subjects were exposed to the kind of discrimination required by each of the four tasks, and they were familiar with the testing procedures and apparatus. In the practice lists and during the experiment, test conditions were revealed only after the study list had been presented, thus minimizing the possibility of subjects' adopting different learning strategies for different test conditions.

The selection of words and the assignment of the words to conditions were random. The eight study-test trials were presented in one of eight different orders for each subject. Over the eight orderings, each test type/presentation rate combination occurred equally often in each of the eight positions.

Stimulus presentation and response collection were controlled by a DEC PDP-11/34 computer. The stimuli were presented on VT-220 and VT-240 terminals.

Results

Table 1 gives the hit rates and false alarm rates averaged over subjects for each test/presentation rate condition, plus the average of the d' 's for individual subjects for each condition (for subjects who showed hit rates equal to 1.0 or false alarm rates of zero, estimates were calcu-

lated with the procedure described by Murdock & Ogilvie, 1968). Statistical tests were calculated on hit rate minus false alarm rate (HR - FAR) rather than d' (note, however, that the pattern of results is the same for both the linear [HR - FAR] and nonlinear [d'] dependent measures). A test type \times presentation rate \times test order ($4 \times 2 \times 8$) analysis of variance (ANOVA) showed no effect of order, but a large effect for test type [$F(3, 120) = 45.53$, $MS_e = .029$, $p < .0001$], and presentation rate [$F(1, 40) = 195.76$, $MS_e = .035$, $p < .0001$], as well as a significant test type \times presentation rate interaction [$F(3, 120) = 5.46$, $MS_e = .033$, $p < .005$]. The results can be summarized as pair $>$ single \geq cued $>$ associative, for both the fast and slow presentation rate conditions. The trend toward a negative cuing effect was not statistically significant for fast or slow presentation rates.

The effect of increasing study time is shown in Table 1 in terms of the ratio of d' for slow and fast presentation rates, d'_s/d'_f , and in differences in HR - FAR for slow and fast presentation rates. Both measures show the effect of increasing study time to be roughly twice as large for associative recognition as for the other conditions. An ANOVA of the differences showed that the differences were reliable [$F(1, 141) = 5.95$, $MS_e = .083$, $p < .001$], and Tukey's honestly significant difference (HSD) was .153, indicating that the effect of increasing presentation rate for associative recognition differed from all the other test types, but that the effects of presentation rate for single-item, cued, and pair tests were not different.

Discussion

A very small, statistically insignificant negative cuing effect was shown for both fast and slow presentation rates. This result replicates earlier results of Clark and Shiffrin (1987) and Gillund and Shiffrin (1984). One interpretation of this result is that subjects simply ignored the cue, despite instructions to use it. Moreover, as will be shown in the next section, all of the models with the exception of SAM have mechanisms that will produce a large cuing advantage if the cue is used. Thus, advocates of these models must explain why subjects who are encouraged to use the cue ignore it. Within the SAM framework, attention to the cue produces a deficit, so it may plausibly be argued that subjects might learn to ignore it.

The hypothesis that storage of associative information requires more study time than item information was supported by data showing that the increase in performance

Table 1
Average Hit and False Alarm Rates Plus d' for Experiment 1

	Single-Item		Cued		Pair		Associative	
	HR	FAR	HR	FAR	HR	FAR	HR	FAR
Fast	0.80	0.17	0.77	0.17	0.84	0.11	0.73	0.44
d'	2.01		1.98		2.67		0.85	
Slow	0.90	0.07	0.90	0.09	0.95	0.03	0.87	0.15
d'	3.11		2.86		3.65		2.27	
d'_s/d'_f	1.55		1.44		1.37		2.67	
$(HR - FAR)_s - (HR - FAR)_f$.24		.21		.19		.43	

with longer presentation times was much larger for associative recognition than the other test conditions. At the fast presentation rate, associative recognition was considerably worse than the other recognition tasks, but at the slow presentation rate associative recognition performance was much closer to the other tasks.

The increase in associative recognition performance suggests that the A-B associations are stronger at the slow presentation rate. However, the strengthening of A-B associations still did not produce a cuing advantage. This result poses a puzzle that will be addressed later in the article.

MODELS

SAM (Gillund & Shiffrin, 1984), TODAM (Murdock, 1982), Pike's (1984) Matrix model, MINERVA 2 (Hintzman, 1984), and Humphreys's (1978) ICM are described below. With the exclusion of the ICM, these models constitute a class of models called *interactive cue global matching models*.

Models in this class share two characteristics: (1) They base recognition decisions entirely on the familiarity of the test probe (without appeal to any additional processes like search). This familiarity represents information accessed directly from all items in memory (constrained to list items) on the basis of a global match (or global activation) of the test item to all items in memory. This global match differs from earlier local match familiarity models, in which only the representation of the test item is accessed (see, e.g., Kintsch, 1970). (2) Multiple cues in these models are combined interactively such that the match (or activation) given by cues A and B is not a linear function of the match of the separate word cues.

The ICM proposed by Humphreys (1978) does not combine cues in this interactive manner, but rather posits dependencies such that retrieval of associative information is contingent on retrieval of item-specific information. Although conceptually different from any of the interactive cue global matching models, the Independent Cue Model predicts many of the same results.

The models were fit to the hit rates and false alarm rates shown in Table 1. For SAM, TODAM, and the Matrix model, this was done by deriving expressions for the expectation and variance of the familiarity distributions for target and distractor cases in each test condition at the fast and slow presentation rates. For these models, normal (but not necessarily equal-variance) distributions are assumed. For MINERVA 2, predictions were obtained by means of Monte Carlo simulation, because closed-form expressions for means and variances have not been derived. For the ICM, response probabilities are derived directly.

For all models, fits were evaluated with the chi-square statistic (including probabilities of misses and correct rejections, which are not shown in Table 1). The data yield 16 degrees of freedom. For each model, the degrees of freedom for the fit is 16 minus the number of free parameters.

Although the predictions of all models differ significantly from data, the chi-square values give a useful index of relative fit across models. Also, despite the high chi-square values, the predictions for hit and false alarm rates are quite close for some of the models. The proportion of variance accounted for (r^2) was quite high for all model fits, all r^2 s being above .97.

The best-fit predictions of each model are given in Table 2, and the corresponding parameters and chi-square values are given in Table 3. Table 2 gives hit and false alarm rates and d 's calculated from those hit and false alarm rates. (The d 's were not used for fitting the models; they are listed as a convenient summary statistic. Since all the d 's in Table 2 are calculated from average hit and false alarm rates, they will not correspond to the data d 's in Table 1.) The model-fitting algorithm minimized chi-square error for the response frequencies for hits, misses, false alarms, and correct rejections, not d 's.

SAM

In the SAM model, individual words are stored separately in units called *images*.² Recognition is based on the

Table 2
Best Fits to Experiment 1 for SAM, TODAM, Matrix, MINERVA 2, and ICM

	Single-Item			Cued			Pair			Associative		
	HR	FAR	d'_s	HR	FAR	d'_q	HR	FAR	d'_p	HR	FAR	d'_a
Fast Presentation Rate												
Data	.80	.17	1.80	.77	.17	1.69	.84	.11	2.22	.73	.44	0.76
SAM	.82	.16	1.91	.76	.17	1.66	.84	.12	2.17	.69	.50	0.50
TODAM(2)	.77	.19	1.62	.77	.17	1.69	.85	.10	2.32	.70	.47	0.60
Matrix	.72	.23	1.32	.77	.18	1.65	.88	.09	2.52	.74	.44	0.79
MINERVA 2	.69	.28	1.08	.77	.18	1.65	.86	.10	2.36	.81	.33	1.32
ICM	.77	.25	1.42	.83	.25	1.62	.84	.06	2.22	.68	.47	0.54
Slow Presentation Rate												
Data	.90	.07	2.76	.90	.09	2.62	.95	.03	3.53	.87	.15	2.16
SAM	.92	.06	2.96	.91	.09	2.68	.95	.03	3.53	.71	.30	1.08
TODAM(2)	.86	.11	2.31	.92	.07	2.88	.96	.02	3.81	.81	.20	1.72
Matrix	.79	.16	1.80	.92	.08	2.81	.97	.02	3.94	.83	.19	1.83
MINERVA 2	.87	.11	2.35	.92	.08	2.81	.96	.02	3.81	.89	.14	2.31
ICM	.87	.12	2.58	.94	.12	2.62	.95	.01	3.87	.83	.16	1.94

Table 3
Parameter Values for Best Fits to Experiment 1 for SAM, TODAM, Matrix,
MINERVA 2, and ICM models

SAM	$b_f = .0628, b_s = .1700$ $c_f = .1002, c_s = .1800, d = .0100, w_q = .16$ $\chi^2(2) = 69.90, \chi^2/df = 34.95$
TODAM(2)	$\gamma_i(f) = .5500, \gamma_i(s) = 2.5016, \gamma_a(f) = .1270, \gamma_a(s) = 1.7984, N = 190$ $\chi^2(3) = 45.72, \chi^2/df = 15.24$
Matrix	$\mu_f = 4.1, \mu_s = 4.6875$ $\sigma_f = 3.801, \sigma_s = 5.9625, N = 493$ $\chi^2(3) = 187.50, \chi^2/df = 62.50$
MINERVA 2	$e_f = .355, e_s = .604, N = 20$ $\chi^2(5) = 140.48, \chi^2/df = 28.096$
ICM	$\alpha_f = .695, \alpha_s = .856, \beta_f = .354, \beta_s = .641$ $g_{1f} = .248, g_{2f} = .518, g_{3f} = .734$ $g_{1s} = .115, g_{2s} = .161, g_{3s} = .671$ $\chi^2(6) = 134.78, \chi^2/df = 26.956$

Note—Parameters are described in the text. Criterion parameters are not listed for SAM, TODAM, MINERVA 2, and Matrix models.

connection strengths between stored images and the cues used to probe memory (including the test item). These cue-to-image strengths are combined multiplicatively and summed over all N images in memory to give a measure of familiarity,

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

In Equation 1, the Q_i refer to the cues used to probe memory, including the test item(s) and context, and w_i is a cue weight indicating the proportion of limited resources allocated to cue Q_i . The limitation on resources is instantiated by the constraint that $\sum w_i$ is a constant, usually set to 1. Generally, memory is probed with the test item(s) and context as cues. The context cue is assumed to focus retrieval on the study list, and in previous applications (Clark & Shiffrin, 1987), the weight on context has been set to 0.5.

Each cue-to-image strength is a random variable with expectation determined by parameter values. The variance is approximated by a three-point distribution. Specifically, the value for any $S(Q_i, I_j)$ with mean x is $x(1+v)$, x , or $x(1-v)$, each with a probability of 1/3 (usually v is set equal to .5).

The expectations for $S(Q_i, I_j)$ are given by four parameters. The strength between the context cue and image I_j is denoted a . The strength between Q_i and its own image I_i is a self-strength, denoted c . The strength between a cue and the image of a word with which it has been rehearsed is an interitem strength denoted b . A residual

strength, d , is assumed between a cue and an image with which it has not been rehearsed.

Assuming independence between images (i.e., the variability in $S(Q_i, I_j)$ is uncorrelated for all i and j), expressions can be derived for the expectation and variance of the familiarity distribution for each of the different test probes. Hit and false alarm rates were obtained by assuming that familiarity distributions are normal, and by calculating the proportion of target and distractor distributions above a criterion that varied as a free parameter.

The free parameters in the model were as follows: b_f and b_s are interitem strengths for fast and slow presentation rates, c_f and c_s are self-strengths for fast and slow rates, and d is a residual strength. The residual strength applies to words not rehearsed together, so the value of d is determined only by preexperimental factors (but see Shiffrin, Ratcliff, & Clark, 1990) and thus did not vary with presentation rate. (It is reasonable to assume that context strength increases with presentation time; however, if only list items are activated by the probe cues, the context parameter cancels out of all the d' equations, so the context strength parameter is ignored in the present work.)

The context cue w_c is given a fixed weight of .5, and the remaining weight of .5 is evenly divided among the test word cues, except in the case of cued recognition where a weight w_q is given to the cue,³ and w_t is given to the test item, where $w_t = .5 - w_q$. The other weights are as follows: w_s denotes the weight on the item cue in single-item recognition, w_p denotes the weight on each item cue in double-item tests (pair and associative recog-

niton); w_s is set to .5, and w_p is set to .25. Only w_q is free to vary.

Since the decision criterion for responding "old" varied freely for each of the eight study-test conditions, a total of 14 parameters varied freely.

The predictions for d' for the four test conditions are as follows:

$$d'_s = (\alpha_s/N^{1/2})[(c/d)^{w_s} + (b/d)^{w_s} - 2] \quad (2)$$

$$d'_q = (\alpha_q/N^{1/2})[(c/d)^{w_q}(b/d)^{w_t} + (b/d)^{w_q}(c/d)^{w_t} - (c/d)^{w_q} - (b/d)^{w_q}] \quad (3)$$

$$d'_p = (\alpha_p/N^{1/2})2[(c/d)^{w_p}(b/d)^{w_p} - 1] \quad (4)$$

$$d'_a = (\alpha_a/N^{1/2})2[(c/d)^{w_p}(b/d)^{w_p} + 1 - (c/d)^{w_p} - (b/d)^{w_p}] \quad (5)$$

The α terms are constants whose values depend only on the cue weights. For $w_s = w_c = .5$ and $w_p = .25$, $\alpha_s = 3.22$, and $\alpha_p = \alpha_a = 3.70$. The value of α_q depends on the free parameter w_q ; $\alpha_q = \alpha_s$ if w_q is zero, and $\alpha_q < \alpha_s$ for all nonzero w_q . For the best-fit value of $w_q = .16$, $\alpha_q = 3.63$. The derivation of α terms is described in Appendix A.

It can be seen from Equations 2 and 3 that cued recognition performance cannot exceed single-item recognition. Note that $w_s = w_q + w_t$, and all the terms in parentheses are greater than or equal to 1. The α terms have almost no effect on d' . By symmetry, Equation 3 reaches a maximum when w_q equals zero. At that point, $d'_s = d'_q$, so it is clear that $d'_s \geq d'_q$. This point can be verified by setting w_q to zero in Equation 3. If w_q is zero, then $\alpha_q = \alpha_s$, the weight on the test item in cued recognition will be equal to the weight on the test item in single-item recognition ($w_t = w_s$), and all terms raised to the w_q power are equal to 1. Equations 2 and 3 are then identical.

Although the model clearly deviates from the data [$\chi^2(2) = 69.90$], several aspects of the data are captured in the fit. The model makes accurate predictions for old-new discrimination (single-item, cued, and pair tests) at fast and slow presentation rates. However, associative recognition performance is underpredicted at the fast presentation rate, and particularly at the slow presentation rate, where the error in prediction contributes a chi-square of 52.24 to the total error chi-square of 69.90; 77% of the total error comes from this comparison. There is something happening in associative recognition that the model in its present form is missing. We will return to this issue in the general discussion. It is useful to note at this point that the self-strength parameter c rises by a factor of 1.8 from the fast to the slow rate, but that the inter-item strength parameter b rises by a factor of almost 3, suggesting accentuation of associative information at the slower rates.

TODAM

A word is represented as a normalized vector of independent feature values, each with expectation equal to

zero. We will distinguish between stimuli and their corresponding vector representations with upper and lower case. Thus, Stimulus A is represented by vector \mathbf{a} . For a given AB pair, three vectors are stored, \mathbf{a} and \mathbf{b} , and the convolution of \mathbf{a} and \mathbf{b} , which represents the association between \mathbf{a} and \mathbf{b} , and is denoted $\mathbf{a} * \mathbf{b}$. Unlike SAM, items are not stored separately, but rather are summed together into a single vector, \mathbf{M} , which represents a composite memory. Specifically,

$$\mathbf{M}_j = \alpha \mathbf{M}_{j-1} + \gamma_1 \mathbf{a}_j + \gamma_2 \mathbf{b}_j + \gamma_3 (\mathbf{a}_j * \mathbf{b}_j), \quad (6)$$

where \mathbf{M}_j is the memory vector after the j th presentation, \mathbf{a}_j and \mathbf{b}_j are item vectors, $(\mathbf{a}_j * \mathbf{b}_j)$ is the a-b association, α is a forgetting parameter, and γ_1 , γ_2 , and γ_3 are weights. In the original description of the model, all γ were set equal to 1.0, although in some applications (see Weber, 1988), they have been constrained to sum to 1.0. We allowed the weights to vary freely, with the only restriction that $\gamma_1 = \gamma_2$. In accord with this assumption, we denote the weight on an item vector (γ_1 , γ_2) as γ_i and the weight on an associative vector (γ_3) as γ_a . The forgetting parameter α does not play a substantive role for our studies, so it is not considered further (i.e., it is set to 1.0 for convenience).

Recognition is based on the familiarity of the test probe, which is given by the match of the test probe vector to the memory vector. This match is computed as the inner dot product of the two vectors.

Two versions of TODAM are considered, denoted T-1 and T-2, which vary with respect to the information used to probe memory.⁴ T-1 assumes that single-item tests probe with the single vector, and that all double-item tests probe with the convolution of the two test item vectors. This model can be rejected out of hand, because it predicts essentially equal performance for all double-item tests. This prediction has been verified previously by Weber (1988).

Model T-2 also assumes that for a single-item test, memory is probed with the single test vector. For the cued tests, memory is probed with the sum of the test item vector and the pair convolution vector. For the pair tests, memory is probed with the sum of the single vectors corresponding to both test items, and the convolution vector for the test pair. For the associative tests, only the convolution vector is used as a probe (since single items cannot help).

Thirteen parameters were varied for Model T-2; $\gamma_i(f)$ and $\gamma_i(s)$ are storage weights for single-item vectors at fast and slow presentation rates, and $\gamma_a(f)$ and $\gamma_a(s)$ are the corresponding weights for convolution (associative) vectors. The length of the single-item vector, N , did not vary with study time. Eight decision criteria varied freely for each study-test condition. The means and variances for each condition are given in Appendix B.

Expressions for d' for single-item, cued, pair, and associative recognition are given as follows:

$$d'_s = \gamma_i(A)^{1/2} \quad (7)$$

$$d'_q = (\gamma_i + \gamma_a) / \{A + B + [(N+1)/N^2](\gamma_i^2 + \gamma_a^2)\}^{1/2} \quad (8)$$

$$d'_p = (\gamma_a + 2\gamma_i) / (B + 2A)^{1/2} \quad (9)$$

$$d'_a = \gamma_a \left/ \left[\left(B + \frac{2N+2}{N^2} \right) (\gamma_i^2 + \gamma_a^2) \right]^{1/2} \right., \quad (10)$$

where

$$A = \left[(2/N)\gamma_i^2 + \gamma_a^2 \left(\frac{3N^2+1}{4N^3} \right) \right] K,$$

$$B = \left[\left(\frac{3N^2+1}{2N^3} \right) \gamma_i^2 + \left(\frac{2N^2+1}{3N^3} \right) \gamma_a^2 \right] K,$$

and K is the number of pairs.

From Equations 7 and 8, it is clear that a cued advantage is predicted for large γ_a relative to γ_i . Cued recognition is based on a mixture of item and associative information, so that as the relative contribution of associative information increases, so too will the predicted positive effect of the cue. The value of γ_a is constrained by paired and associative recognition. Lowering values for γ_a can produce a cuing disadvantage but only at the cost of underpredicting performance for associative recognition. This pattern may be seen in Table 2. The overall fit of T-2 shown in Table 2 is relatively good, although predictions deviate from the data [$\chi^2(3) = 45.72$]. At fast and slow presentation rates, single-item recognition performance is slightly underpredicted, and cued recognition performance is slightly overpredicted. Particularly at the slow presentation rate, the relationship between single-item and cued recognition is mispredicted. Like SAM, associative recognition performance is underpredicted at the slow presentation rate, although the error is not as great. Perhaps more to the point, TODAM predicts a bigger advantage for cued over associative recognition than that observed, especially at the slow presentation rate.

Despite these mispredictions, the fit of the model is very good. Perhaps this is not surprising, given the freedom of choosing values for γ_i and γ_a . Item and associative information are independent in TODAM, so the relation between single-item and cued recognition can be adjusted freely; increasing the contribution of associative information has no effect on single-item recognition. Note that γ_i increases by a factor of five (from .55 to 2.5) from the fast to the slow presentation rate, and γ_a increases by a factor of 14 (.127 to 1.8). At the fast presentation rate, $\gamma_i/\gamma_a = .55/.127 = 4.33$. At the slow presentation rate, this ratio is $2.5/1.8 = 1.39$. Thus, according to TODAM, associative information is accentuated at slower presentation rates. This is the same pattern as that shown by the SAM model: in SAM, the slow/fast ratios of interitem strengths were much larger than the ratios of self-strengths.

Matrix Model

Words are represented as vectors. Values of vector elements are random variables, each with expectation equal

to μ and variance equal to σ^2 . Associations are represented by vector multiplication (outer product); specifically, the association for AB is given by multiplication of the a column vector by the b row vector to form a matrix. The a-b association matrices are summed together cell by cell into one distributed memory, represented as the $N \times N$ matrix M .

Recognition is based on the familiarity or match of the test item matrix with the memory matrix. As in TODAM, the match is given by the dot product. For single-item tests, a test matrix is constructed by taking the product of the test item vector and a vector of ones. Equations for d' (and for the means and variances) are very complicated and do not reduce to simple enough expressions to merit their incorporation in the text; they are given in Appendix C.

The Matrix model can produce a cuing advantage or disadvantage. As σ^2 is increased relative to μ , an increasing cuing advantage is predicted. In fact, increasing σ^2 relative to μ produces performance increases in all test conditions. The reason for this is that recognition is based on the relative match of target and distractor probes to memory. As σ^2 increases relative to μ , the difference in the match of targets and distractors is expanded.

Thus, like TODAM, the Matrix model predicts a correlation between associative recognition performance and a positive cuing effect. This behavior in the model can be seen in Figure 1, which shows predictions for d'_i , d'_q and d'_a , varying μ from 3 to 6 while σ is constant at 3.0. Vector length was set at 100, 200 (not shown), and 500. Thus, an increasing cuing advantage should be associated with increasing associative recognition performance, a result that is not found.

In fitting the Matrix model to Experiment 2, 13 parameters were free to vary: μ_f , μ_s , σ_f^2 , and σ_s^2 , the mean and variance of the match values for test matrix cell $t(i,j)$ to memory matrix cell $M(i,j)$ for fast and slow presentation rates, N the vector length (which does not vary with presentation rate), and the eight decision criteria.

The fit of the model was relatively poor [$\chi^2(3) = 187.5$]. The predictions are given in Table 2. The main source of the error is clear: single-item recognition is underpredicted particularly at the slow presentation rate. The reason for this is also clear. Unlike TODAM and SAM, item and associative information are not stored separately, and their strengths cannot vary independently. Except in the single-item case, the Matrix model exhibits a difficulty also seen in TODAM: The predicted difference between cued and associative conditions is too large, especially at the slow presentation rate.

MINERVA 2

Word pairs are represented as a single vector, which is the concatenation of two item vectors with Elements 1 to $N/2$, representing item A, and Elements $N/2 + 1$ to N , representing item B (this is consistent with the representation of category name—exemplar pairs in Hintzman, 1986). Vector elements may take on values of +1, -1, or 0. The expectation for a given feature is zero, and the

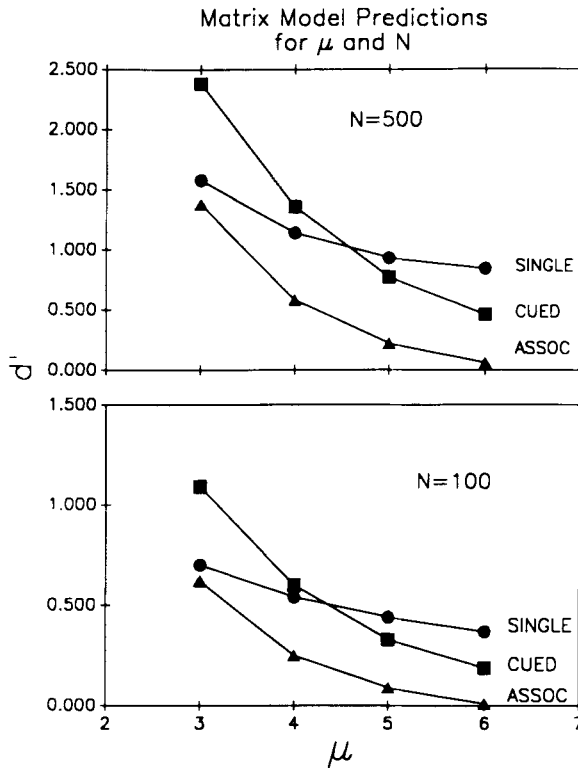


Figure 1. Matrix model predictions for cued, single, and associative recognition, holding σ constant and varying μ and vector length N . Recognition d 's are calculated from hit and false alarm rate predictions.

variance depends on the probability of a nonzero value. Each pair vector is stored separately in memory: MINERVA 2 stores K pairs as K separate memory vectors $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_K$.

An index of familiarity is given by taking the vector dot product of the test item vector \mathbf{t} and each memory vector \mathbf{m}_i , cubing it, and summing these values across all memory vectors. The familiarity of a test item \mathbf{t} is given as

$$F(\mathbf{t}) = \sum_{i=1}^N 1/N_R \left[\sum_{k=1}^K \mathbf{t}(k)\mathbf{m}_i(k) \right]^3, \quad (11)$$

where the N_R is the number of cells for which either $\mathbf{t}(k)$ or $\mathbf{m}_i(k)$ is nonzero.

For single-item tests, the test vector contained elements of the test item vector, and the other half of the test vector contained zeros. Double-item tests concatenated the two single-item vectors for each test type.

Cubing the dot product match of test vector and memory vector introduces a nonlinearity into the model, which allows the discrimination of intact test pairs from rearranged test pairs. This same nonlinearity, however, predicts a substantial positive cuing effect.

Eleven parameters varied freely: e represents the probability of correctly encoding feature j in the stimulus and storing it in cell j of its corresponding memory vector.

This encoding parameter is assumed to vary with presentation rate, giving two free parameters, e_1 and e_s . The length of the memory vector, N , also varied, but it did not vary with presentation rate. The decision criterion also varied freely in each of the eight study-test conditions.

Derivations for expectations and variances of familiarity have not been developed for MINERVA 2, although their development is under way (Sheu, 1990). Thus, we obtained predictions by Monte Carlo simulation. Each set of predictions is based on 500 simulations.

The model did not yield a good fit relative to other models [$\chi^2(5) = 140.48$], although the fit was better than that of the Matrix model. Also, the pattern of misprediction was different from that for other models. Whereas SAM, TODAM, and the Matrix model tended to underpredict associative recognition performance, MINERVA 2 overpredicted it, particularly at the fast presentation rate. Like the Matrix model, MINERVA 2 underpredicted single-item recognition performance, but unlike the Matrix model, MINERVA 2 underpredicted single-item recognition performance for the fast, rather than for the slow presentation rate. MINERVA 2, like the Matrix model, does not store item and associative information separately. Relative performance of single- and double-item tests is constrained. In particular, MINERVA 2 predicts a considerable cuing advantage, contrary to the data.

Independent Cue Model

The ICM was first proposed by Humphreys (1976, 1978), and a similar model was later proposed by Mandler (1980). For the present purposes, we will focus here on the Humphreys model. As noted previously, this model is quite different from the others, but it makes many of the same predictions.

In the ICM, item information provides evidence that an item, B, has been presented in some context (i.e., on the list). Relational information indicates that two items, A and B, were rehearsed together (as a pair). The probability of retrieving item information is α , and the probability of retrieving relational information given that item information (for some item) is retrieved is β . The probability that a given test item, B, will be recognized as old when tested alone, $p(\text{Rg B}|\text{B})$, or in a rearranged pair, $p(\text{Rg B}|\text{CB})$, is

$$p(\text{Rg B}|\text{B}) = p(\text{Rg B}|\text{CB}) = \alpha. \quad (12)$$

The probability that B will be called old when tested in an intact pair, $p(\text{Rg B}|\text{AB})$, is

$$p(\text{Rg B}|\text{AB}) = \alpha + (1 - \alpha)\beta\alpha. \quad (13)$$

Thus, even if item information is not retrieved for B, it may still be recognized if A is recognized and used as a cue to retrieve AB relational information. A cued recognition advantage is clearly predicted when $\beta > 0$, which it must be to predict $d'_a > 0$, and a cued disadvantage cannot be predicted.

The probabilities for retrieving item and relational information vary with presentation rate, giving four free

parameters, α_f , α_s , β_f , and β_s . We assumed six guessing parameters. The probability of guessing that a single item distractor is old is g_1 . Subjects may also guess old for an AB probe if item information is retrieved for only one test item. This probability is g_2 . Subjects may also guess that a rearranged distractor is old given that item information is retrieved for both items with probability g_3 . These guessing probabilities varied independently for fast and slow presentation rates.

The fit [$\chi^2(6) = 134.78$] is not as good as that of SAM or TODAM, but it is better than that obtained with the Matrix model and just slightly better than that for MINERVA 2. In some conditions, the model predicts hit and false alarm rates to be too high, which contributes to the error. However, when hit and false alarm rates are converted to d' , the fit of the model improves. As with MINERVA 2 and the Matrix model, the main error for the ICM occurs in predicting single-item recognition performance to be too low relative to that on double-item tests: predictions for single-item recognition are too low, and predictions for cued and pair recognition are too high. The error is not as great as it is for MINERVA 2 or the Matrix model, particularly when the fit is evaluated in terms of d' .

Finally, note that for the ICM, associative (relational) information is accentuated at slower presentation rates: β increased from .354 to .641 for fast and slow presentation, while α increased from .695 to .856. Slow/fast ratios were 1.81 β_s/β_f and 1.23 for α_s/α_f .

Summary

Overall fit. Of the models considered here, SAM and TODAM give the best fits to data. Both models underpredicted associative recognition performance at the slow presentation rate (SAM's error was larger than TODAM's), and TODAM incorrectly predicted a cuing advantage, particularly for the slow presentation rate. The fits for the Matrix model and MINERVA 2 were much poorer.

In evaluating the performance of the models, one cannot overlook that the predictions are generated with different numbers of free parameters. Direct comparisons of performance that take into consideration the differing numbers of free parameters are difficult to interpret. Although goodness-of-fit statistics can be used to evaluate the fit of a single model with varying numbers of free parameters, comparison across different models is not as straightforward. In many cases, it is not clear how additional parameters would be incorporated into the model. For MINERVA 2, for example, the fit would clearly improve were there an additional parameter that weighted item and associative information; however, it is not conceptually clear how such a parameter would be incorporated into the model.

Nonetheless, one can obtain an index of the *efficiency* of each model by dividing the chi-square by its degrees of freedom; these statistics are also listed in Table 3. It can be seen that χ^2/df is smallest for TODAM, indicating that it was the most efficient model. MINERVA 2 and

the ICM also have relatively low values for χ^2/df because each had relatively few free parameters, even though the overall fits were not very good. The SAM model, while giving a good fit, was not very efficient. The Matrix model showed the largest value for χ^2/df .

Cued and single-item recognition. The Matrix model, MINERVA 2, and the ICM clearly mispredict the effect of cuing, each predicting a substantial advantage of cued over single-item recognition. The picture is more complicated for SAM and TODAM. At the fast presentation rate, the cued-single results do not distinguish between TODAM and SAM. TODAM produced a small cuing advantage, and SAM produced too large a cuing disadvantage. However, at the slow presentation rate, the models diverge. SAM still predicts a cuing disadvantage that is just slightly too large, but TODAM incorrectly predicts a large cuing advantage.

TODAM and the Matrix model can predict positive or negative cuing effects, depending on parameters; however, they predict an increasing cuing advantage as associative recognition performance increases. Associative recognition performance increased markedly at the slow presentation rate. Consequently, both TODAM and the Matrix model incorrectly predict a cuing advantage, particularly at the slow presentation rate.

The SAM model, on the other hand, cannot predict a cuing advantage. SAM predicts a negative effect for two reasons: (1) the cue is nondiagnostic in distinguishing targets from distractors, since it is common to both test probes, and (2) retrieval capacity in SAM is limited. Thus any resources given to process the cue must be taken from the test item, and because the cue is nondiagnostic, no return is made on that resource investment.

Representation of item and associative information. The models differ in their representation of item-specific and associative information. In SAM, the two kinds of information are represented by self-strength and interitem parameters, c and b . The self-strength represents the cue's connection to its own representation in memory, and the interitem strength represents the cue's connection to the representation of a word with which it has been rehearsed. These parameters are allowed to vary independently. Similarly, TODAM assumes that item and associative information are stored in separate vectors (within the composite memory) and assumes separate weights γ_i and γ_a on item and associative (convolution) vectors.

While TODAM and SAM both distinguish between item-specific and associative information, they differ in an important respect. In TODAM, associative information plays relatively little role in recognition of single items because the expected dot product match between the convolution vector and one of its component vectors is zero, that is, $E[(\mathbf{a} * \mathbf{b}) \cdot \mathbf{a}] = 0$. Thus, predictions for item and associative recognition are completely unconstrained for TODAM. This is not the case for SAM. The familiarity of the test item is a sum of strengths between the test item A and each image in memory. One of these strengths is the connection from A to image B. The larger this inter-

item connection is, the larger will be the familiarity of A. This factor also underlies the cuing deficit in SAM: A-B associative information contributes to the familiarity of A even when B is not present. Thus, as long as A and B share the weight that would otherwise go to A, adding B as a cue does not add to the familiarity of A as the sole item cue. The ICM is similar to SAM and TODAM in that item and associative information are separate. However, their relationship during retrieval is quite constrained—relational information cannot be retrieved unless item information is retrieved. The ICM fit was worse than for TODAM or SAM, but better than for MINERVA 2 or the Matrix model.

For MINERVA 2 and the Matrix model, the constraints appear too severe. Item and associative information are not stored separately in these models. The Matrix model stores pairs by vector multiplication, and MINERVA 2 stores them as the concatenation of two item vectors. Thus, for both models, the match of a single test item to memory is considered a partial match to its pair. This helps to explain why single-item recognition predictions are so low for these models.

MINERVA 2 makes an additional misprediction not made by any other model; it overpredicted associative recognition performance, especially at the fast presentation rate. In fact, it predicted associative recognition to be better than single-item recognition, contrary to current and previous reports (Clark & Shiffrin, 1987; Humphreys, 1976). Again, item and associative information are not stored separately, and intact-rearranged discrimination is produced by cubing activation. It is probably this cubing assumption that produces such high associative recognition performance.

Presentation rate hypothesis. We hypothesized that item and associative information are stored at different rates, and specifically, that it takes more time to store associative information than it does to store item information. This hypothesis is supported by the empirical and theoretical results. The hypothesis presupposes that item and associative information are stored separately, and models that do not make that assumption could not fit the data. Each model that represents item and associative information separately (TODAM, SAM, and the ICM) showed estimated parameters consistent with relatively more storage of associative information at slow presentation rates. Despite this, all three models underpredicted the magnitude of the increase in associative recognition performance with presentation rate and underpredicted the level of associative recognition performance at the slower presentation rate.

Thus, there are two main findings that the models as a group have some problems in handling: the failure to find an advantage for cued recognition (even when associative recognition improves markedly), and the rather large improvement exhibited by associative recognition when the presentation time is increased (causing the cued and associative performance levels to become fairly close). It may well be that the models will be forced to

incorporate additional forms of associative information, or additional types of retrieval processes, in order to handle such effects.

EXPERIMENT 2

In Experiment 2, we explored conditions that might produce a cuing advantage. In Clark and Shiffrin (1987, Experiment 2) and in Experiment 1 of the present study, low-frequency words were used; this may not have been conducive to the storage and retrieval of the kind of associative information needed to produce a cuing advantage (although it did produce high levels of associative recognition performance).

If one is to obtain a cuing advantage, retrieval gains must exceed attentional cost. There are at least three ways in which this might happen: (1) A higher order AB unit might be stored (such as the $a*b$ convolution; see also Gronlund, 1986; Shiffrin, Murnane, Gronlund, & Roth, 1988); (2) the A cue could be used to retrieve B (via recall perhaps), similar to the independent cue models of Humphreys (1978) and Mandler (1980); or (3) the cue word might bias the encoding of the test word (Clark & Shiffrin, 1987; Light & Carter-Sobell, 1970). These factors may be minimized for low-frequency words. For example, Clark (1992) has shown that associative recognition performance is better for high-frequency than for low-frequency words. Experiment 2 therefore examines cuing effects for low- and high-frequency words.

A second factor may have been operating in previous experiments. Subjects may have switched *retrieval* strategies between test conditions in such a way that associative information was utilized for associative recognition, but not for cued recognition (since it isn't necessary to do the task). Such strategy switching would allow high levels of associative recognition performance without a cuing advantage for old-new recognition. However, switching strategies between associative recognition and cued recognition would be more difficult if the test trials were not blocked by type, but rather varied from trial to trial. A mixed rather than blocked testing procedure was used in Experiment 2, to minimize strategy switching between test conditions.

Also, while there was little evidence that the cued-single-item recognition relationship varied with study time for low-frequency words in Experiment 1, there was strong evidence consistent with the hypothesis that it takes more time to store associative information, and that proportionally more associative information accumulates with additional study time. Additional study time may produce a cuing advantage for high-frequency words, even though it did not produce an advantage for low-frequency words.

Experiment 2 also varied the number of cues in cued recognition. Presumably, monotonic functions should be obtained. In cases in which one cue harms performance slightly (perhaps with low-frequency words at fast presentation rates), two cues should do more harm, and in

cases in which cues are helpful (high-frequency words at slow presentation rates), more cues should be more helpful. Words were studied in triples rather than in pairs, and zero, one, or two cues were available at test.

Method

Subjects. The subjects were 122 students at the University of California, Riverside, participating to fulfill an introductory psychology course requirement. They were run in groups of 1-5, with each session lasting about 50 min.

Procedure. The subjects were presented with eight study lists, each consisting of 36 word triples (ABC, DEF, GHI, . . . , etc.). Each list consisted of all high-frequency or all low-frequency word triples, presented at 1.0, 2.5, 4.0, or 5.5 sec/triple (which was constant within a given list). All words were sampled from Thorndike and Lorge's (1944) and Kučera and Francis's (1967) word frequency norms. High-frequency words were defined as having 50+ occurrences per million, and low-frequency words were defined as having less than 4 occurrences per million.

A recognition test followed a mental arithmetic task. All items used for distractors were of the same frequency as items from the study lists. Single, cued (with one or two cues), and associative trials were mixed together in a 32-trial test phase following each study list. For each test condition, four targets and four distractors were tested. The procedures for a given test trial are described below for single, cued, and associative recognition.

Single. A list item was selected from one of the three positions of a study triple, or a new word was presented. The subject was to indicate whether the word was on the list.

Cued. Cued tests consisted of one or two list words from a triple with either a list word from the same triple or a new word. Cued-1 tests were designated as AB or AX for target and distractor trials, respectively; Cued-2 tests were designated as ABC and ABX. For all cued tests, the subject's task was to say whether the designated test word was from the list. The position of the cue(s) and test item varied across the three positions with equal probability. For 70 subjects, a row of question marks identified the test word. For the remaining 52 subjects, the question marks were not used, and the cue word was marked by a line under it. This manipulation had no effect, and it is not considered in the data analysis. As in Experiment 1, the subjects were fully informed regarding the relation of the cue(s) to the test item, they were told that the cue would be useful in making recognition decisions, and they were encouraged to use them.

Associative. Each test consisted of an intact study triple (ABC) or a rearrangement of three list words from three different study triples (AEI). The subjects were to indicate whether each test triplet was intact or rearranged.

As in Experiment 1, each study list triple contributed to only one test trial. The subjects were prompted as to the nature of the test type immediately before each test trial. The position within the triple of old test items and cues was maintained between study and test. Test trials were self-paced and independent for each subject. The subjects were run in groups of 1-6.

Prior to data collection, the subjects were given detailed instructions and a practice study and test list containing all recognition test conditions.

The order of list presentation, order of test types within a test phase, and selection of words from the word pool and assignment to conditions were all randomly determined for each subject. Stimulus presentation and response collection were controlled by ATARI 1040ST computers, which ran independently for each subject.

Results

Overall analysis. Average hit and false alarm rates, plus hit rates minus false alarm rates (HR-FAR), are shown in Table 4 for each condition. An analysis of variance (ANOVA) computed on hit rates minus false alarm rates showed increases in performance with presentation time [$F(3,363) = 108.85$, $MS_e = .149$, $p < .0001$], a main effect of test type [$F(3,363) = 36.84$, $MS_e = .105$, $p < .0001$], and a very small statistically insignificant effect for word frequency [$F(1,121) = 3.08$, $MS_e = .106$, $p < .09$].

However, word frequency did interact with test type [$F(3,363) = 19.31$, $MS_e = .106$, $p < .0001$]. Recognition performance was higher for low-frequency than for high-frequency words for single-item [$F(1,121) = 28.39$, $MS_e = .106$, $p < .0001$] and for Cued-1 recognition [$F(1,121) = 10.31$, $MS_e = .105$, $p < .005$]. No word frequency effect was shown for Cued-2 recognition, and the word frequency effect was reversed for associative recognition [$F(1,121) = 19.16$, $MS_e = .124$, $p < .0001$]. Discussion of this word frequency interaction will shed light on other findings to be discussed; thus, further discussion of the interaction will be temporarily deferred.

Single-item-cued recognition comparisons. Clark and Shiffrin (1987) showed that old-new discrimination decreased slightly as cues were added (e.g., single-item > Cued-1 > Cued-2). Clearly, this trend was not replicated here. A separate ANOVA comparing single-

Table 4
Average Hit Rates and False Alarm Rates, Experiment 2

Presentation Time (seconds)	Single-Item			Cued-1			Cued-2			Associative		
	HR	FAR	HR-FAR	HR	FAR	HR-FAR	HR	FAR	HR-FAR	HR	FAR	HR-FAR
Low-Frequency Words												
1.0	.53	.30	.23	.53	.39	.14	.57	.38	.19	.62	.56	.06
2.5	.58	.26	.32	.64	.26	.38	.69	.24	.45	.67	.50	.17
4.0	.65	.25	.40	.68	.25	.43	.70	.21	.47	.75	.42	.33
5.5	.66	.19	.47	.73	.26	.47	.74	.20	.54	.73	.40	.33
<i>M</i>	.61	.25	.36	.65	.29	.36	.68	.26	.42	.69	.47	.22
High-Frequency Words												
1.0	.56	.45	.11	.53	.44	.09	.59	.35	.24	.59	.45	.14
2.5	.59	.38	.21	.60	.31	.31	.64	.25	.39	.69	.35	.34
4.0	.63	.30	.33	.64	.30	.34	.77	.25	.52	.68	.30	.38
5.5	.63	.30	.33	.70	.27	.43	.75	.22	.53	.73	.30	.43
<i>M</i>	.60	.36	.24	.62	.33	.29	.69	.27	.42	.67	.35	.32

item, Cued-1, and Cued-2 recognition showed that cuing helped for both high-frequency [$F(2,242) = 40.76$, $MS_e = .100$, $p < .0001$] and low-frequency words [$F(2,242) = 6.06$, $MS_e = .094$, $p < .003$]. It can be seen from Table 4 that for low-frequency words, collapsing over presentation rate, there was no difference between single-item and Cued-1 recognition. Thus, the overall cuing advantage was due to better recognition with two cues. For high-frequency words, the 0.048 advantage for Cued-1 over single-item recognition was not statistically significant [$F(1,121) = 1.38$, $MS_e = .100$]. However, the .13 advantage of Cued-2 over Cued-1 was significant [$F(1,121) = 24.40$, $MS_e = .100$, $p < .001$].

For high-frequency words, test type (excluding associative recognition) did not interact with presentation time [$F(6,726) = 1.10$, $MS_e = .105$, $p > .36$]. For low-frequency words the interaction was marginally significant [$F(6,726) = 1.99$, $MS_e = .097$, $p < .065$]. At the fastest presentation rate, for both high-frequency and low-frequency words, single-item recognition was better than Cued-1. Only for low-frequency words did this difference approach significance [$t(121) = 1.80$, $p < .08$].

Effects of study time. We wished to know (1) whether the rise in performance due to increasing presentation rate varied across the different test conditions, and (2) whether the effect of cues varied with either presentation rate or the level of associative recognition performance. Each of these will be discussed in turn.

Presentation rate did not interact with test type [$F(9,1,089) = 1.56$, $MS_e = .010$, $p > .10$], indicating that the rate at which performance increased with presentation rate did not vary across test types. The slopes of the functions relating performance to presentation time ranged from .051 to .072. The flattest slopes were for single-item recognition, which is consistent with the hypothesis that proportionally more associative information is stored with increased study time; however, the slope differences were extremely small and statistically unreliable.

For low-frequency words, associative recognition performance was much lower than for the other conditions at the fastest presentation rate, which was consistent with Experiment 1. However, in Experiment 2, associative recognition performance didn't "catch up" to the level of performance in the other conditions when presentation time was increased, as had occurred in Experiment 1. The pattern for high-frequency words was quite different from that for low-frequency words. At the 1-sec presentation rate, associative recognition was slightly better than single-item or Cued-1 recognition, and it maintained that advantage as presentation time increased.

Presentation time had only a very small effect on the relationship between cued and single-item recognition. Only at the fastest presentation rate was there evidence for a cuing disadvantage. For presentation rates longer than 1-sec/triple, the cued-single-item relationship was relatively constant.

Relationship between cuing and associative recognition. We also looked at the relationship between cuing effects and associative recognition. There was little evidence for any relationship. First, associative recognition performance increased steadily with increased presentation time, but the effect of cuing was little changed by increases in presentation rate. If one looks at mean levels of performance, the only evidence for a link between cuing effects and associative recognition performance was that the sole cuing deficit of just marginal reliability occurred under conditions that rendered associative recognition performance near chance.

A correlational analysis was also performed. For each subject, the differences in performance between Cued-1 and single-item recognition (Q1-S), and between Cued-2 and single-item recognition (Q2-S) were calculated by subtracting HR-FAR scores. Separate correlation coefficients were calculated, relating each of these to associative recognition (A) performance. These are denoted $r_{Q1,A}$ and $r_{Q2,A}$, for the one- and two-cue cases. If the cued-single-item relationship is mediated by associative information, then positive cuing should co-occur with high associative recognition scores, and both correlations should be positive. Sixteen separate correlations were calculated: $r_{Q1,A}$ and $r_{Q2,A}$ for high- and low-frequency words, and for each presentation rate. All of these correlations were near zero, except for the longest presentation time and high-frequency words where $r_{Q1,A} = .180$, and $r_{Q2,A} = .184$, which are both significantly different from zero [$t(120) = 1.99$ and 2.05 , respectively, $p < .05$]. The interpretation of significance must be done with caution, of course, since the omnibus test that *any* of the r s is different from zero fails to show a significant result $\chi^2(119) = 11.686$, $p > .10$].

One might cautiously conclude that there is limited evidence for a correlation between cuing effects and associative recognition based on: (1) the only marginally reliable cuing deficit occurred when associative recognition performance was near chance, and (2) for high-frequency words, $r_{Q1,A}$ and $r_{Q2,A}$ are significantly greater than zero. However, the bulk of other results that do not show a relationship between cuing effects and the level of associative recognition performance suggest that any relationship must be extremely weak—at least in the present study.

Discussion

Experiment 2 shows that the effect of adding cues on old-new recognition judgments is influenced by word frequency and by the number of additional word cues. For high-frequency words, old-new recognition increased as the number of cues increased. For low-frequency words, the increase in performance due to adding cues was minimal: adding a single cue had no effect (Cued-1 performance equaled single-item recognition performance), and the increase in performance with two cues was modest. That the effect of cuing depends on word frequency and the number of cues argues against the hypothesis that sub-

jects simply ignore the cue. For this hypothesis to be tenable, it must be argued that subjects selectively ignore the cue only for low-frequency words, only when the presentation rate is at least 2.5 sec, and only when one cue is used.

Contrary to what was found in Experiment 1, the increase in associative recognition performance did not increase disproportionately relative to that for item recognition. The result from Experiment 1 seems reliable, however, and it was replicated in two additional experiments done with word pairs (see Note 1). One reason for the discrepancy between Experiments 1 and 2 may be that the associative recognition task in Experiment 2 was considerably easier. Discrimination of ABC from AEI (in which the words in the rearranged triple come from three different study triples) is easier than discrimination of AB from AD. For the ABC-AEI discrimination, the AEI triple can be rejected on the basis of any one of the three pairwise connections.

Increases in study time also had very little effect on the relationship between cued and single-item recognition. Only at the fastest presentation rate did a cuing deficit occur. Beyond that, cuing effects were relatively constant with increases in presentation rate.

The evidence that cuing effects were correlated with associative recognition performance was also quite limited: the cuing deficit occurred only under conditions that rendered associative recognition performance marginally above chance. However, a correlational analysis showed very little evidence overall that the effect of cuing was correlated with the level of associative recognition performance. Moreover, for low-frequency words, there was no effect of adding a single cue, although associative recognition performance increased steadily with study time.

This pattern of results shows that the effect of adding cues, and the relationship between cuing effects and associative recognition, is not simple. The lack of a relationship between cuing effects and associative recognition performance suggests two separate associative factors. Moreover, the associative factor that underlies intact-rearranged discrimination in associative recognition may not be the same associative factor that produces cuing effects in old-new discrimination.

It may be, for example, that associations are represented at two levels. At one level, the association between A and B is a connection between two distinct memorial units; at another level, the A-B association is a single higher order unit. This bilevel representation of associative information was proposed earlier within the framework of the SAM model by Shiffrin et al. (1988).

Related to the hypothesis of higher order units is the hypothesis that a word's encoding depends on the other words that are presented and studied with it in the same group. To take an extreme example, the encoding of *diamond* would be different in the context of *ruby* than it would in the context of *baseball*. If such effects occur, then repeating a studied group at test would lead to simi-

lar encoding and hence better performance. Such a mechanism was incorporated in the model of Clark and Shiffrin (1987), to help account for higher than predicted performance when all items in a studied group were presented together at test.

Alternatively, the two aspects of associations may reflect differences in retrieval processes. Within the framework of a single-level model, either associative information may be accessed by global matching, or associative connections may be utilized in a recall-like memory search. Recognition models of this type, which combine recall-like retrieval processes with direct-access familiarity computations, have also been proposed (Atkinson & Juola, 1974; Humphreys, 1978; Mandler, 1980; Wolford, 1971).

The word frequency \times test type interaction is also consistent with a two-factor association model. Typical results show that high-frequency words are recalled better than low-frequency words, but that low-frequency words show better performance for item recognition (see Gregg, 1976, for a review). Word frequency interactions have been demonstrated previously by Clark (1992), who showed that a single-factor version of the SAM model could not simultaneously produce a low-frequency word advantage for item recognition and a high-frequency word advantage for associative recognition.

In Experiment 2, the low-frequency advantage for single-item old-new recognition disappeared as additional word cues were added, and it reversed to a high-frequency word advantage for associative recognition. The clear implication here is that there is an increasing associative contribution as cues are added, which by the nature of the task is maximized for associative recognition. On the basis of the independence of cuing effects and associative recognition, it would seem that there are at least two associative factors. How many associative factors are responsible for the shift in the frequency effect remains an open question.

An additional associative factor may have been induced by strategies that subjects used in the mixed testing procedure. This procedure may have induced them to adopt an "associative" strategy, since for any block of trials, three out of four test trials involved the presentation of word groups for which associative information was important. Such a strategy may have been nonoptimal for single-item recognition, and it may have produced a decrease in performance for that condition relative to the others. This factor might explain why cuing effects were almost always in the positive direction, contrary to the results in Experiment 1, and it would also be consistent with the finding that for high-frequency words, associative recognition was better than single-item recognition. In Experiment 3, performance in mixed conditions was directly compared with performance in blocked test conditions. If the pattern of results is the same for mixed and blocked testing, it can be assumed that the results of Experiment 2 were not due to subjects' adopting an associative strategy.

EXPERIMENT 3

Method

Subjects. The subjects were 130 undergraduates from the same population as that for Experiment 2.

Materials and Procedure. Eight study lists were presented, each consisting of 18 study triplets presented at a rate of 4.0 sec/triplet. Four lists consisted of high-frequency words, and four consisted of low-frequency words. Each study list was followed by the arithmetic distractor task that was used in Experiments 1 and 2. Each study list was then followed by a recognition memory test. The test probes were the same as those in Experiment 2 (single-item, Cued-1, Cued-2, associative).

For each subject, four study lists were followed by a blocked test phase and four by a mixed test phase. The blocked test condition corresponded to the test condition in Experiment 1; only one kind of discrimination was made within one block of test trials. The mixed test condition was like that in Experiment 2; the four different test types were mixed into one test sequence. As in Experiment 2, a prompt appeared just prior to each test trial, informing the subject of the test type.

In a mixed test phase, two target and two distractor trials were presented for each of the four test types, for a total of 16 tests. In the blocked test phase, eight target and eight distractor trials were presented for each test condition—except for associative recognition, which had four target and four distractor trials.

Results

Average hit rates and false alarm rates, as well as hit rates minus false alarm rates, are listed in Table 5 for each condition. Blocked testing led to better overall performance than did mixed testing [$F(1,294) = 3.92$, $MS_e = .278$, $p < .051$]. An overall main effect was also shown for test type [$F(3,294) = 12.59$, $MS_e = .050$, $p < .0001$]. There was not a main effect of word frequency [$F(1,98) = 0.18$, $MS_e = .072$]. However, the test type \times word frequency interaction was significant [$F(3,294) = 23.09$, $MS_e = .053$, $p < .0001$], replicating results from Experiment 2. No other interactions approached statistical significance.

For low-frequency words, cued recognition was only slightly better than single-item recognition, but the advantage was not statistically reliable, either for blocked [$F(2,98) = 1.24$, $MS_e = .035$] or for mixed [$F(2,98) =$

.15, $MS_e = .040$] testing conditions. For high-frequency words, the cuing advantage was reliable for both mixed [$F(2,98) = 6.59$, $MS_e = .035$, $p < .005$] and blocked [$F(2,98) = 9.20$, $MS_e = .052$, $p < .0005$] testing conditions. Even for high-frequency words, the increase in performance with one cue (Cued-1 – Single) was small, and not statistically significant, either for mixed [$t(49) = .30$] or for blocked [$t(49) = 1.57$] test conditions.

The word frequency \times test type interaction is a replication of the word frequency effect reversal shown in Experiment 2. Eight separate comparisons of high- and low-frequency words for mixed and blocked testing and for each test type showed the following for both mixed and blocked testing: a low-frequency word advantage for single-item and Cued-1 recognition, no difference due to word frequency for Cued-2 recognition, and a reversal of the word frequency effect—that is, a high-frequency word advantage for associative recognition. The statistical analyses are summarized in Table 5.

Discussion

Experiment 3 replicated the results of Experiments 1 and 2. For low-frequency words, there was no increase in recognition performance due to adding cues. However, for high-frequency words, recognition performance increased with the number of cues. Experiment 3 showed that the cuing \times word frequency interaction is not produced by an “associative” strategy induced by mixing test types within a block of test trials. It is also noteworthy that for high-frequency words the associative performance level was virtually equal to that of the Cued-2 condition, a finding that poses problems for most models (for example, SAM, TODAM, and the Matrix model all predicted too large a difference for the Experiment 1 data, and the present difference is considerably smaller).

GENERAL DISCUSSION

In the present research, we investigated the effects of adding cues on old-new item recognition by examining cued and single-item recognition, and we investigated the

Table 5
Average Hit Rates, False Alarm Rates and HR-FAR for Experiment 3

	Single-Item			Cued-1			Cued-2			Associative		
	HR	FAR	HR-FAR	HR	FAR	HR-FAR	HR	FAR	HR-FAR	HR	FAR	HR-FAR
Mixed												
LF	.66	.17	.49	.74	.20	.54	.74	.20	.54	.79	.47	.32
HF	.64	.24	.40	.67	.26	.41	.73	.18	.55	.78	.27	.51
<i>F</i>			4.02			8.17			.09			17.12
<i>p</i>			.06			.01			n.s.			.0001
Blocked												
LF	.77	.20	.57	.80	.20	.60	.82	.22	.60	.76	.40	.36
HF	.72	.29	.43	.72	.22	.50	.78	.18	.60	.79	.21	.58
<i>F</i>			3.19			5.76			.02			10.24
<i>p</i>			.001			.05			n.s.			.005

Note—*F* values with $df = 1,49$ and *p* values are for comparisons of HR – FAR for high-frequency (HF) and low-frequency (LF) words.

relation of associative recognition to other types of multiple-items tests. The relationship between cued, single-item, and associative recognition has important implications for current models of recognition.

The SAM model predicts only negative cuing effects, i.e., that cued recognition performance is always worse than performance for single-item recognition. Humphreys's (1978) ICM and Hintzman's (1984) MINERVA 2 predict positive cuing effects. For Murdock's (1982) TODAM and Pike's (1984) Matrix model, cuing effects covary with associative recognition performance.

Experiment 1 showed that for low-frequency words, cued recognition was slightly worse than single-item recognition for both fast and slow presentation rates, although the differences were small and statistically insignificant. This is consistent with previous findings of small, statistically unreliable cuing deficits (Clark & Shiffrin, 1987; Gillund & Shiffrin, 1984). The absence of cuing facilitation is clearly inconsistent with predictions of MINERVA 2 and the ICM, which both predict cuing advantages. The result also appears to be inconsistent with TODAM and the Matrix model, because the relationship between single-item and cued recognition remained constant while associative recognition performance showed large improvements with increased study time. The small negative cuing result is consistent with the SAM model's limited capacity assumption.

It is important to evaluate the models, not simply on the basis of the cued and single-item recognition data, but on that of the whole data set. Each of the models listed above was fit to the data of Experiment 1. TODAM gave the best fit, followed by the SAM model. The fit of TODAM was good, despite the fact that there was a constant cuing deficit, presumably because the magnitude of the deficit was small. The fits for the other models were substantially poorer. In general, models that make a clear distinction between item and associative information (TODAM, SAM, ICM) fit better than models that do not (MINERVA 2, Matrix model). One reason for this is that in Experiment 1 increases in performance with presentation time were much larger for associative recognition than for any of the other test conditions.

Two factors were examined in Experiment 2, presentation rate and word frequency. Overall, positive rather than negative cuing effects were shown. These positive cuing effects were larger for high-frequency words than for low-frequency words. For low-frequency words, the positive cuing effect was found only when two cues were used. A negative cuing effect occurred only for low-frequency words at the fastest presentation rate. Cuing effects were unrelated to the level of performance for associative recognition.

Experiments 2 and 3 yielded large effects of word frequency. Single-item and Cued-1 recognition showed higher performance for low-frequency words, which is consistent with the standard findings for old-new item recognition. Cued-2 recognition showed a null result for word frequency, and associative recognition showed an advantage

for high-frequency words. This pattern of results suggests that the contribution of associative factors increases as cues are added, and that it is maximized for associative recognition.

The overall pattern of results for Experiments 2 and 3 is obviously inconsistent with the original version of the SAM model, which predicts only cuing deficits. Beyond this, all the models have difficulty handling the cuing results, the single-item results, and the associative recognition results, taken together. In particular, associative recognition performance increases with presentation time, whereas cuing effects (null or positive) are relatively insensitive to presentation time. The results suggest that the factor that produces cuing advantages is not the factor that underlies associative recognition performance. One explanation may be that different kinds of associative components are operating.

There are at least three ways in which associative factors may be operating in the recognition tasks used in these experiments. One proposal is based on higher order associative units, another is based on the operation of recall-like retrieval processes, and the third posits encoding match factors. We did not apply the models to the data of Experiments 2 and 3, because it was clear that the results were complicated and problematic for all of the models. Three general ways of modifying the models are discussed below; the specifics of how to instantiate multiple associative components, however, remains to be worked out.

Higher Order Units

Two types of associative information may be stored: information linking two different stored traces, and information incorporating the members of a pair or a triple into a single, higher order unit (see Gronlund, 1986; Shiffrin et al., 1988).

In SAM, *linking* information (i.e., the *b* parameter) contributed to single-item, cued, and associative recognition; thus, strengthening linking information will improve performance in all conditions and will not lead to differential improvement for different conditions. However, adding higher order associative information will produce a cuing advantage, because the cost, if any, of processing the additional cue or of forming a joint cue may be small in comparison with the gain produced by the exact match of the test unit to the stored unit.

The following assumptions seem reasonable concerning the relation of associative units and word frequency. First, associative units are more likely to be stored for high-frequency word groups than for low-frequency word groups. Second, for word triples, only one associative unit is stored for each triple, if one is stored at all. And third, the partial match to the associative unit for a group of low-frequency words is very small, compared with a partial match for high-frequency words. These assumptions are consistent with the results of Experiments 2 and 3, which showed that cuing effects were quite small for low-frequency words and that the cuing advantage in Experi-

ment 2 for low-frequency words occurred only when two cues were given, but not when only one cue was given.

TODAM, MINERVA 2, and the Matrix model store associations in terms of a single representation, rather than as a link between items. Associations are represented as a single convolution in TODAM, as a single vector in MINERVA 2, and as a single matrix in the Matrix model. These representations are similar to the higher order unit assumption in SAM. However, only in TODAM is the associative information distinct from the item-specific information. TODAM also differs from the higher order unit SAM model in that SAM has dual representation of associations: as separate units, and as links between items. Adding dual representations to TODAM and the Matrix model might enable these models to predict the small differences between cued and associative conditions that were found for high-frequency words.

The dual nature of associations might also reflect a difference in level of processing: associative information may be represented both graphemically and semantically. Subjects may discriminate intact pairs from rearranged pairs on the basis of such factors as the number of spaces between words, as well as the overall shape of the pair, even if the meanings of the words are unknown. This kind of associative information may play a relatively larger role for low-frequency than for high-frequency words.

Recall Processes in Recognition

This hypothesis involves the possibility of recall's being used to supplement global recognition processes. Retrieval of associative information may involve a search process similar to that proposed by Raaijmakers and Shiffrin (1980) for recall. This assumption is central to Humphreys's (1978) ICM and a similar model proposed by Mandler (1980). This proposal is particularly tempting in light of the word frequency interaction shown in Experiments 2 and 3. The standard word frequency effect reverses for associative recognition (performance was better for high-frequency words), a pattern normally found for recall tasks. Also, the cuing effects in Experiments 2 and 3 were very small for low-frequency words.

The fact that cuing advantages in Experiments 2 and 3 were very small for low-frequency words is consistent with a familiarity-plus-search model: additional cues facilitate recognition performance mainly for high-frequency words, because high-frequency words are better recall cues than low-frequency words. For low-frequency words, the cues are not helpful and the resources given to the cue are wasted. The same hypothesis helps explain the failure to obtain cuing advantages in Experiment 1 (in which low-frequency words were used). Given this, the increase in associative recognition performance with presentation time must be due to some factor other than recall.

Encoding Match

Another proposal that must be considered has been termed the encoding-match or change-of-meaning hypoth-

esis (Light & Carter-Sobell, 1970). Words tested in a verbal context that is different from their study context may be encoded differently than they were at study, and hence they will be less familiar than words tested in an intact verbal context. Reder, Anderson, and Bjork (1974) have proposed that encoding-match effects are larger for high-frequency words than for low-frequency words, since they have more meanings. This differential encoding match hypothesis is consistent with the present result that negative cuing occurs for low-frequency words and that positive cuing occurs for high-frequency words, and with the high-frequency word advantage for associative recognition. Explanations based on encoding match are not associative in the way that the higher order unit and recall hypotheses are. For cued recognition, the cue and the test item do not combine to form an association, but rather the cue biases the encoding of the test item. Thus, the effect is at the time of encoding rather than at retrieval.

Evidence from similar paradigms argues against the sufficiency of encoding match as an explanation of cuing effects. Specifically, Humphreys (1976) and Humphreys and Bain (1983) required subjects to make old-new decisions for the individual words in intact and rearranged test pairs. Encoding match factors operate at the level of words, not associations; thus double-miss rates are predicted to be lower for intact than for rearranged test pairs (see Humphreys and Bain for details of the argument). Contrary to this prediction, they have shown that double-miss rates are equal for intact and rearranged test pairs. Moreover, Clark (1992) has shown that this equality is maintained for both high- and low-frequency words.

Summary

It is possible, if not likely, that each of the three possible mechanisms may contribute to the pattern of results from the present studies. Higher order units, recall during recognition, and matching of encodings at study and test may all contribute to cuing advantages and interactions of these with word frequency effects.

In addition, various kinds of strategies may be involved, some of which may be quite complex. For instance, in some cases, subjects may initially probe with item information only, and if that does not allow discrimination, they may then probe with associative information (Atkinson & Juola, 1974; Humphreys & Bain, 1991). Further investigation regarding the contribution of these factors in recognition is reserved for future research.

REFERENCES

- 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: Vol. 1. Learning, memory, and thinking* (pp. 242-293). San Francisco: W. H. Freeman.
- CLARK, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, *20*, 231-243.
- CLARK, S. E., & SHIFFRIN, R. M. (1987). Recognition of multiple-item probes. *Memory & Cognition*, *15*, 367-378.
- DAPALITO, F., BARKER, D., & WIAIT, J. (1972). The effects of con-

- textual changes on component recognition. *American Journal of Psychology*, **85**, 431-440.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recall and recognition. *Psychological Review*, **91**, 1-67.
- GREGG, V. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall & recognition* (pp. 183-216). London: Wiley.
- GRONLUND, S. D. (1986). *Multi-level storage and retrieval: An empirical investigation and theoretical analysis*. Unpublished doctoral dissertation, Indiana University.
- HINTZMAN, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, **16**, 96-101.
- HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.
- 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. (1983). Recognition memory: A cue and information analysis. *Memory & Cognition*, **11**, 583-600.
- HUMPHREYS, M. S., & BAIN, J. D. (1991). Episodic and strength components in recognition memory. In W. Hockley & S. Lewandowsky (Eds.) *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 309-330). Hillsdale, NJ: Erlbaum.
- HUMPHREYS, M. S., PIKE, R., BAIN, J. D., & TEHAN, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, **33**, 36-67.
- KINTSCH, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 333-374). New York: Academic Press.
- KUČERA, H., & FRANCIS, W. N. (1967). *A computational analysis of present-day American English*. Providence, RI: Brown University Press.
- 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.
- 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. (1989). Learning in a distributed memory model. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium on Cognition* (pp. 69-106). Hillsdale, NJ: Erlbaum.
- MURDOCK, B. [B., JR.], & LAMON, M. (1988). The replacement effect: Repeating some items while replacing others. *Memory & Cognition*, **16**, 91-101.
- MURDOCK, B. B., JR., & OGILVIE, J. C. (1968). Binomial variability in short-term memory. *Psychological Bulletin*, **70**, 256-260.
- PIKE, R. (1984). A comparison of convolution and matrix distributed memory systems. *Psychological Review*, **91**, 281-294.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and memory* (Vol. 14, pp. 207-262). New York: Academic Press.
- REDER, L. M., ANDERSON, J. R., & BJORK, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, **102**, 648-656.
- SHEU, C.-F. (1990). *A note on the multiple-trace memory model without simulation*. Unpublished manuscript.
- SHIFFRIN, R. M., MURNANE, K., GRONLUND, S. D., & ROTH, M. (1988). On units of storage and retrieval. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium on Cognition* (pp. 25-68). Hillsdale, NJ: Erlbaum.
- SHIFFRIN, R. M., RATCLIFF, R., & CLARK, S. E. (1990). The list strength: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 179-195.
- SLAMECKA, N. J. (1975). Intralist cueing of recognition. *Journal of Verbal Learning & Verbal Behavior*, **14**, 630-637.
- 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.
- TULVING, E., & THOMSON, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, **87**, 116-124.
- UNDERWOOD, B. J. (1974). The role of the association in recognition memory. *Journal of Experimental Psychology Monographs*, **102**, 917-939.
- WEBER, E. U. (1988). Expectation and variance of item resemblance distributions in a convolution-correlation model of distributed memory. *Journal of Mathematical Psychology*, **32**, 1-43.
- WOLFORD, G. (1971). Function of distinct associations for paired-associate performance. *Psychological Review*, **78**, 303-313.

NOTES

1. Two other minor variations on Experiment 1 were also carried out. The pattern of results was essentially the same for all three cases, so only Experiment 1 is reported. In one of the unreported variations, the cue in cued recognition was on the left on half of the trials and on the right for the other half (and the test item was identified with question marks above it). In the second variation, fast and slow presentation rates were varied at .75 and 8.0 sec/pair. The patterns of results for these two experiments were essentially identical to those of Experiment 1, indicating that the results are very reliable.

2. The SAM model was originally developed as a search model of recall by Raaijmakers and Shiffrin (1980). The recognition model does not involve a search mechanism, but takes the same name as a matter of theoretical heritage.

3. The term *cue* is used throughout both in a generic sense and in the specific sense of referring to the cue in cued recognition test pairs. The dual usage of the term is unavoidable. We will try to avoid ambiguous use of the term as much as possible.

4. We have also looked at a model in which the single-item test vectors are weighted by γ_i and the convolution vector is weighted by γ_a prior to summing and probing. The difficulties in handling the cued disadvantage seem only to increase.

There are various versions of TODAM that were not applied to the data. In particular, there are two aspects of the model that we did not investigate. First, Weber (1988) introduced weights at both storage and retrieval, which could vary independently. Making this assumption would certainly have produced a better fit, but it probably would also have resulted in a rather unconstrained version of the model with many additional free parameters. Second, effects of presentation time could also have been modeled in terms of probabilistic encoding (Murdock, 1989; Murdock & Lamon, 1988). Increasing vector weights γ_i and γ_a should produce similar results, and in addition, it allows storage to vary independently for item and associative information, an aspect that seems crucial for the present data.

APPENDIX A
Expectations and Variances for SAM

Equations are given for the expectations and variances for targets and distractors for each of the four test conditions. In the equations, the following parameters are used: a , b , c , and d are context, interitem, self-, and residual retrieval strengths. Cue weights are as follows: w_c for context; w_s for the test item in single-item recognition, w_p for each item in a double-item test (pair, associative); w_q and w_t are the weights on the cue and test item in cued recognition. The $\theta(w)$ terms are adjustment factors, reflecting the fact that $E[X^2] \neq E[X]^2$. We assume that for any strength with mean M , the probability that the true strength equals $y_i M$ is p_i . That is, $P[S = y_i M] = p_i$. $\theta(w)$ is then defined: $\theta(w) = \sum y_i^w p_i$, for any w .

Given the variance assumptions from the text, θ terms are calculated as follows: $\theta(w) = [(1+\nu)^w + 1 + (1-\nu)^w]/3$, where $\nu =$

5. The α terms in the text are then calculated from the θ terms that appear in the equations below.

For single-item recognition, target and distractor expectations are

$$E[A] = \theta(w_c)\theta(w_s)a^{w_c}[b^{w_s}+c^{w_s}+(N-2)d^{w_s}], \quad (A1)$$

$$E[X] = \theta(w_c)\theta(w_s)Na^{w_c}d^{w_s}, \quad (A2)$$

and target and distractor variances are

$$\text{VAR}[A] = a^{2w_c}[\theta(2w_c)\theta(2w_s)-\theta^2(w_c)\theta^2(w_s)] \cdot [(N-2)d^{2w_s}+b^{2w_s}+c^{2w_s}], \quad (A3)$$

$$\text{VAR}[X] = [\theta(2w_c)\theta(2w_s)-\theta^2(w_c)\theta^2(w_s)]Na^{2w_c}d^{2w_s}. \quad (A4)$$

For cued recognition, expectations for target and distractor are

$$E[AB] = \theta(w_a)\theta(w_q)\theta(w_i)a^{w_c}[b^{w_a}c^{w_i}+c^{w_a}b^{w_i}+(N-2)d^{w_a}d^{w_i}], \quad (A5)$$

$$E[AX] = \theta(w_c)\theta(w_q)\theta(w_i)a^{w_c} [b^{w_a}d^{w_i}+d^{w_i}c^{w_a}+(N-2)d^{w_a}d^{w_i}]. \quad (A6)$$

The variances are

$$\text{VAR}[AB] = \psi_a a^{2w_c}[(N-2)c^{2w_a}b^{2w_i}+b^{2w_a}c^{2w_i}+d^{2w_a}d^{2w_i}], \quad (A7)$$

$$\text{VAR}[AX] = \psi_a a^{2w_c}[(N-2)c^{2w_a}d^{2w_i}+b^{2w_a}d^{2w_i}+d^{2w_a}d^{2w_i}], \quad (A8)$$

where

$$\psi_a = \theta(2w_c)\theta(2w_q)\theta(2w_i) - \theta^2(w_c)\theta^2(w_q)\theta^2(w_i). \quad (A9)$$

For pair recognition,

$$E[AB] = \theta(w_c)\theta^2(w_p)a^{w_c}[2b^{w_p}c^{w_p}+(N-2)d^{2w_p}], \quad (A10)$$

$$E[XY] = \theta(w_c)\theta^2(w_p)a^{w_c}Nd^{2w_p}, \quad (A11)$$

$$\text{VAR}[AB] = \psi_p a^{2w_c}[2c^{2w_p}b^{2w_p}+(N-2)d^{4w_p}], \quad (A12)$$

$$\text{VAR}[XY] = \psi_p a^{2w_c}Nd^{4w_p}, \quad (A13)$$

where

$$\psi_p = \theta(2w_c)\theta^2(2w_p) - \theta^2(w_c)\theta^4(w_p). \quad (A14)$$

And for associative recognition, $E[AB]$ and $\text{VAR}[AB]$ are the same as listed above for pair recognition.

$$E[AD] = \theta(w_a)\theta^2(w_p)a^{w_c}[b^{w_p}d^{w_p}+c^{w_p}d^{w_p}+(N-4)d^{2w_p}], \quad (A15)$$

$$\text{VAR}[AD] = \psi_p a^{2w_c}[(2c^{2w_p}d^{2w_p}+b^{2w_p}d^{2w_p}+(N-4)d^{4w_p})]. \quad (A16)$$

APPENDIX B

Expectations and Variances for TODAM

Each pair AB adds $\gamma_i \mathbf{a} + \gamma_a \mathbf{b} + \gamma_a (\mathbf{a} * \mathbf{b})$ to the memory vector \mathbf{M} . Single-item tests probe with the single test vector, cued tests probe with the convolution plus the test vector, pair tests probe with the convolution plus both test vectors, and associative tests probe with the convolution vector.

In the equations, γ_i and γ_a are weights on item and associative information, K is the number of pairs, and N is the item vector length.

For all test conditions, the expectations for distractors are zero: $E[X] = E[AX] = E[XY] = E[AD] = 0$. For single-item recognition,

$$E[A] = \gamma_i. \quad (B1)$$

The variances for target and distractor are

$$\text{VAR}[A] = A + \left(\frac{1}{N} \gamma_i^2 + \frac{N+1}{N^2} \gamma_a^2 \right), \quad (B2)$$

$$\text{VAR}[X] = A, \quad (B3)$$

where

$$A = \left[\frac{2}{N} \gamma_i^2 + \frac{3N^2+1}{4N^3} \gamma_a^2 \right] K. \quad (B4)$$

For cued recognition,

$$E[AB] = \gamma_i + \gamma_a, \quad (B5)$$

$$\text{VAR}[AB] = \text{VAR}[A] + B + \left(\frac{2N+2}{N^2} \gamma_i^2 + \frac{4N+3}{N^2} \gamma_a^2 \right), \quad (B6)$$

$$\text{VAR}[AX] = A + B + \frac{N+1}{N^2} (\gamma_i^2 + \gamma_a^2), \quad (B7)$$

where

$$B = \left[\frac{3N^2+1}{2N^3} \gamma_i^2 + \frac{2N^2+1}{3N^3} \gamma_a^2 \right] K. \quad (B8)$$

For pair recognition,

$$E[AB] = 2\gamma_i + \gamma_a, \quad (B9)$$

$$\text{VAR}[AB] = \text{VAR}[AB] + \text{VAR}[A], \quad (B10)$$

$$\text{VAR}[XY] = 2A + B. \quad (B11)$$

For associative recognition,

$$E[AB] = \gamma_a, \quad (B12)$$

$$\text{VAR}[AB] = \text{VAR}[AB] - \text{VAR}[A], \quad (B13)$$

$$\text{VAR}[AD] = B + \frac{2N+2}{N^2} (\gamma_i^2 + \gamma_a^2). \quad (B14)$$

APPENDIX C

Expectations and Variances for Matrix Model

Each item is stored as a vector with N elements. A word pair AB is represented as the matrix given by vector multiplication of the \mathbf{a} and \mathbf{b} vectors. Memory is represented as the sum of K such matrices representing the K pairs. Recognition is based on the match of the test item to the memory matrix \mathbf{M} , given by the matrix dot product of the test matrix \mathbf{T} with the memory matrix \mathbf{M} .

The expectation of the match of a given cell in \mathbf{M} to the corresponding cell in \mathbf{T} is μ and the standard deviation of this match is σ , both of which may vary as free parameters. Assuming that the cell match distribution is normal in form, the second, third, and fourth raw moments are

$$m_2 = \mu^2 + \sigma_2.$$

$$m_3 = \mu^3 + 3\sigma^2\mu,$$

$$m_4 = 3\sigma^4 + 6\sigma^2\mu^4.$$

From these, expectations and variances of familiarity distributions for the single-item, cued, pair, and associative recognition can be derived. The equations for the variances are quite complicated. For the sake of simplicity, all of the equations for expectations are given first, followed by the treatment of variances. Expectations and variances are identical for targets for cued, pair, and associative recognition. Thus, expectations are given for A, X, AB, AX, XY, and AD.

$$E[A] = n^2 m_2 \mu + (K-1)n^2 \mu^3, \quad (C1)$$

$$E[X] = Kn^2 \mu^3, \quad (C2)$$

$$E[AB] = n^2 m_2^2 + (K-1)n^2 \mu^4, \quad (C3)$$

$$E[AX] = n^2 m_2 \mu^2 + (K-1)n^2 \mu^4, \quad (C4)$$

$$E[XY] = Kn^2 \mu^4, \quad (C5)$$

$$E[AD] = 2n^2 m_2 \mu^2 + (K-2)n^2 \mu^4. \quad (C6)$$

To obtain variances for familiarity of each test probe, variances and covariances must be found for the match of the test matrix to each pair matrix in the summed memory matrix. The total variance is the sum of the variances and covariances of the separate matches of the test matrix to each of the individual pair matrices. These component variances and covariances are the V and C terms in the equations listed below.

$$\text{VAR}[A] = V_s + (K-1)v + 2(K-1)C_s + (K-1)(K-2)C, \quad (C7)$$

$$\text{VAR}[X] = Kv + K(K-1)C, \quad (C8)$$

$$\text{VAR}[AB] = V_2 - (K-1)V_0 + 2(K-1)C_2 + (K-1)(K-2)C_0, \quad (C9)$$

$$\text{VAR}[AX] = V_1 + (K-1)V_0 + 2(K-1)C_1 + (K-1)(K-2)C_0, \quad (C10)$$

$$\text{VAR}[XY] = KV_0K(K-1)C_0, \quad (C11)$$

$$\text{VAR}[AD] = 2V_1 + (K-2)V_0 + 2C_2 + 4(K-2)C_1 + (K-2)(K-3)C_0. \quad (C12)$$

The V and C terms are defined as follows: V and V_s are variances for nonmatching and matching single vectors, and C and C_s are the corresponding covariances. For double-item tests, V_0 , V_1 , and V_2 are variances for zero, one, and two matches, and C_0 , C_1 , and C_2 are the corresponding covariances. Equations for these variance and covariance terms are listed as follows:

$$V = n^2(m_2^2 - \mu^4)(m_2 - \mu^2) + n^3[(m_2^2 - \mu^4)\mu^2 + (m_2 - \mu^2)\mu^4], \quad (C13)$$

$$V_s = [nm_4 + n(n-1)m_2^2][nm_2 + n(n-1)\mu^2] - n^4 m_2^2 \mu^2, \quad (C14)$$

$$V_0 = n^2(m_2^2 - \mu^4)^2 + 2n^3[(m_2^2 - \mu^4)\mu^4], \quad (C15)$$

$$V_1 = n^2(m_4 - m_2^2)(m_2^2 - \mu^4) + n^3[(m_4 - m_2^2)\mu^4 + (m_2^2 - \mu^4)m_2^2], \quad (C16)$$

$$V_2 = n^2(m_4 - m_2^2)^2 + 2n^3[(m_4 - m_2^2)m_2^2]. \quad (C17)$$

And for covariances:

$$C = n^3 \mu^4 (m_2 - \mu^2), \quad (C18)$$

$$C_s = n^3 \mu^3 [m_3 - m_2 \mu], \quad (C19)$$

$$C_0 = n^2 \mu^4 (m_2 + \mu^2) [m_2 + \mu^2 (2n - 1)], \quad (C20)$$

$$C_1 = \mu^3 \{ n^2 (m_2 - \mu^2) (m_3 - m_2 \mu) + n^3 [(m_3 - m_2 \mu) \mu^2 + (m_2 - \mu^2) m_2 \mu] \}, \quad (C21)$$

$$C_2 = n^2 \mu^2 (m_3 - m_2 \mu) [m_3 + m_2 \mu (2n - 1)]. \quad (C22)$$

(Manuscript received September 10, 1991;
revision accepted for publication January 10, 1992.)