An analysis of interference in recognition memory

NANCY L. BOWLES and MURRAY GLANZER
New York University, New York, New York 10003

A stimulus-sampling model of recognition memory is presented that predicts both proactive and retroactive interference. To test the predictions of the model, a recognition memory experiment was carried out using a standard proactive-retroactive design with a forced-choice task. Both accuracy and latency were measured. The data showed, as predicted, equal proactive and retroactive effects on accuracy, providing solid support for the model. The interference effects are interpreted in terms of the model as arising from an increase in indirect marking, the marking of shared stimulus elements in words other than the study word. The model has two parameters representing the rate of indirect marking for high-frequency and low-frequency words and two parameters reflecting the efficiency of direct marking. The latency results completely paralleled the accuracy findings, showing proactive and retroactive effects. A possible extension of the model to handle latencies is considered.

On an intuitive level, there is considerable agreement among recognition theorists on the source of interference effects in recognition memory. While it is couched in different language, the fundamental concept proposed is a single one: Interference arises out of overlapping elements in study-list and test-list items. Underwood and his associates (e.g., Underwood & Freund, 1970) propose that study items elicit implicit associational responses. Such responses result from elements of the study items shared with other items. Shepard (1961), in his trace model of recognition, postulates that there is some probability of diffusion of trace elements from presented stimuli to other related stimuli. Bower's (1972) stimulus-sampling model of the recognition process has a parameter that represents the amount of overlap between stimulus elements of old and new items. Glanzer and Bowles (1976) use an indirect marking parameter to refer to the tagging of overlapping elements in unpresented stimuli in their stimulus-sampling model of recognition.

Models such as these provide the means to identify the locus of interference effects by offering mechanisms of the underlying memory processes. In particular, the
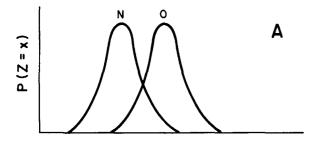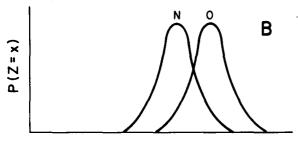
Glanzer and Bowles (1976) model makes strong predictions concerning interference effects. It will be presented here in detail.

A MODEL FOR INTERFERENCE IN RECOGNITION MEMORY

Glanzer and Bowles (1976) proposed a stimulus-sampling model of the recognition process to account for the word-frequency effect in recognition memory. This model is based on Bower's (1972) model, which was derived from Estes' (1955) stimulus fluctuation theory. The model assumes that a stimulus is a set of stimulus elements. In the present discussion, the stimuli will be words. When a word is shown on a study list, a subset of its elements is sampled, and the elements in that subset are directly marked. Some shared elements from other related words, which were not presented, may also be sampled and thus marked indirectly for those words. At test time, another sample of elements is taken. The decision as to whether the word is old or new depends on the number of directly and indirectly marked elements in the sample. The number of marked elements in a sample, Z, is a random variable that may be represented as in Figure 1A. The figure shows hypothetical distributions of the number of marked elements in a sample for old words that were shown on the study list and for new words that were not on the study list. In a forced-choice task, the subject chooses the word that has the largest number of marked elements in its sample. The model has parameters reflecting the rate of indirect marking, the size of the subset sampled, and the number of stimulus elements associated with each class of words.

The original Bower (1972) model included a parameter that reflected the fluctuation of stimulus elements

Figure 1. Theoretical distributions of the number of marked elements. Panel A shows the placement of the distributions for the control (short-list) conditions. Panel B gives the placement of the distributions for the experimental (long-list) conditions. Abbreviations: N = new words; O = old words.

between the subset of stimulus elements that were sampled and the remaining unsampled stimulus elements. Over time, the sampled elements would be distributed throughout the total set of elements; that is, the probability of drawing an element that had been drawn before would reach some asymptotic level. It is assumed in the present application that, with an average lag of 120 items between study and test, this asymptote has been reached. The fluctuation parameter is not included in this discussion.

The proportion of marked elements in an old word is specified as

$$p(O) = p(N) + [1 - p(N)](s/S), \qquad (1)$$

where $p(N)$ is the proportion of indirectly marked elements for a new word, s is the number of elements sampled when a word is presented, a constant, and S is the total number of stimulus elements associated with each member of a given class of words. (S has been shown to be greater for high-frequency words than for low-frequency words; Glanzer & Bowles, 1976.)

The subject's decision in a recognition task is determined by the number of marked elements sampled at test time. The sampling process here is a Bernoulli process, and the distribution of outcomes can be described by the binomial distribution. The probability of obtaining x marked elements in a sample of size s for word A is

$$P(Z=x) = \binom{s}{x} [p(A)]^x [1 - p(A)]^{s-x}, \qquad (2)$$

where $p(A)$ is the proportion of marked elements for a word, A.

It is assumed that in a forced-choice recognition memory test the subject chooses the word with the larger number of marked elements. In the case of a tie, the subject resamples. The resampling continues until two different values are found for A and B and a choice can be made. The choice can be expressed as

$$P(A,B) = P(Z_A > Z_B) + P(Z_A = Z_B)P(Z_A > Z_B)$$

$$+ P(Z_A = Z_B)^2 \cdot P(Z_A > Z_B)$$

$$+ P(Z_A = Z_B)^3 \cdot P(Z_A > Z_B) + \dots,$$

$$= P(Z_A > Z_B)\sum_{i=1}^{\infty} [P(Z_A = Z_B)]^{i-1},$$

where Z_A is the number of marked elements in the sample for word A and Z_B is the number of marked elements for word B. This is an infinite series which has the solution

$$P(A,B) = \frac{P(Z_A > Z_B)}{1 - P(Z_A = Z_B)}. \qquad (3a)$$

Assuming that Z_A and Z_B are binomially distributed, then

$$P(A,B)$$

$$= \frac{\sum_{r=1}^{s}\left\{\sum_{x=r}^{s}\binom{s}{x}[p(A)]^x[1-p(A)]^{s-x}\binom{s}{x-r}[p(B)]^{x-r}[1-p(B)]^{s-x+r}\right\}}{1 - \left\{\sum_{x=0}^{s}\binom{s}{x}[p(A)]^x[1-p(A)]^{s-x}\binom{s}{x}[p(B)]^x[1-p(B)]^{s-x}\right\}}.$$

$$(3b)$$

The extension of the model to handle interference effects is straightforward. Up to this point, there has been no discussion of the number of list items a subject has seen at the time a new or old item is viewed. That number did not play a critical role in the aspects of the theory discussed thus far. In order to discuss interference effects, this factor must be made explicit. The notation will be extended such that $p(O,n)$ and $p(N,n)$ represent the proportion of marked elements in an old word and a new word, respectively, after n items have been presented for study. For a new word, $p(N,n)$ is the proportion of indirectly marked elements after n study items. If it is assumed that indirect marking continues at the same rate over the next n trials, then the proportion of indirectly marked elements after 2n study trials will be

$$p(N,2n) = p(N,n) + [1 - p(N,n)] \, p(N,n). \qquad (4)$$

Equation 4 can be used to estimate the proportion of indirect marking for the interference conditions from that for the control conditions.

Interference is expressed as an increase in the amount of indirect marking. This increase results in a shift in the location of the distributions of marked elements, as depicted in Figure 1B. It can be demonstrated that under the assumptions of the model, the shift in the distributions results in a reduction in the difference between the mean of the distribution for old words and the mean of the distribution for new words. From Equation 1, $p(O,n) - p(N,n) = \alpha - \alpha p(N,n)$ represents the difference (D_1) between the means of the old and new distributions for a list of length n, where $\alpha = s/S$ and represents the direct marking rate. From Equations 1 and 5, $p(O,2n) - p(N,2n) = \alpha - 2\alpha p(N,n) + \alpha p^2(N,n)$ is the difference (D_2) between the means of the old and new distributions for a list of length 2n. The change in the differences for the short list and the long list ($D_1 - D_2$) is $\alpha p(N,n) - \alpha p^2(N,n)$, which will be positive for $0 < p(N,n) < 1$. This reduction in the size of the difference between the means results in reduced discriminability between old and new words. Thus, interference, arising from an increase in list length, will lead to a decrease in recognition performance.

Discriminability is typically considered to be a function not only of the difference between the means of the underlying theoretical distributions, but also of their variances. In the familiar framework of a signal detection theory, which is based on the assumption of normal distributions, d' is the measure of discriminability. A measure of discriminability analogous to d' can be calculated for binomial distributions. This measure can be expressed as

$$d'_b = [Np(old) - Np(new)] / \sqrt{Np(new)q(new)}, \qquad (5)$$

where Np(old) is the mean number of marked elements for the distribution of old words, Np(new) is the mean number of marked elements for the distribution of new words, and Np(new)q(new) is the variance of the distribution of new words. For a binomial distribution, as p increases from 0 to .50, the variance increases. Since $p(N,2n)$ will be greater than $p(N,n)$, the variance will be greater for the interference conditions than the control conditions, assuming that $p(N,2n)$ is less than .50. An increase in the variance results in a decrease in

the measure of discriminability. Thus, according to the present model, both the reduced difference between the means of the underlying theoretical distributions and the increased variance will contribute to reduced discriminability under interference conditions.

A second prediction from the model is that proactive interference effects and retroactive interference effects will be equal. Interference effects in the model are due to an increase in indirect marking. The model predicts that it makes no difference whether the additional indirect marking occurs before a given word is studied (proactive interference) or after it is studied (retroactive intereference). It can be demonstrated that the difference between the means of the old and new distributions will be the same for the proactive and retroactive conditions.

The assumptions of the model result in two strong predictions concerning interference in recognition memory: (1) There will be interference in recognition memory experiments, and (2) proactive effects and retroactive effects will be equal. In terms of the present model, interference is expected to result from the indirect marking process, which leads to a reduction in the difference between the means of the distributions of old and new items and to an overall increase in the number of marked elements, Z, for both distributions (Figure 1B). It does not matter whether the indirect marking occurs before or after a particular target word is studied. Therefore, both proactive and retroactive interference are predicted. The present experiment was designed to test these predictions.

Response latencies were collected as well as accuracy data in order to examine interference effects on this measure. Although the existing model does not handle response latencies, a possible extension of the model to predict latencies will be considered later.

METHOD

Design

A standard proactive-retroactive design, illustrated in Table 1, was used. Subjects in the proactive control condition studied 120 target words and were tested on those words. Subjects in the proactive experimental condition saw 240 words, 120 filler words followed by 120 target words, and were then tested. In the retroactive control condition, subjects were given 120 target words and a counting task and were then tested. The counting task equated the time between study and test for the retroactive control and experimental conditions. The retroactive experimental condition was exactly like the proactive experimental condition except that subjects were tested only on

Table 1
Experimental Design

	Proactive				Retroactive		
	Study List				Study List		
Condition	Filler	Target	Test List	Condition	Target	Filler	Test List
Control		120 Words	120 Pairs	Control	120 Words	120 Numbers	120 Pairs
Experimental	120 Words	120 Words	120 Pairs	Experimental	120 Words	120 Words	120 Pairs

the first 120 words and the second 120 served as fillers. Within each condition, proactive and retroactive, the average study-test interval was the same for control (short-list) and experimental (long-list) groups. The length of the test list was the same for all four conditions: 120 test pairs. It can be assumed, then, that any difference between a control group and its experimental group is due to the difference in the number of study words processed.

Subjects

Eighty New York University students participated in the experiment. The students received either credit in their introductory psychology course or $2.50 for their participation. All subjects had been speaking English since they were in the first grade of school or earlier.

Materials

Ten lists of 60 words each were prepared. Five of the lists were made up of words of high normative frequency with Kučera-Francis (1967) counts of 40 occurrences per million words or higher. The ranges and the distributions of frequency were balanced across lists. The other five lists were made up of low-frequency words with Kučera-Francis counts of 5 occurrences per million words or less. All 10 lists were matched for word length, with words varying from 3 to 10 letters. All lists were matched for number of abstract and concrete words. Study lists consisted of half high-frequency words and half low-frequency words randomly mixed. Both study and test lists were randomized individually for each subject. Assignment of lists to conditions was balanced across subjects so that every list served equally often to furnish target, distractor, and filler items in the experimental and control conditions.

For each subject, words were randomly assigned to one of six types of test pair: HO-HN, HO-LN, LO-HN, LO-LN, LO-HO, and HN-LN, where H represents a high-frequency word, L a low-frequency word, O an old (target) word, and N a new (distractor) word. (For example, HO-LN represents a high-frequency old word paired with a low-frequency new word.) The first four types of pairs are the standard test pairs, consisting of an old word from the study list and a new word. The last two types of pairs are called "null pairs." These pairs were either two old words of different frequency (both from the study list) or two new words of different frequency (neither on the study list). Performance on the standard test pairs reflects the differences between the means of the old and new theoretical distributions; the null choices reflect differences between the means of the high-frequency and low-frequency distributions (see Figure 2). The use of the null pairs provides an important check on the internal consistency of the model.
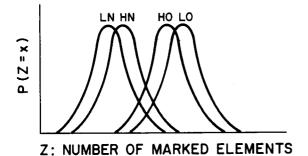

Figure 2. Theoretical distributions of the number of marked elements for high- and low-frequency words. Abbreviations: LN = low-frequency new words; HN = high-frequency new words; HO = high-frequency old words; and LO = low frequency old words. (Based on parameter values from Glanzer & Bowles, 1976.)

Procedure

Each subject saw a series of words presented one at a time on a CRT screen. A word remained on the screen for 2 sec, with 500 msec of blank screen between words. Stimulus presentation and response collection were controlled by computer.

The recognition test was a two-alternative forced-choice procedure in which two words appeared on the screen one above the other. Subjects pressed the upper or lower button on the response box to indicate which word they thought had been on the list. The correct choice (or predicted choice, in the case of the null trials) was the upper word for one-half of the trials and the lower word for one-half. The test pair remained on the screen until a response was made. Subjects were told that they must always pick one of the words and that they should guess if they did not know which word was on the list. Both the responses and the latencies were recorded.

In the retroactive control condition, the subjects saw 120 target words followed by 120 three-digit numbers. They counted and then reported the number of even numbers.

RESULTS

The theoretical analysis of the results will be preceded by a general, nontheoretical analysis of the characteristics of the proportions of choices and latencies. Some additional supporting data will also be introduced.

General Analysis of Response Proportions

The mean proportions of choices for each of the test pairs in each of the four conditions are shown in Table 2. Since both paid and unpaid subjects were used in this experiment, preliminary analyses were carried out separately for these two groups. Analyses of proportions correct and response latencies for paid and unpaid subjects showed no difference in performance between the two groups. The data were therefore combined in all subsequent analyses. Standard notation is used such that P(LO,HN) gives the proportion of responses in which a low-frequency old word was chosen when paired with a high-frequency new word. That is, the first member of the pair designates the class of the word selected as a response, high frequency or low frequency and old or new. The second member designates the class of the word with which the selected word was paired in the forced-choice test. It can be seen that there is little difference between the two control conditions, proactive and retroactive. There is also little difference between the two experimental conditions, proactive and retroactive. There is, however, a consistent decline in accuracy between the control conditions and the experimental conditions. A three-way analysis of variance was carried out on the proportions of correct choices for the standard pairs using the arcsin transformation $(Y' = 2 \arcsin \sqrt{Y})$ to obtain homogeneity of variance. The main effect of list length was significant $[F(1,76) = 8.266, p = .006]$, as well as the type of test pair $[F(3,228) = 19.678, p < .001]$. (The latter result reflects the standard word-frequency effect in recognition.) There was, however, no difference between the proactive conditions and the retroactive conditions

Table 2
Mean Proportions of Choices for the Difference Model

	Null Pairs		Standard Pairs			
	P(HN,LN)	P(LO,HO)	P(HO,HN)	P(HO,LN)	P(LO,HN)	P(LO,LN)
Control Condition						
Observed—Proactive	.68	.68	.79	.86	.87	.92
Observed—Retroactive	.68	.64	.80	.84	.82	.90
Predicted	.63	.63	.78	.86	.86	.92
Experimental Condition						
Observed—Proactive	.64	.60	.75	.80	.80	.82
Observed—Retroactive	.67	.62	.75	.82	.80	.86
Predicted	.64	.58	.71	.82	.77	.86

(F < 1). There were no significant interactions. A separate analysis of variance was carried out on the two control conditions as an additional check on their equivalence. There was no difference between the proactive control and retroactive control groups (F < 1).

A separate analysis of the proportion data for the null pairs showed no effect due to list length [F(1,76) = 3.009, p = .087], type of interference (F < 1), or type of pair [F(1,76) = 3.663, p = .06]. There were no reliable interactions.

General Analysis of Response Latencies

The mean response latencies for the standard test pairs for each of the four conditions for correct responses and for incorrect responses are shown in Table 3. The latency results for correct responses parallel those for the proportion data. Response latencies were longer after a longer list had been studied. There were no differences between the proactive groups and the retroactive groups. An analysis of variance showed that there was a significant difference in latency due to list length [F(1,76) = 7.646, p = .008] and to type of test pair [F(3,228) = 13.536, p < .001], but no difference in latency between proactive and retroactive conditions (F < 1). There were again no significant interactions. As was found for the proportion data, there were no consistent differences in mean response latency between

the proactive and retroactive control conditions. A separate analysis of variance of the proactive and retroactive control conditions for the latencies yielded a value of F < 1. The two control conditions were clearly equivalent.

The mean latencies for error responses were based on a very small number of observations. However, all of the observed differences in these latencies were in the same direction as those for correct responses. It took longer to make an error response after seeing a long list than after a short list, and there was no difference between proactive and retroactive conditions. A separate analysis of the latencies for error responses was performed. Missing data, resulting from cells in which there were no errors, were estimated on the basis of row and column means according to a procedure outlined by Winer (1962). The degrees of freedom were therefore reduced by the number of cells estimated. (Latencies for error responses for one subject were not available due to a computer failure and were therefore omitted from this analysis.) The main effect of list length failed to achieve significance [F(1,75) = 1.158, p = .286]. This is not surprising, since the number of observations was small and the variances were very large.

A separate analysis of the latency data for the four null pairs (HN-LN, LO-HO, LN-HN, HO-LO) showed no effect due to list length [F(1,75) = 3.196, p = .078] or

Table 3
Mean Latencies (in Milliseconds) for the Difference Model

	Null Pairs				Standard Pairs							
	HN,LN		LO,HO		HO,HN		HO,LN		LO,HN		LO,LN	
	C	I	C	I	C	I	C	I	C	I	C	I
Control Condition												
Observed—Proactive	2855	2931	1774	2005	1979	2607	2019	2927	1697	2846	1864	3566
Observed—Retroactive	2559	2946	1788	1962	2117	2790	1987	2923	1753	2692	1848	3214
Predicted	2654	2768	1805	2187	2105	2662	2094	2770	1708	2667	1687	2772
Experimental Condition												
Observed—Proactive	3180	3035	2405	2634	2651	3296	2390	2945	2180	3179	2257	3640
Observed—Retroactive	3005	3213	2134	2611	2634	3132	2347	3481	2460	3127	2242	3572
Predicted	3043	3281	2359	2633	2511	3079	2475	3292	2241	3092	2187	3297

Note—C = mean latencies for the correct responses; I = mean latencies for incorrect responses (unpredicted for null pairs).

type of interference (F < 1). Only the type of null pair had a significant effect [$F(3,225) = 51.742$, $p < .001$]. There were no significant interactions.

Additional Supporting Results

The equality of proactive and retroactive interference effects is a strong prediction of the model. This equality, which was observed in both the accuracy and the latency data, has not been demonstrated before. Some investigators have found proactive interference (McCormack & Swenson, 1972; Schulman, 1974; Shepard & Teghtsoonian, 1961); some have failed to find such effects (Donaldson & Murdock, 1968; Shepard & Chang, 1963). Only one study examined both proactive and retroactive interference in recognition (Schulman, 1974). Schulman found proactive interference but failed to find retroactive interference in a standard recognition memory paradigm. The findings of the present study are strengthened by the fact that no difference between proactive and retroactive interference is detected in either accuracy or latency. Schulman reports only accuracy data. Further support for the present results can be obtained from data available in a separate study.

A replication of the proactive control condition was carried out as part of a completely separate set of experiments that is currently in progress.[1] Fourteen subjects between the ages of 16 and 25 years were recruited from universities in the Boston area and were paid $10 for their participation. A comparison was made of both accuracy and latency between words on the first half of the study list and words on the second half. If proactive and retroactive interference are indeed equal, there should be no difference between performance on the first half of the list and performance on the second half. The mean proportion of errors was .17 (SD = .10) for words from the first half of the list and .16 (SD = .11) for words from the second half. Mean latency for the first half was 1,492 msec (SD = 378) and for the second half, 1,473 msec (SD = 342). There was clearly no difference in performance between the two halves of the list in either accuracy or latency. These results are offered as supporting evidence for the equality of proactive and retroactive interference.

ANALYSIS IN TERMS OF THE MODEL AND DISCUSSION

The model presented offers an underlying process to account for interference in recognition memory: indirect marking. On the basis of the model, it was predicted that indirect marking would result in equal proactive and retroactive interference effects. While it is, of course, not possible to assert the null hypothesis that the proactive and retroactive effects are equal, the data clearly fail to reject such a hypothesis. With an F value less than 1 for the comparison of proactive and retroactive effects on both accuracy and latency measures, even increasing the power of the test by increasing the α level would not result in rejection of the equality hypothesis. It should further be noted that the failure to find a difference between the proactive and retroactive conditions cannot be attributed to lack of power in the present data, since the differences due to word frequency and due to interference were clearly detected.

The symmetry of the interference effects strongly suggests that the interference is operating on items in storage. If interference affected the efficiency of the encoding of a new stimulus during study, then only proactive interference would be observed. If interference operated only in the retrieval of an item, then only retroactive interference would be observed. When both proactive and retroactive effects are observed, it is necessary either to propose two different mechanisms or to identify one which would produce the observed symmetry. The increase in overall marking fills the latter role.

Parameter Estimation

The computer program STEPIT (Chandler, 1969) was used to obtain minimal chi-square parameter estimates. The data that were fitted were the mean response proportions and correct and incorrect latencies for each of the six types of test pairs for each of the four conditions: proactive control, retroactive control, proactive experimental, and retroactive experimental. Four of the eight parameters estimated have already been discussed in detail. These four parameters were: $p(HN,n)$, the proportion of indirectly marked elements for high-frequency words on a control (short) list; $p(LN,n)$, the proportion of indirectly marked elements for low-frequency words on a control (short) list; α, which is $s/S(H)$ in the model and represents the efficiency of direct marking for high-frequency words; and β, which is $s/S(L)$ and represents the efficiency of direct marking for low-frequency words. The size of the sample of elements taken at study and test, s, was set at the value of 4 estimated from Glanzer and Bowles (1976) data. The parameters α and β were assumed to be the same for all four test conditions. The values of $p(HN,2n)$ and $p(LN,2n)$ for the experimental conditions were determined directly from the model using Equation 4. Thus, four estimated parameters were sufficient to predict the 24 observed proportions.

Predicting Response Latencies

The symmetry of proactive and retroactive effects observed in the accuracy data is mirrored in the latency data. Latencies are longer for the experimental conditions than for the control conditions, and they are the same for proactive and retroactive interference conditions. This result provides further evidence that the effects of interference on processing are of the same magnitude for both proactive and retroactive sources of interference.

There is considerable literature on choice reaction time in a variety of perceptual tasks (e.g., Audley, 1960; Link, 1975; Smith, 1968). There have also been a

number of attempts to model latencies in a yes-no recognition memory task (e.g., Atkinson & Juola, 1973; Ratcliff, 1978; Ratcliff & Murdock, 1976). There are no models at this time, however, to handle latencies in a forced-choice recognition memory task.

Three types of models were considered in the attempt to predict the latency data. The class of random-walk models was rejected because of the failure of these models to handle the increase in latency for error responses observed in the present data (Audley & Pike, 1965). A simple accumulator model discussed by Audley and Pike (1965) seemed promising because it predicted an increase in latency for error responses. The predicted increase, however, was very small relative to the present data, and the model was rejected for that reason. Finally, a difference model was tested that was highly successful in predicting both correct and error latencies.

The Difference Model

The difference model is based on the idea that latency is related to the difference in the number of marked elements between the two final samples from the stimulus words in the test pair. This difference can be viewed as playing a role similar to that of a stimulus factor in a standard choice reaction time task. It moves a subject more or less strongly toward a given response. The size of this difference is directly related to the distance between the underlying theoretical distributions of old and new items, which has been used successfully to model recognition latencies in a yes-no task (e.g., Atkinson & Juola, 1973). Within each of the two conditions, control and experimental, latency is a decreasing function of the theoretical difference in the number of marked elements. STEPIT was used to fit this model to the data. Latency was calculated as a decreasing linear function of the expected value of the difference between the number of marked elements in the samples from the test pair:

$$L_{ijk} = c_i - k_i d_{ijk}, \qquad (6)$$

where i specifies the two conditions, control and experimental, j indexes the six types of test pairs, and k indexes the response that is being predicted. The constants c_i and k_i specify the intercept and slope for each condition, and d_{ijk} is the mean difference in the number of marked elements. The response proportions, the latencies for correct responses, and the latencies for error responses for each of the four conditions provided 72 data points to be fit with eight parameters to be estimated. The chi-square statistic, which was minimized, was that used by King and Anderson (1976) for the simultaneous evaluation of proportion and latency data:

$$\chi^2 = \sum_{i=1}^{4} \sum_{j=1}^{6} \sum_{k=1}^{2} \left[\left(\frac{L_{ijk} - L'_{ijk}}{s_{ijk}} \right)^2 + \frac{N(P_{ijk} - P'_{ijk})^2}{P'_{ijk}} \right], \quad (7)$$

where i indexes the four conditions, j indexes the six types of response pairs, and k indexes the member of the pair that was chosen. L'_{ijk} is the predicted latency for a choice of k in Pairing j for Condition i; L_{ijk} is the observed mean response time; s_{ijk} is the standard error; P_{ijk} is the observed mean proportion of response of k; P'_{ijk} is the predicted proportion of k responses; N is the number of responses in each cell, 400 here. The parameter estimates were: $\alpha = .11$, $\beta = .19$, $p(HN,n) = .06$, $p(LN,n) = .04$, $c(control) = 6,164$, $c(experimental) = 6,991$, $k(control) = 3,221$, and $k(experimental) = 3,349$. The model provides an excellent fit to the data with a chi-square value of 66 with 64 degrees of freedom. This value will be obtained by chance more than 20% of the time.

The latency data can be well handled as a function of the difference in the number of marked elements between the old and new distributions. However, no mechanism is provided to account for this relationship. Furthermore, the excellent fit is obtained by allowing the slope and the intercept to vary between the experimental and control conditions. While the slopes are very close for the two conditions (3,221 and 3,349), there is a substantial difference between the intercepts (6,164 and 6,991). A similar result was obtained by Koppell (1977). Koppell predicted latencies in a yes-no recognition memory task as a logarithmic function of the separation between test item and criterion along a theoretical decision axis. He found an increase in the intercept of the latency function with list length, but the slopes were the same for long and short lists. The theoretical interpretation of such a shift in intercept is not clear.

Interference and Response Bias

Shepard and Teghtsoonian (1961) reported an increase in false alarms throughout the list in a continuous recognition memory task. They interpreted this result as evidence of proactive interference. Donaldson and Murdock (1968) replicated the Shepard and Teghtsoonian (1961) study and analyzed the results within a signal detection framework. They reported no change throughout the list in discriminability as measured by d'. The observed increase in false alarms was attributed to a criterion shift. The reason for such a shift can be clarified in the context of the present indirect marking model of interference. According to the view proposed here, both old and new distributions are shifting higher on the familiarity continuum as the list is processed (see Figure 1) because of directly and indirectly marked elements. The distributions of old and new items move slightly closer together, but the major effect is the shift of the distributions higher on the continuum. If the subject kept the same criterion placement in a yes-no task while the distributions shifted higher on the continuum, this would result in an increase in both hits and false alarms. This result was observed by Donaldson and Murdock (1968) and was interpreted within a signal detection analysis as indicating a criterion shift. They offered no theory to account

for such a shift. The present model can account for these results. If it is assumed that the subject maintains a constant criterion, the model predicts an increase in both hits and false alarms due to a shift in the location of the underlying distributions. What Donaldson and Murdock viewed as an unmotivated shift in criterion is viewed here as a result of the subject's keeping the criterion fixed while the distributions shift as a result of successive list items.

Conclusions

Both proactive and retroactive interference effects have been demonstrated in a standard proactive-retroactive recognition memory paradigm. Interference is seen to affect both accuracy in a forced-choice task and response latency. These results are interpreted within the framework of a stimulus-sampling model of recognition memory processing. Indirect marking of words related to the target words is offered as a theoretical construct for interference. It is proposed that exposure to items in the experiment results in an increase in the marking of other items. The increase in marking level reduces discriminability between stimuli and increases processing time.

The stimulus-sampling model considered here has now been shown to handle the effects of two different variables quite well: word frequency and list length. The parameters vary in ways that are amenable to psychological interpretation and consistent with related findings in the literature. In the case of word frequency, the indirect marking parameter is higher for high-frequency words than for low-frequency words, which can be interpreted in light of the fact that high-frequency words have more associates than do low-frequency words (Paivio, Yuille, & Madigan, 1968). The direct marking parameter is lower for high-frequency words, which has been interpreted as reflecting the fact that high-frequency words have more meanings than do low-frequency words, a larger set of elements from which a sample is taken (Glanzer & Bowles, 1976). In the present study, the indirect marking parameter varies with list length, as predicted by the model. As was pointed out earlier, the notion of some construct like indirect marking to account for interference is neither new nor unique to the present model. What is offered here is an explicit mechanism of the recognition process that has been quantified and tested in two different experiments.

The problem of evaluating the goodness of fit of any model is a difficult one. It is particularly difficult in the case of the present model because there are no other explicitly defined models of the forced-choice recognition memory paradigm with which to compare it. One approach to this problem will be to modify the model so that it can be tested in a yes-no paradigm. An important criterion for evaluating the usefulness of a model must be its ability to generalize to a variety of para-

digms. A test of the model in a yes-no paradigm will contribute to this evaluation and also make it possible to compare its success with that of other models of the yes-no recognition task. At the present point in its evaluation, the model appears to be a potentially productive tool for the analysis of recognition memory processing.

REFERENCES

ATKINSON, R. C., & JUOLA, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

AUDLEY, R. J. A stochastic model for individual choice behavior. *Psychological Review*, 1960, 67, 1-15.

AUDLEY, R. J., & PIKE, A. R. Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology*, 1965, 18, 207-224.

BOWER, G. H. Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C: Winston, 1972.

CHANDLER, J. P. STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, 14, 81-82.

DONALDSON, W., & MURDOCK, B. Criterion change in continuous recognition memory. *Journal of Experimental Psychology*, 1968, 76, 325-330.

ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychological Review*, 1955, 62, 145-154.

GLANZER, M., & BOWLES, N. Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 21-31.

KING, D. R. W., & ANDERSON, J. R. Long-term memory search: An interesting activation process. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 587-605.

KOPPELL, S. Decision latencies in recognition memory: A signal detection theory analysis. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, 3, 445-457.

KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, R.I: Brown University Press, 1967.

LINK, S. W. The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, 1975, 12, 114-135.

McCORMACK, P. D., & SWENSON, A. L. Recognition memory for common and rare words. *Journal of Experimental Psychology*, 1972, 95, 72-77.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph*, 1968, 76(1, pt. 2).

RATCLIFF, R. A theory of memory retrieval. *Psychological Review*, 1978, 85, 59-108.

RATCLIFF, R., & MURDOCK, B. B., JR. Retrieval processes in recognition memory. *Psychological Review*, 1976, 83, 190-214.

SCHULMAN, A. I. The declining course of recognition memory. *Memory & Cognition*, 1974, 2, 14-18.

SHEPARD, R. N. Application of a trace model to the retention of information in a recognition task. *Psychometrika*, 1961, 26, 185-203.

SHEPARD, R. N., & CHANG, J. J. Forced-choice tests of recognition memory under steady-state conditions. *Journal of Verbal Learning and Verbal Behavior*, 1963, 2, 93-101.

SHEPARD, R. N., & TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, 62, 302-309.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, 69, 77-110.

UNDERWOOD, B. J., & FREUND, J. S. Word frequency and short-

term recognition memory. *American Journal of Psychology,* 1970, **83**, 343-351.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

NOTE

1. The data from the study that is reported in this paper were not available in the form necessary to carry out this post hoc analysis. The new data reported here were collected at the Mental Performance and Aging Laboratory at the VA Outpatient Clinic in Boston, Massachusetts.