Psychological Review
1988, Vol. 95, No. 4, 528–551

0033-295X/88/$00.75

Judgments of Frequency and Recognition Memory in a Multiple-Trace Memory Model

Douglas L. Hintzman
University of Oregon

The multiple-trace simulation model, MINERVA 2, was applied to a number of phenomena found in experiments on relative and absolute judgments of frequency, and forced-choice and yes–no recognition memory. How the basic model deals with effects of repetition, forgetting, list length, orientation task, selective retrieval, and similarity and how a slightly modified version accounts for effects of contextual variability on frequency judgments were shown. Two new experiments on similarity and recognition memory were presented, together with appropriate simulations; attempts to modify the model to deal with additional phenomena were also described. Questions related to the representation of frequency are addressed, and the model is evaluated and compared with related models of frequency judgments and recognition memory.

Although memory for specific events (episodic memory) and memory for abstract concepts (generic memory) seem quite different intuitively, experimental evidence for different underlying systems is sparse (see McKoon, Ratcliff, & Dell, 1986; Ratcliff & McKoon, 1986; Tulving, 1986). One suggestion has been that the two systems are affected differently by repetition, with multiple occurrences establishing multiple traces in episodic memory but strengthening a single representation in generic memory. A primary purpose behind the simulation model, MINERVA 2 (Hintzman, 1984), is to test this notion indirectly by attempting to account for performance in both episodic and generic memory (Hintzman, 1978) tasks using the same multiple-trace mechanism. Application of the model to generic memory has focused on concept learning, as represented in the laboratory by the schema abstraction, or classification learning task (Hintzman, 1986b). The present article describes how the model can be applied to memory for presentation frequency—a quintessentially episodic memory task—and to recognition memory, which is treated as a special case of memory for frequency.

The model, MINERVA 2, is an outgrowth of theoretical ideas regarding effects of repetition on memory that have been stated less rigorously elsewhere (Hintzman, 1976; Hintzman & Block, 1971; Hintzman, Grandy, & Gold, 1981). Because the inspiration for these ideas came from experiments in which subjects judge from memory aspects of an item's presentation—most particularly, its frequency—it is important in evaluating the model to establish how well it deals with data obtained in frequency-judgment experiments.

The organization of the present article is as follows: The first section presents the model's basic assumptions, which are primarily concerned with similarity, repetition, and retrieval. The second section describes how the model accounts for several experimental results that have been reported in the literature on memory for frequency and recognition memory. In the third section, new experiments are presented that test predictions of the model concerning similarity and recognition memory. The fourth section describes a slightly more elaborate version of the model that includes an intertrace resonance process (Hintzman, 1986b) and applies this model to further results on memory for frequency. The fifth section briefly discusses attempts to deal with additional phenomena by constructing special versions of the model, with varying success. Finally, the general discussion evaluates the model and compares it with related models of the same tasks, and addresses issues concerning the representation and encoding of frequency information and its relation to recognition memory.

The Model

MINERVA 2 is primarily concerned with long-term or secondary memory, although there is also assumed to be a temporary buffer store or primary memory, whose function is to communicate with secondary memory. As a multiple-trace model, MINERVA 2 assumes that each experienced event is represented in memory by its own trace. From a theoretical perspective, secondary memory is seen as a vast collection of episodic memory traces, the majority of which were formed outside the experimental context. As will be shown later, however, contextual information specified in the retrieval cue can greatly suppress the activation of traces formed in nonspecified contexts. Thus, in principle, the effect of extraexperimental traces on experimental performance can be reduced arbitrarily close to zero, simply by increasing the amount of contextual information in the retrieval cue. To make simulation manageable, therefore, extraexperimental traces were ignored in the present simulations, on the assumption that they would have only negligible effects on performance on the experimental tasks.

For mathematical simplicity, a specific event is represented

This research was supported by National Science Foundation Grants BNS-7824987 and BNS-8403258.

Correspondence concerning this article should be addressed to Douglas L. Hintzman, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

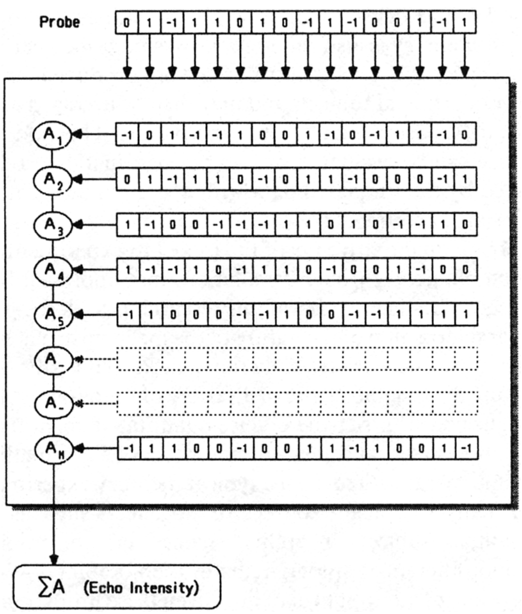

Figure 1. Activation of traces in secondary memory by a probe. (The level of activation of each trace, A_i, is determined by its feature-by-feature similarity to the probe. Echo intensity is the sum of the A_i values.)

in MINERVA 2 as a vector of feature loadings having the values $+1$, 0, and -1. The array labeled *probe* at the top of Figure 1 shows how an event is represented. There are N features, $j = 1 \cdots N$ ordered from left to right, and every feature is assigned a value in the vector representing each event. A value of 0 indicates that, for the event in question, the indicated feature is either irrelevant or unknown. One could view the elements of the vector as connections, linking the feature nodes at a lower level with a single-event node at a higher level. Values of $+1$ and -1 could then be interpreted, respectively, as excitatory and inhibitory links.

Encoding an event entails copying the event vector into secondary memory, represented by the large box in Figure 1. In the model, each individual feature is stored with probability L, the learning rate, and so encoding may be imperfect. If a particular feature is not stored, the value entered into the trace is 0. The parameter L is applied independently to each feature in every event. Thus, when $0 < L < 1$, traces will match their original events to varying degrees, and multiple traces of a repeated item may not be exactly the same. Typical traces of several events are shown in the secondary memory box in Figure 1.

Retrieval is never spontaneous, but is always produced by a retrieval cue or probe originating in primary memory. A probe is assumed to activate all memory traces in parallel, and the traces are assumed to respond simultaneously, producing a single composite *echo* that emanates back from secondary memory into primary memory. This composite response will typically differ for different probes because the contribution of each trace to the echo depends on its similarity to the probe. (Note that probes are completely specified; i.e., the parameter L does not apply to a probe.)

The similarity of a given trace, i, to the probe is given by

$$S_i = \sum_{j=1}^{N} P_j T_{i,j} / N_i, \tag{1}$$

where P_j is the value of feature j in the probe, $T_{i,j}$ is the value of feature j in trace i, and N_i is the number of features relevant to the comparison of the probe and trace i. (Feature j is relevant if either $P_j \neq 0$ or $T_{i,j} \neq 0$; thus, $N_i = N - Z_i$, where Z_i is the number of features for which both $P_j = 0$ and $T_{i,j} = 0$.) The numerator of this function is a version of Tversky's (1977) similarity metric. S_i behaves much like a Pearson r, being zero when trace i is orthogonal to the probe and $+1$ when the two are identical. Values of S_i approaching -1 are mathematically possible, but they are extremely unlikely and would have no particular theoretical meaning in the work presented here.

The degree to which a trace is activated is a positively accelerated function of its similarity to the probe. The simulations reported here used the activation function,

$$A_i = S_i^3. \tag{2}$$

The nonlinearity of the A_i function allows retrieval to be quite selective: In principle, all secondary memory traces are activated by the probe, but the response of secondary memory as a whole is dominated by those traces that most closely match the probe. Note that the expression for A_i preserves the sign of S_i, so that a trace can have negative activation. The relation between A_i and S_i is shown in Figure 2. The operative range of the function in the present simulations is roughly that contained in the unshaded area of the graph.

The simultaneous activation of all traces by a probe produces an echo that has two properties, *intensity* and *content*. The intensity of the echo is found by summing the activation levels of all traces:

$$I = \sum_{i=1}^{M} A_i, \tag{3}$$

where M is the number of traces in secondary memory. The more traces there are that match the probe and the more closely

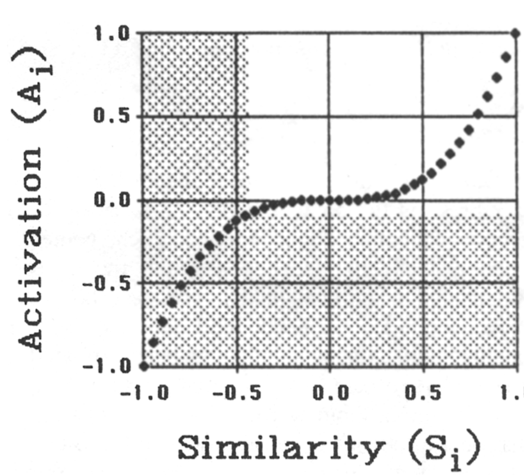

Figure 2. Trace activation as a function of similarity to the probe. (The operative range lies in the unshaded region.)

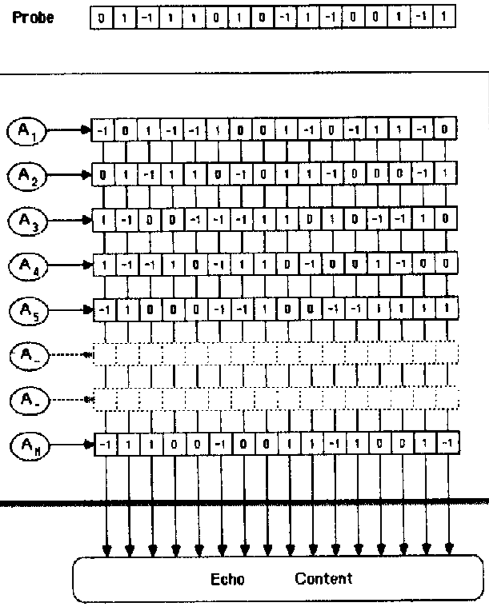

Figure 3. Determination of echo content. (After traces have been activated by the probe, they pass activation to their constituent features. Echo content is the pattern of activation across features induced by all secondary memory traces in parallel. A_i = activation of trace *i*.)

they match it, the greater will be the echo intensity. If there are no secondary memory traces similar to the probe, then the value of I should be near 0. Echo intensity serves as a kind of familiarity signal and is used in modeling frequency judgments and recognition memory (Hintzman, 1984, 1987). Most of the predictions of concern here are based on echo intensity as computed by Equation 3.

The content of the echo is the summed content of all secondary memory traces, each weighted by its activation level. One can imagine the set of *M* episodic trace nodes, each passing its positive or negative activation downward through a network of positive and negative links to the set of *N* features. The activation of each feature is the net result of its input from all traces of which it is part.

Echo content is a vector **C**, whose *j*th element is given by

$$C_j = \sum_{i=1}^{M} A_i T_{i,j}. \tag{4}$$

C_j values can be positive or negative, and their profile across features represents the content of the information retrieved from secondary memory. Determination of echo content is depicted schematically in Figure 3. Because the traces most strongly activated by a probe—and, hence, the echo—can contain information not in the probe itself, MINERVA 2 is capable of associative recall. This capability has been exploited in applying the model to other tasks (Hintzman, 1984, 1986b, 1987), but not in the present simulations.

Simulation with MINERVA 2 uses a Monte Carlo technique. For each simulated subject, event vectors are generated randomly, with each feature sampled independently from the set -1, 0, $+1$. The probabilities of -1 and $+1$ are parameters of the model. The determination of which features are to be encoded in memory is also random. As was described earlier, encoding occurs with probability *L*. Data are produced for a large number of artificial subjects and may then be averaged and otherwise analyzed as are data from experiments. The model's performance can be evaluated by comparing functional relations displayed by the simulated data with corresponding data from appropriate experiments.

Although quantitative fits of the model to experimental data could be obtained through simulation, that approach has not been used. The general aim of this research is to develop a general characterization of the nature of the mechanism subserving performance in a variety of memory tasks. Thus, the goal of predicting the general nature of a variety of functional relations outweighs concern over the precise quantitative form that any one relation takes, and task-specific processes are matters for later refinement. Moreover, the typical memory experiment includes many significant sources of variance—such as differences among subjects in ability, changes in motivation over time, and differences among items in memorability—that the model does not attempt to explain. Given these limits, quantitative data fitting seems inappropriately precise. Crude parameter adjustments have been made in much of the present work, but their purpose is to bring simulated data into the same range as human data and to facilitate visual comparisons, not to achieve quantitative fits.

Frequency Judgments and Recognition Memory

Consider the effect of frequency on echo intensity (Equation 3). If there is no trace in secondary memory that matches the probe (more precisely, if all encoded events were generated by a random process orthogonal to the one that generated the probe), then some of the A_i values will be slightly positive and others will be slightly negative, and obtained values of I will be distributed around zero, because its expected value, $E(\sum A_i) = 0$. If there is just one trace in secondary memory that matches the probe, then the one corresponding A_i value will be strongly positive and the others will be near zero; if there are two, then two A_i values will be strongly positive and the rest near zero. As a general rule, if the learning parameter, *L*, is constant, the expected value of I increases linearly with presentation frequency. Furthermore, because *L* is applied independently to each feature of each event, target traces in secondary memory have goodness-of-encoding variances that are additive. Thus, both the mean and the variance of echo intensity increase with repetition.

A simulation run illustrates these effects. The procedure was as follows: (a) Each item was represented by a vector of 20 randomly generated feature values of $+1$, 0, and -1, with probabilities $\Pr\{-1\} = \Pr\{+1\} = \Pr\{0\} = 1/3$. A list of 20 such items was encoded in secondary memory, using $L = .50$. Four items were stored in secondary memory at each of five frequencies, 1–5, and so *M*, the number of traces in secondary memory, was 60. (b) Testing was carried out using each original item as a probe and also using 4 more randomly generated items to represent the condition, frequency = 0. Traces of the test items themselves were not added to secondary memory. (c) I values for each frequency, 0–5, were accumulated in a histogram having

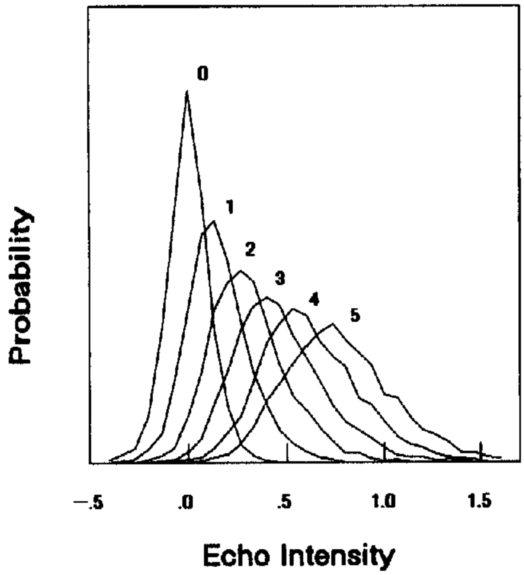

Figure 4. Typical echo intensity distributions for test probes having frequencies of 0–5.

interval widths of .067. (d) The entire procedure was repeated 1,000 times for 1,000 simulated subjects. The resulting distributions of intensity values are shown in Figure 4. It is clear from Figure 4 that the mean and variance increase with frequency.

The model's structure constrains relations among intensity distributions, but parameter settings affect the distributions' quantitative characteristics. Exploration of the effects of four main parameters of the model reveal the following:

1. With a rise in the learning rate, L, differences among means increase and variances decrease (due to increasing N_i in Equation 1).

2. As N increases, N_i also increases, and so S_i and A_i become more stable. Thus, although means are not affected by increasing N, variances decline.

3. The greater are $\Pr\{+1\}$ and $\Pr\{-1\}$, the larger is N_i. Thus, the effects of increasing $\Pr\{+1\}$ and $\Pr\{-1\}$ are similar to those of increasing N.

4. When $\Pr\{+1\} = \Pr\{-1\}$, as in all simulations reported here, randomly generated items tend to be orthogonal to one another. As the ratio of $\Pr\{+1\}$ to $\Pr\{-1\}$ drifts away from 1 in either direction, items become more similar to one another on average. Mean I values rise and so do variances—the latter causing a general increase in overlap among distributions.

Implicit in these generalizations is the fact that there is considerable trade-off among parameters. The primary determinant of performance on both frequency-judgment and recognition tasks is the overlap among distributions, and this can be influenced by manipulating L or N, or even $\Pr\{+1\}$, $\Pr\{-1\}$, and $\Pr\{0\}$. As a practical matter, then, none of the above parameters are identifiable in the sense of having distinctive effects on task performance. This being the case, for most of the following simulations the value of N was set to a convenient value at the outset, and preliminary parameter adjustments were carried out only with L.

Frequency Discrimination

Intensity distributions such as those in Figure 4 can be used to produce data for both the frequency-discrimination and the numerical frequency-judgment tasks. Because test items were generated independently with constant L, forced-choice frequency-discrimination performance can be determined directly from the distributions. The proportion of times the intensity of a B item exceeds that of an A item can be estimated by the following rule:

$$\Pr\{B > A\} = \sum^{k} \Pr\{I_A = k\}\cdot[\Pr\{I_B > k\} + .5\cdot\Pr\{I_B = k\}],$$

where I_A and I_B are the intensities produced by the A and B items, respectively, and k indexes the intervals.

Applying this rule to the distributions of Figure 4 yielded the forced-choice data shown in the main panel of Figure 5. The general pattern is that performance improves with increasing difference between the larger and smaller frequencies; but if the difference is held constant, performance declines as the two frequencies increase. This description is typical of frequency-discrimination accuracy data. For comparison purposes, the inset of Figure 5 shows data from an experiment by Hintzman and Gold (1983), who tested subjects with two instructions: one to choose the item with the larger frequency and one to choose the item with the smaller frequency. Because the results suggested that the two different wordings may have induced opposite response biases, the data from the two conditions have been combined in Figure 5.

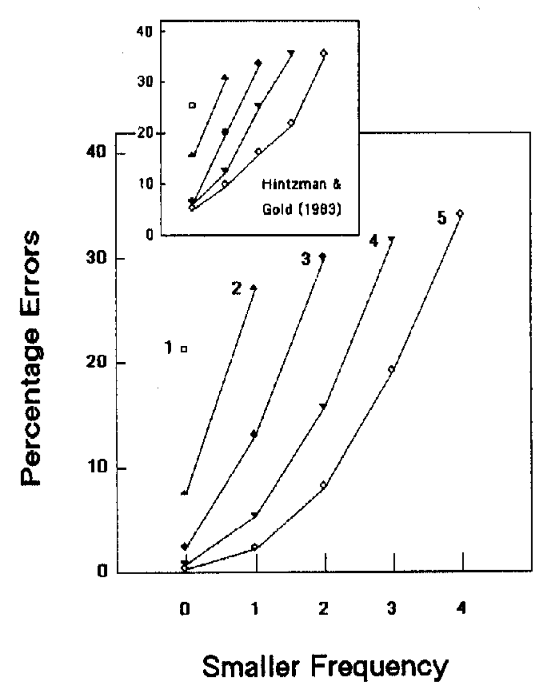

Figure 5. Simulated frequency-discrimination accuracy as a function of the smaller and the larger of the two compared frequencies (parameter = larger frequency). (Inset: experimental data.)

Absolute Frequency Judgments

To predict numerical frequency judgments, one can assume that the echo intensity scale is partitioned by several criterion values, c. Thus, if $I > c_5$, the test item is assigned a frequency judgment of 5; if $c_4 < I < c_5$, the item is given a judgment of 4; and so on. Precise predictions are problematic because criterion settings undoubtedly vary from subject to subject and may even vary within a subject over the course of an experimental session. However, criteria can be set arbitrarily and adjusted in an attempt to approximate empirical frequency judgment distributions.

Most of the frequency-judgment distributions in the literature are less than ideal for comparison with a model; the numbers of observations are usually too small, and many studies use an open-ended or inappropriately truncated judgment scale. An exception is a well-controlled study by Hockley (1984, Experiment 3), who obtained more than 46,000 frequency judgments from 4 subjects, each tested over a series of 12 sessions. Judgments were made during list presentation, at short lags, and responses were restricted to 5 alternatives. The judgment distributions from Hockley's Table 8 (p. 237) are shown in the inset of Figure 6. For comparison, a set of echo-intensity distributions was obtained from MINERVA 2, using the frequencies 0–4 and simulating 250 subjects, for 1,000 observations per distribution. As before, N (the number of features) was 20. After several pilot runs for parameter adjustments, L was set at .80, and the 4 judgment criteria on the echo-intensity scale were set at $c_1 = 0.17$, $c_2 = 0.67$, $c_3 = 1.33$, and $c_4 = 2.00$. The resulting judgment distributions are shown in the main panel of Figure 6. In both the experimental and simulated data, errors are restricted to neighboring frequencies, and generalization increases as frequency goes up. Overall, the experimental and simulated data agree quite well.

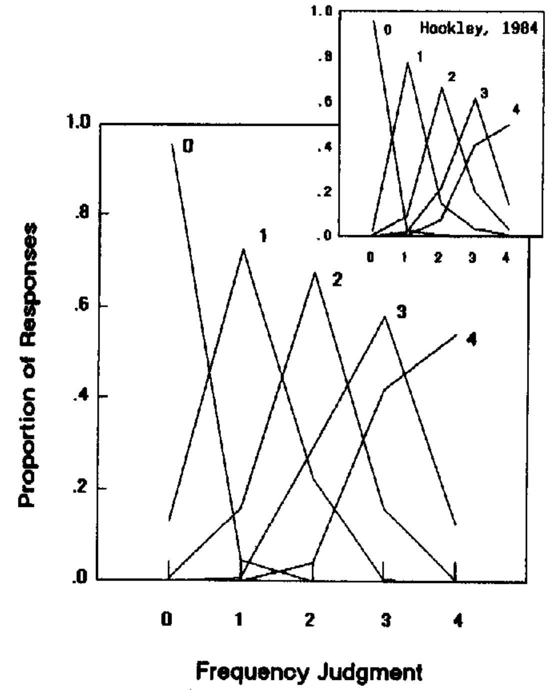

Figure 6. Simulated frequency judgment distributions. (Parameter = frequency. Inset contains experimental data.)

Recognition Memory

Recognition judgments can be seen as a special case of memory for frequency, in which the decisions are focused on the lower end of the intensity scale (Figure 4). In this view, a forced-choice recognition memory test is simply a frequency-discrimination test in which one member of the test pair—the new item, or lure—has a frequency of zero. A yes–no recognition test is equivalent to a frequency-judgment test with two response alternatives (frequency = 0 vs. frequency > 0), and therefore one criterion. Recognition confidence ratings can be modeled by setting several criteria, for different degrees of confidence, in the region in which the distributions for frequency = 0 and frequency = 1 overlap.

In general, the model is consistent with analyses of recognition memory based on signal detection theory (e.g., Banks, 1970). The noise (frequency = 0) distribution originates in the activation of traces of list items by new probes, or lures. Typically, an individual trace will be only slightly activated by a lure, but because echo intensity is the sum of the A_i values, a new test item will sometimes produce an intensity high enough to suggest that the item's trace is in secondary memory. There are several interesting consequences of the way the noise distribution is produced by MINERVA 2. Two obvious ones, explored later, are list-length effects and the false recognition of lures that resemble items in the presentation list.

The increase in variance from frequency = 0 to frequency = 1 implies that binormal plots of operating characteristic (ROC) curves derived from recognition confidence ratings should have slopes of less than 1. Word-recognition data reprinted in Swets (1986) are clearly consistent with this prediction; however, odor-recognition data have slopes of about 1, which suggests that frequency = 0 and frequency = 1 variances were the same.

Forgetting

One implication of the view that recognition memory is a special case of memory for frequency is that forgetting functions for the two tasks should be the same. To demonstrate this, the model was run on a frequency-discrimination task nearly identical to the one that produced the data shown in Figure 5. A total of 500 subjects were simulated on the task using a learning rate of $L = .60$; subsequently, the same was done using $L = .30$. The learning rates were such that in the second run, the number of features stored was about one half that in the first run. This difference can be used to simulate forgetting, under the assumption that one cause of forgetting is trace decay. A learning rate of .30 is equivalent to learning with $L = .60$, followed by forgetting with probability .50, in which a "forgotten" feature value of +1 or −1 reverts to 0 (cf. Hintzman, 1986b). The large panel of Figure 7 shows the forgetting curves for several representative conditions from the simulation run.

There are well-known dangers in comparing forgetting curves that fall at different levels of the dependent variable (Loftus,

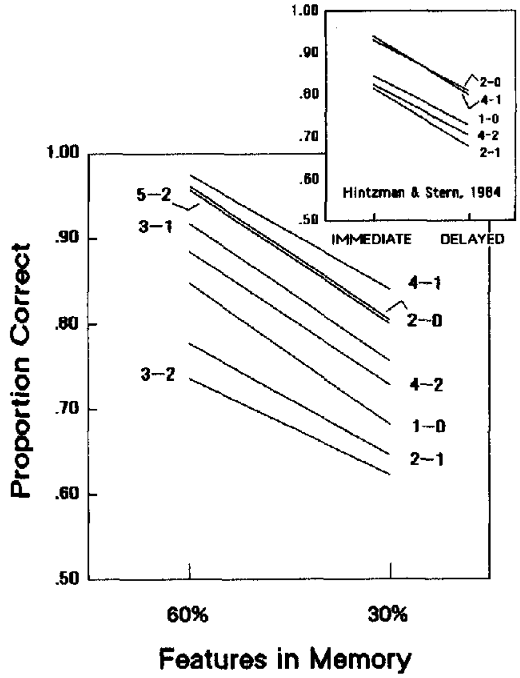

Figure 7. Simulated forgetting curves for forced-choice recognition (1–0 and 2–0 curves) and frequency discrimination (all others). (Inset contains experimental data.)

1985; Underwood, 1954, 1964). One solution is to generate a family of forgetting curves encompassing a range of levels for each of the conditions to be compared, so that different forgetting rates can appear as the convergence or crossing of the two sets of curves. This method was used by Hintzman and Stern (1984) to compare forgetting rates in forced-choice recognition and frequency discrimination. In their Experiment 2, Hintzman and Stern (1984) had subjects make five kinds of frequency-discrimination judgments, two of which involved frequency = 0 items and therefore corresponded to recognition-memory tests. Testing was done either 10 min or 2 weeks after presentation of the list. The inset of Figure 7 shows the data. There was no reliable difference in forgetting rate between recognition decisions and decisions involving higher frequencies (compare the 2–0 curve with that of 4–1, for example, and 1–0 with 4–2).

Hintzman and Stern (1984) interpreted the data as suggesting that "the increments (or traces) left by successive repetitions are all lost at the same rate" (p. 412). That statement accurately describes the decay process underlying the simulated data of Figure 7; and because the simulated data mimic the human data in showing two essentially parallel sets of curves, the simulation supports the interpretation given by Hintzman and Stern (1984).

It is apparent in Figure 7 that MINERVA 2 did not order the various conditions exactly as the human subjects did, and some comment is in order as to what this may mean. First, note that Hintzman and Stern (1984) warned against comparisons of the levels of their curves (as opposed to the forgetting rates) because counterbalancing across levels was incomplete. Second, different frequency-discrimination experiments often order conditions differently (cf. Hintzman, 1969; Hintzman et al., 1981; Hintzman & Stern, 1984); even within the same experiment, test instructions may affect the ordering (Hintzman & Gold, 1983). Third, the ordering of comparisons involving different frequencies, such as 1–0 versus 4–2, is certain to be very sensitive to subtle changes in the variances and shapes of the underlying distributions. An example, demonstrated in the next section, is that different comparisons are affected differently by list length. Fourth and finally, the ordering of conditions in Figure 7 was obtained using a constant value of L, and different orderings can be obtained under the plausible assumption that attention, and therefore L, declines systematically across repetitions. For these reasons, although the discrepancies between orderings in the main panel and the inset of Figure 6 are worth noting, it is not clear that they signify any fundamental problems with the model.

List-Length Effects

Although list length has been found to affect recognition memory in several experiments (e.g., Bowles & Glanzer, 1983; Gillund & Shiffrin, 1984; Legge, Grosmann, & Pieper, 1984; Strong, 1912), studies of its effect on memory for frequency do not seem to have been done. To explore the influence of list length on recognition memory and frequency discrimination in MINERVA 2, four simulation runs were compared. One consisted of the data plotted in Figure 5. In that simulation, four replications of the frequencies 1–5 were stored in secondary memory, for a total list length of 60. The three additional simulations used the same parameters, except that the numbers of replications stored were 1, 2, and 3, yielding list lengths of 15, 30, and 45, respectively. Enough subjects were simulated in each of these new runs to give 2,000 observations per data point (vs. 4,000 for list length = 60). Several representative conditions have been selected for display in Figure 8.

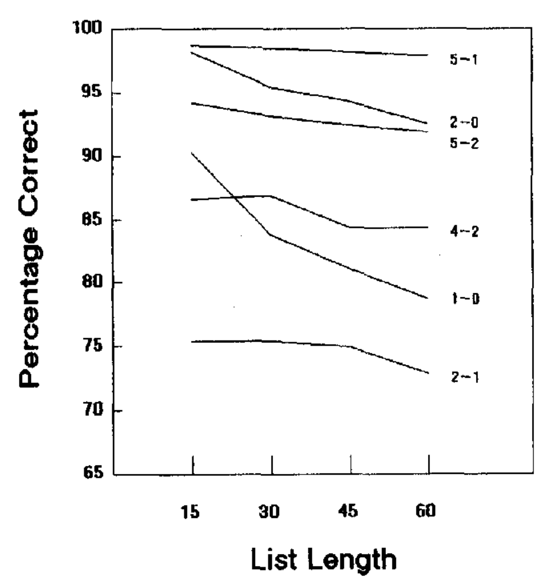

Figure 8. Simulated effects of list length on forced-choice recognition (1–0 and 2–0 curves) and frequency discrimination (all others).

The basic observation is that list length had a much stronger effect on recognition decisions (1–0 and 2–0) than on discriminations among nonzero frequencies. The apparent reason is that recognition accuracy is strongly affected by the variance of ΣA_i for nontarget traces, and this variance increases linearly with list length. Discrimination among frequencies greater than zero, in contrast, depends more heavily on variation in the goodness of encoding of the traces of the items being compared—a factor that list length does not affect. Nontarget traces do have an effect, but their influence decreases with increasing frequency and even at frequency = 1 is proportionally small. As a result, recognition performance deteriorates more rapidly with increasing list length than does frequency discrimination. The difference in slopes is evident in the convergence and crossing of the curves (providing a contrast to the null effects of forgetting shown in Figure 7). This prediction of differential effects of list length on recognition and frequency-discrimination accuracy cannot be related directly to existing data, and therefore requires experimental test.

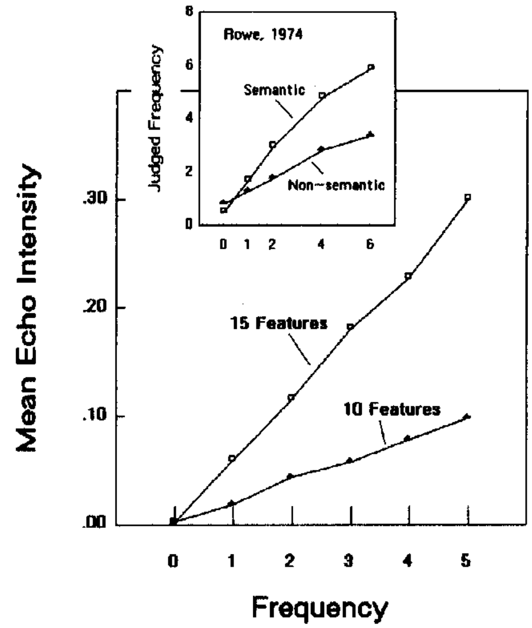

Figure 9. Effect of number of features available for encoding on echo intensity. (Probes had 25 features each. Inset contains experimental data on levels of processing and judged frequency.)

Orienting Tasks

Several published studies have failed to find an effect of intentionality on memory for frequency; that is, subjects instructed ahead of time to remember the items do no worse on a frequency-judgment test than do subjects instructed to remember the frequencies of the items. Other instructional manipulations, however—particularly those directed at the subject's orienting task—do have effects. In experiments reported by Rowe (1974) and Rose and Rowe (1976), subjects made judgments about words during presentation that can be characterized, in levels-of-processing terms (Craik & Lockhart, 1972), as either *shallow* (judgments of visual or phonemic properties) or *deep* (judgments of semantic properties). Frequency judgments were higher and more accurate after a semantic orienting task than after a shallow orienting task. Fisk and Schneider (1984) reported similar results.

A review of the large literature on levels of processing and memory would be out of place here, inasmuch as MINERVA 2 does not offer a new interpretation of effects of orienting tasks. As concepts have evolved, the notion that encoding processes can be characterized by their level and that durability of the product of processing improves with increasing depth (Craik & Lockhart, 1972) has given way to various views, many of which emphasize the distinctiveness of the memory trace (see Cermak & Craik, 1979). Commenting on the proceedings of a conference on levels of processing held in the late 1970s, Craik (1979) used distinctiveness to explain levels of processing effects as follows:

> If shallow codes are characterized as sharing many features with past encodings (possibly because of a relatively restricted repertory or "alphabet" of encoding operations), then . . . the information contained in the present encoding (B) would overlap substantially with information from past encodings (A). In this case, even when there is a high degree of compatibility between the trace B and the retrieval information C, performance will not be high, since the retrieval information also "overlaps" with many past encodings. Put in a slightly different way, there is more cue overload for shallow encodings. (pp. 456–457)

Essentially the same sort of mechanism that Craik (1979) described can be used to explain effects of orienting task on frequency judgments in MINERVA 2. To demonstrate this, a simulation was done in which 25-feature items were encoded in secondary memory with frequencies ranging from 1 to 5. The 25 features were divided into one set of 10 and another set of 15—the first set corresponding to a shallow encoding level (Craik's "more restricted alphabet"), and the second to a deeper level. Four items of each frequency were stored in secondary memory. For one half of these (the shallow encoding condition), the 10 shallow features were learned with $L = .60$ and the 15 deep ones with $L = .00$. For the other half (the deep encoding condition), the two learning rates were reversed. Testing, which included frequency = 0 items, was always done using all 25 features of each probe. Mean echo intensities are plotted in Figure 9. For comparison purposes, the inset panel shows mean numerical-frequency judgments from Rowe's (1974) Experiment 2.

The difference between the 10-feature and 15-feature curves shown in Figure 9 derives primarily from different degrees of activation, owing to differential overlap of target traces with the probe. Of course, the discrimination of frequencies within the 10- and 15-feature conditions could be equal despite the difference between the slopes of the mean intensity curves, if variability were proportional to the mean. That was not the case, however. Standard deviations of the distributions underlying the 15-feature curve were only about 1.5 times as large as those underlying the 10-feature curve, which is not great enough to outweigh the difference in slopes. Frequency-discrimination data were derived from the same simulation run. In overall pattern, both the 15- and 10-feature data resembled Figure 5, but the proportion of correct choices, collapsed over frequency comparisons, was higher for the 15-feature than for the 10-feature condition (.83 vs. .69). Thus, the model predicts that both mean numerical-frequency judgments and frequency discrimination will be affected by orienting task.

This demonstration does not break new ground in explaining levels of processing effects, but it does show that, given the trace distinctiveness interpretation, the effects are compatible with the assumptions of MINERVA 2. A related point is that experiments on "appropriateness of processing" (Morris, Bransford, & Franks, 1977) should also be generally consistent with the model because the efficiency of retrieval depends on the match between encoded features and features of the retrieval probe.

One simplifying assumption of the preceding simulation should be mentioned. The probes consisted of both shallow- and the deep-level features, but words *as stimuli* do not have semantic features; meaning is not available until some retrieval has taken place. It may be assumed that, initially, a probe made up only of the word's sensory features activates preexperimental traces that link the sensory features with the semantic features of the word. Associated information, returned in the content of the echo, can be used to construct a secondary probe consisting of both types of features (see Hintzman, 1986b). It is this secondary probe that the simulation takes as its starting point. Of course, the same preliminary retrieval process must take place during the learning phase of an experiment if the semantic features of a word are to be encoded in the new memory trace. It was assumed here that this happens during the deep-level, but not during the shallow-level orienting task.

Selective Retrieval: List-Specific Frequency Judgments

Subjects can make fairly accurate list-specific judgments of the presentation frequencies of words in two different lists, even when not informed beforehand about the nature of the test. In Experiment 3 of Hintzman and Block (1971), words had frequencies of 0, 2, or 5 in List 1, combined orthogonally with 0, 2, or 5 in List 2. On the test, subjects made separate List 1 and List 2 frequency judgments for each word. For both sets of judgments combined, target-list frequency accounted for about 90% and nontarget-list frequency for only about 9% of the variance among judgment means. Correct response rates on the two lists were roughly symmetrical and have been combined in the inset panel of Figure 9. Obviously, such interlist discrimination would not be possible if judgments of frequency rested solely on probing with the test item and gauging the intensity of the echo. Some means of differentiating lists is required.

Two ways in which traces from different lists might be differentiated are preactivation and postactivation selection. The former can be implemented by attaching appropriate contextual features to the probe and thereby restricting the set of memory traces that it will most strongly activate. The latter would require activating all traces that match the probe, and then somehow apportioning overall echo intensity to the different lists. Preactivation selection seems potentially the more useful of these two techniques, and it is the one that has been explored in the present work. It should be pointed out that the contextual features that are appropriate to the target list are not necessarily the contextual features that would be present during testing. Thus, in order for the preactivation scheme to work, the system must be able to use the instructions to retrieve the discriminating features from memory, and then to add them to the probe. This assumption does not seem unreasonable, but because it is beyond the present capabilities of MINERVA 2, these simulations simply assumed perfectly constructed list-specific probes.

To simulate list-specific frequency judgments, 25 random, 20-feature vectors were generated, one for each cell of a 5×5 design representing the orthogonal combination of frequencies 0–4 in List 1 and frequencies 0–4 in List 2. The number of traces of an item was the sum of its two frequencies; thus, a total of 100 traces were encoded in secondary memory. Appended to each 20-feature trace were additional contextual features, which will be called *list tags*. At the theoretical level, one might prefer to postulate several kinds of temporally related contextual features, none of which discriminates perfectly between lists (cf. Hintzman, Block, & Summers, 1973), but for purpose of demonstration, discrete list tags were assumed.

Several simulations were done, exploring the efficacy of this approach. In some simulations, four list tags were used, and the appended features, in order, were $-1, +1, 0, 0$ for List 1 and $0, 0, -1, +1$ for List 2. The learning probability, L, was applied to list tags just as it was to other features. Two schemes for probing memory were tried. In one (excitation only), the probe consisted of the original item vector, in combination first with the original List 1 tags, yielding a List 1 frequency judgment, and then with the original List 2 tags, yielding a List 2 judgment. In the other scheme (excitation plus inhibition), the appended tags matched those for the target list and were the inverse of those for the nontarget list. Thus, the tags appended to the probe for a List 1 judgment were $-1, +1, +1, -1$, and for a List 2 judgment, $+1, -1, -1, +1$. Excitation-only probes activate target-list tags in memory and are neutral with respect to nontarget-list tags. Excitation-plus-inhibition probes activate target-list tags in memory and suppress nontarget-list tags. In addition to the simulations using four list tags, other simulations were done using eight list tags. Although twice as long as the four-tag sequences, these sequences were generated in the same way. In each simulation, all 25-item vectors were used as probes, once for List 1 and once for List 2, so that each simulated subject responded to a total of 50 probes.

An example of the performance of MINERVA 2 on list-specific frequency judgments is shown in the large panel of Figure 10. In this particular simulation there were 25 simulated subjects, L was .75, and there were eight list tags. Both excitation and inhibition were used in constructing the probes. To avoid arbitrary assumptions about criterion settings, the echo-intensity scale was not partitioned for numerical judgments of frequency. This creates no serious problems, as a monotonic relation between echo intensity and judged frequency can be assumed. Because List 1 and List 2 judgments were symmetrical, they were combined.

Ability to discriminate between lists is reflected in a comparison of the separation of the curves (target frequency) with their slope (nontarget frequency). Thus, if there were no discrimination of frequency according to list membership, the intensity increment for each unit on the abscissa would be the same as the separation between adjacent curves. If the ability were perfect, the curves would be well separated and flat. A simple discrimination index can be calculated by taking the ratio of the variance among means accounted for by nontarget frequency to that accounted for by target frequency. Assuming linear

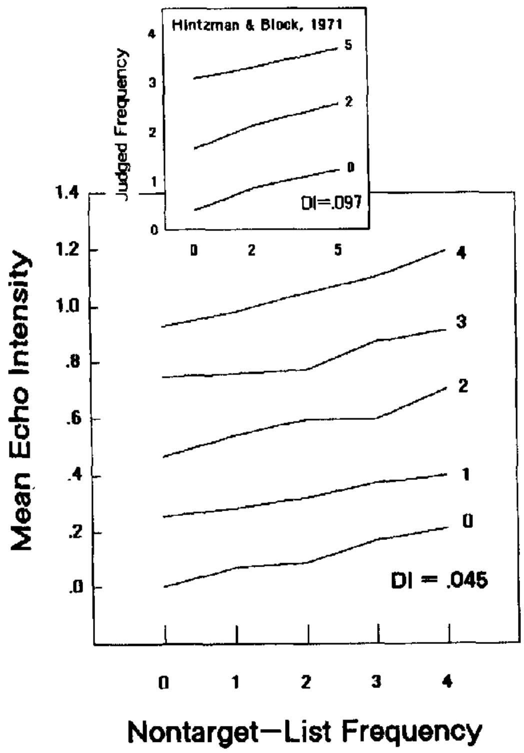

Figure 10. Discrimination between List 1 and List 2 frequencies by the model. (Echo intensities to List 1 and List 2 probes have been combined. A discrimination index, DI, of 0 indicates perfect list discrimination; 1 indicates no discrimination. Inset contains experimental data: mean frequency judgments.)

trends in both cases, the ratio is DI = r_N^2/r_T^2, where r_N and r_T represent correlations of the means with nontarget and target frequencies, respectively. DI = 0 indicates perfect discrimination (i.e., no generalization from the nontarget list), and DI = 1 indicates a complete failure to discriminate. As is shown in Figure 10, the simulated data had a DI of .045, whereas DI for the Hintzman and Block (1971) experiment was .097.

Briefly, the entire set of simulations showed the following: (a) The more list tags that are used, the better is discrimination (e.g., a simulation identical to that in Figure 10 but using just four list tags yielded DI = .207). (b) The higher the learning rate, the better is discrimination (a simulation identical to that of Figure 10, but using L = .50 yielded DI = .065). (c) Constructing probes using excitation alone is not as effective as those using both excitation and inhibition (a simulation identical to that of Figure 10, but without inhibition, yielded DI = .095). It is interesting that there appears to be no way to completely eliminate generalization from the nontarget to the target list by manipulating these parameters; even if nontarget echo intensities are made negative by using inhibition and designating a high proportion of features as list tags, the curves relating intensity to nontarget frequency always have some positive slope.

The capacity for selective retrieval is not restricted to discriminating among lists. Source-specific frequency judgments—for example, judgments of the frequency of internal versus external generation of the same items (Johnson, Raye, Wang, & Taylor, 1979)—could be modeled in exactly the same way. Retrieval in the model is highly context-dependent where generic, as well as episodic, information is concerned (Hintzman, 1986b). An important characteristic of MINERVA 2 is its capacity to determine at the time of retrieval which subset of traces of a particular item will be strongly activated. Frequency judgments do not have to be prestored, but can be generated from memory on demand; in the realm of generic memory, the same holds for concepts (Hintzman, 1986b). Jacoby and Brooks (1984) discussed several advantages of viewing memory in this way.

Recognition and Similarity

The model has several implications for effects of similarity on recognition memory. One is that echo intensity, and therefore the tendency to identify an item as old on a recognition-memory test, should be enhanced if there are items similar to the test item in the list. The phenomenon of false recognition (e.g., Anisfeld & Knapp, 1968)—in which lures that are semantically similar to old items are called old more often than are control items—is consistent with the model. In this regard, MINERVA 2 predicts that the tendency to identify a probe as old should increase with the number list items that partially match the probe, and this should hold for correct, as well as for false recognition. There is some evidence for this tendency where false recognition is concerned (Hall & Kozloff, 1973), but the prediction appears not to have been tested in correct recognition. Both effects are fundamental, as they are expected by several theories of recognition memory (e.g., Bowles & Glanzer, 1983; Gillund & Shiffrin, 1984; Shepard, 1961; Underwood, 1965). Experiment 1 was designed to help fill this gap.

Experiment 1

Method

Materials. The experimental words were 288 familiar nouns (including proper names), 6 falling in each of 48 semantic categories. The categories were selected for high within-category and low between-category similarity. Examples are *booklet, pamphlet, comic book, periodical, magazine, brochure; Scotch, rum, brandy, vodka, whiskey, gin; Jessica Lange, Sissy Spacek, Vanessa Redgrave, Meryl Streep, Sally Field, Debra Winger; minister, priest, rabbi, pastor, preacher, parson; jacket, shirt, coat, sweater, blouse, dress; mouse, prairie dog, groundhog, woodchuck, gopher, chipmunk;* and *Indiana, Wisconsin, Minnesota, Illinois, Michigan, Iowa.*

The 48 categories were divided into four sets of 12, and each set was assigned a presentation frequency of 0, 1, 3, or 5, indicating the number of different category members to appear in the list. Words were arranged randomly in the list, with the constraint that two members of the same category could not appear in close succession. An additional 92 unrelated filler nouns appeared at random list positions. The 200 words in the presentation list were printed in a single column, extending over four pages. To the left of each noun was listed its serial position, and to the right was a blank line for the subject to use in recording an orienting-task response. A single recognition test list was constructed, listing 96 words, randomly ordered and numbered sequentially on the left.

The test list contained two words from each category. For categories having a frequency of 0, both words were new; and for those having a frequency of 1, 3, and 5, one word was old and the other was new. Altogether, eight presentation lists were constructed according to this pattern. Across the eight lists, the presented and nonpresented test items for a category were interchanged, and each category was rotated through the four frequency conditions. In all cases the test list was the same.

Subjects. There were 87 subjects, recruited for course credit from undergraduate psychology classes at the University of Oregon. Subjects were tested in groups. Approximately equal numbers were given each presentation list.

Procedure. Each subject was given a booklet containing (a) instructions to rate nouns on an activity scale ranging from 1 to 5, (b) the presentation list, (c) instructions for a filler task consisting of a sequence of four paper-and-pencil mazes (The mazes were fairly difficult and were intended to occupy all of the subjects for at least 10 min. Subjects were told that if they finished the fourth maze before time was called by the experimenter they were to wait and not turn the page.), (d) the four mazes, and (e) the recognition test page, which included the instruction to circle the number corresponding to each word that had occurred in the presentation list. To allow every subject an opportunity to finish the activity ratings and to provide a short additional retention interval, the experimenter waited 20 min between handing out the booklets and telling subjects to stop working the mazes and turn to the recognition test.

Results

The data in the main panel of Figure 11 are from a simulation that will be described following the presentation of Experiment 2. The inset of the figure shows the hit rates and false alarm rates from Experiment 1. As was expected, hit rates and false alarm rates both increased monotonically with the number of same-category items in the list. For purposes of statistical analysis, the data for each presentation list were combined over subjects and were treated as a macrosubject. Although the linear trend shown by hit rates was small, it was significant, $F(1, 7) = 19.9$, $p < .01$. A similar trend test for false alarms (computed across category frequencies 1, 3, and 5) was highly significant, $F(1, 7) = 69.2$, $p < .001$. Thus, the tendency to call a test item *old* increased with the number of similar items in the list, regardless of whether the test item itself was new or old, as several theories of recognition memory predict.

In MINERVA 2, the difference between the hit and false alarm curves declines as frequency goes up, due to the increasing variances of the underlying distributions. Such a decline is seen in a comparison of the slopes of the two curves in Figure 11, although this could simply reflect a ceiling effect. A signal-detection analysis of the difference between the two curves yields decreasing d' values of 2.1, 1.8, and 1.6 for category sizes 1, 3 and 5, respectively; but this analysis still might be questioned, inasmuch as the computation of d' assumes that the underlying distributions are normal in form. A more direct measure of overall recognition accuracy is therefore to be preferred. This was accomplished in Experiment 2 by using a forced-choice recognition test. There were two versions of this test that yielded another, more interesting result.

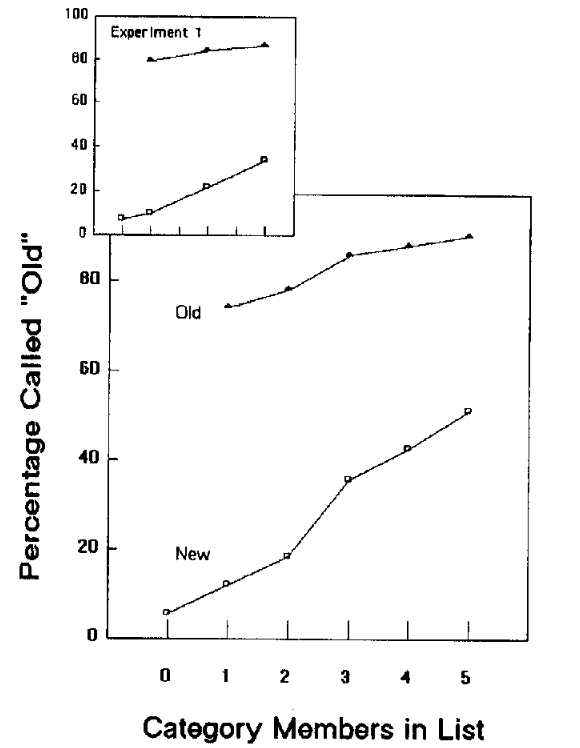

Figure 11. Simulated hit and false alarm rates for related and unrelated test items. (Inset: corresponding data from Experiment 1.)

Experiment 2

Method

The procedure was identical to that of Experiment 1 up to the administration of the recognition test. Each subject was given one of two test forms, both of which were forced-choice versions of the test that was given in Experiment 1. On either form were listed 48 pairs, each consisting of nouns from categories having the same presentation frequency in the original list. One member of the pair was old and one was new, except for the frequency = 0 pairs, for which both members were new. Subjects were instructed to identify one member of each test pair as old. On one of the test forms (related condition), the members of each pair were both from the same semantic category; on the other (unrelated condition), the members came from different categories.

Crossing two test forms with eight presentation lists produced 16 different conditions. A total of 96 subjects, 6 per condition, were recruited as in Experiment 1. Data from 3 subjects were dropped because of their failure to follow instructions. Subjects were tested in groups of varying size.

Results

Correct choice rates are shown in the inset panel of Figure 12. Two effects are evident: First, the ability to choose the old member of a test pair declined as a function of the number of similar items that had appeared in the original list. Second, it was easier to pick the old item if both members of the test pair came from the same category than if the categories were different. Data from subjects having each combination of presentation list and test form were collapsed, yielding 8 macrosubjects for the related test and 8 for the unrelated test. An analysis of variance for planned comparisons confirmed the observations: The effect of category size was significant, $F(1, 14) = 52.0$, $p < .001$, as was the effect of test form, $F(1, 14) = 13.27$, $p < .01$. The interaction of category size and test form did not reach significance, $F(1, 14) = 1.31$.

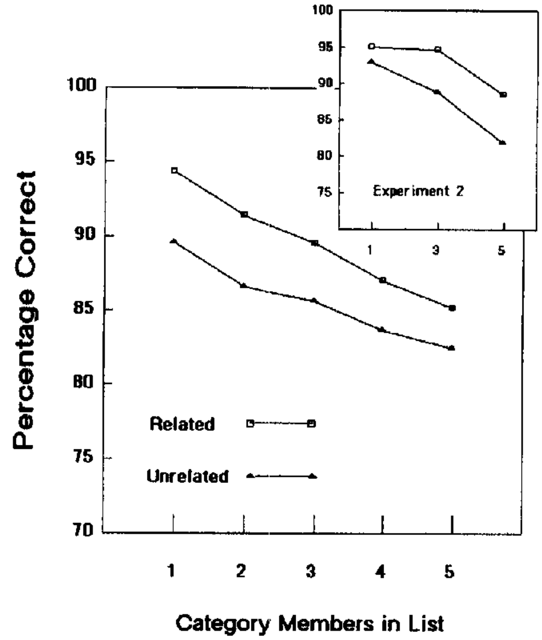

Figure 12. Simulated forced-choice recognition performance for related and unrelated test pairs. (Inset contains data from Experiment 2.)

The effect of category size was expected because an increasing number of category members should increase echo-intensity variability. The effect of related versus unrelated test pairs was also anticipated, partly on theoretical grounds and partly because of similar findings reported earlier in the literature. These matters will receive further discussion after considering simulations of the two experiments.

Simulation of Experiments 1 and 2

The program used for the basic frequency-discrimination task (see Figure 5) was altered so that, instead of having the same item vector stored from 1 to 5 times, from 1 to 5 similar vectors were stored. For each category, a single random prototype was generated, and category members were produced by distorting the prototype. The distortion rule was applied to each feature individually: With probability .30, the prototype feature was replaced with a new one from the set, −1, 0, +1. (Because replacement was random, about one third of the replacement features were the same as the original.) Items for 20 categories were stored in secondary memory in this manner, 4 having each of the category frequencies 1–5. The learning rate was $L = .70$.

To simulate Experiment 1, testing was done first by probing with an original event vector from each category and then with a new member generated from the category prototype. The echo-intensity histograms for all categories were compared, and a single old–new criterion was selected. Hit rates and false alarm rates for all conditions based on this criterion are shown in the main panel of Figure 11. The simulated data are in good qualitative agreement with the results of Experiment 1.

To simulate Experiment 2, testing was done by directly comparing echo intensities from two probes and choosing the one giving the higher value as *old.* Each old probe was tested twice, once against a new distortion of the prototype of the same category (related condition) and once against a new distortion from another category of the same frequency (unrelated condition). A total of 800 subjects were simulated, yielding 3,200 observations per data point. The results are shown in the large panel of Figure 12.

The model reproduced both of the statistically reliable outcomes of Experiment 2: Performance was better on related than on unrelated pairs and, in both cases, decreased with category size. The declines with category size arise from an increase in echo-intensity variance, owing to the moderate but varying similarity of same-category traces to the probe.

The effect of relatedness of the test pairs also is due to variability. To use a familiar analogy, the situation is like the one behind the difference in power between *t* tests for correlated and independent means. In the unrelated condition, the two values being compared, I_A and I_B, are independent. In the related condition, they are correlated because the two probes tend to activate the same subset of memory traces; thus, some of the variance in I is shared variance that does not contribute instability to the difference between the I values. Formally,

$$\mathrm{Var}[I_A - I_B] = \mathrm{Var}[I_A] + \mathrm{Var}[I_B] - 2\,\mathrm{Cov}[I_A, I_B].$$

In the unrelated case, $\mathrm{Cov}[I_A, I_B] = 0$, and in the related case, $\mathrm{Cov}[I_A, I_B] > 0$; thus, $\mathrm{Var}[I_A - I_B]$ is smaller when probes A and B are similar than when they are from different categories.

Although it was not statistically significant, there was a hint of an interaction in the data of Experiment 2 that was not duplicated by the model (see Figure 12). There may have been a ceiling effect in the human data. There is little reason to doubt that the difference between related and unrelated conditions is present even when only one category member was originally stored. Tulving (1981) noted a consistent difference between semantically related and unrelated conditions in several published studies that used verbal materials and demonstrated the phenomenon further using photographs as stimuli. All of those experiments appear to most closely correspond to the frequency = 1 condition here. A similar effect, obtained by manipulating orthographic relatedness, is seen in Hall's (1979) data. It is worth noting that in the unrelated condition of Hall's study, the related-pair lures were not re-paired with targets; instead, two different populations of lures were used. Thus, the overall similarity of lures to list items must have been higher in Hall's related condition than in his unrelated condition—a factor that would tend to make recognition decisions more difficult. Hall's results were complex, but suggest that the relatedness effect was strong enough to overpower this factor when common words were used but not when the stimuli were very rare words.

Tulving (1981) proposed several tentative explanations of the effect of relatedness of test pairs on recognition, but none of them associates the phenomenon with variation in goodness of encoding. It is interesting that a result that falls out of a model as simple as MINERVA 2, with no special assumptions, should appear as such a puzzle when viewed by itself. Almost certainly, many different theories would predict the relatedness effect for the same reason that the present model does. An example is a model recently proposed by Brown (1986) for the specific purpose of accounting for Tulving's (1981) results. Although Brown's model differs from MINERVA 2 in postulating discrete memory states rather than a continuous familiarity dimension, it works analogously in that the familiarity values retrieved by related test pairs are correlated, whereas those retrieved by unrelated pairs are not. What the present treatment demonstrates is that the phenomenon can be derived from general considerations regarding similarity and retrieval; an ad hoc explanation is not required.

Intertrace Resonance Model

As the MINERVA 2 model has been presented so far, similarity effects arise only from relations between traces and the probe. The degree of activation of each trace is solely determined by its overlap with the probe, and traces affect the echo in a strictly additive way. There are other similarity effects, however, that cannot be reproduced by this simple additive model. They require that one take into account the similarity that traces bear to one another, over and above their individual similarities to the probe.

One way to make similarity among traces per se a factor in the model's behavior involves a process of *intertrace resonance,* in which traces that have been activated by the probe can, in turn, pass activation to one another. This secondary activation is a function of feature overlap, as is the primary activation caused by the probe. Exactly such a process was added to MINERVA 2 in order to simulate the Elio and Anderson (1981) generalization results in the schema-abstraction task (Hintzman, 1986b); thus, its invocation here is not entirely ad hoc. Indeed, the idea that traces can be activated by one another, as well as by an explicit probe, is very much in keeping with the spirit of MINERVA 2. Intertrace resonance was not included in the version of the model described up to this point for two reasons. One lies in the belief that the potential of a simple model should be explored before a more complex one is proposed. The other has a purely practical basis: Intertrace resonance greatly increases the amount of time it takes a simulation program to run. Evidence that intertrace resonance is compatible with the results presented in preceding sections is discussed later.

Intertrace resonance can be computed by two algorithms. These differ computationally but not in their basic outcome (Hintzman, 1986b). Both algorithms begin by computing the primary activation of each trace by the probe. One algorithm then computes the secondary activation of each trace, X, by each other trace, Y, according to X, Y similarity and the primary activation level of Y (Hintzman, 1986b, Equation 5). Secondary activation from all sources is then summed. The addition of this algorithm to the model entails an order of magnitude increase in the program's running time. The second method computes the echo content or C vector produced by the initial probe, and constructs a secondary probe from it. When the secondary probe is fed into secondary memory, it produces a secondary echo. With a slight adjustment for normalization when the primary echo is converted into the secondary probe, this algorithm yields the same result as the first one (see Hintzman, 1986b, Appendix), but it is far more efficient, as it increases running time by only a factor of 2.

In simulations of the schema-abstraction task (Hintzman, 1986b), intertrace resonance was used to enhance the retrieval of associative information, as reflected in the content of the echo. Because the present concern is with memory for frequency, we consider the effects of intertrace resonance on echo intensity, rather than echo content.

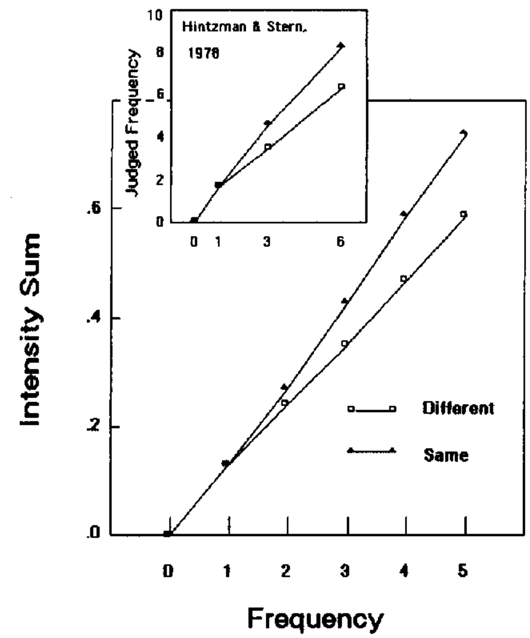

Figure 13. Simulation by the intertrace resonance model of effects of same versus varying context on judgments of frequency. (Intensities of the primary and secondary echoes have been summed. Inset contains experimental mean judged frequencies.)

Contextual Variability and Memory for Frequency

Hintzman and Stern (1978) investigated the effects of varying an item's context on numerical judgments of frequency. As an example of the manipulation, in one experiment they had subjects rate words on six different semantic scales. If a word was repeated, either it was rated on a different scale on each presentation or the rating scale was always the same. When subjects subsequently judged presentation frequencies of the words, their judgments were higher for the same-context condition than for the different-context condition (see the inset panel of Figure 13). In a second experiment, the names of celebrities occurred in sentence frames that either changed when the names were repeated or remained the same. Consistent with the first experiment, judged frequencies for the same-context names were higher than those for different-context names.

The basic MINERVA 2 model, as presented earlier, does not reproduce the context effect reported by Hintzman and Stern (1978). One can assume that each memory trace is divided into an item segment and a context segment, and manipulate these as appropriate for the same- and different-context conditions. But trace activation is a function only of feature overlap between the trace and the probe, and because the probe consists of item-segment information only, similarities among context segments of the traces cannot affect the echo intensity.

To simulate the effect of varying versus constant context on frequency judgments, the intertrace resonance process was added to the basic MINERVA 2 model. Each event vector consisted of 24 features, divided into a 12-feature item segment and a 12-feature context segment. To represent the same-context condition, both the item segment and context segment were identical for each repetition of the item. To represent the different-context condition, the item segment remained the same on each repetition, but the context segment was randomly generated anew. Four frequency = 1 items were stored in secondary memory, as were two items in the same-context condition and two in the different-context condition at each frequency level 2 through 5. Testing was done using probes constructed from item segments alone (i.e., all context-segment features were set to 0).

Effects of intertrace resonance were computed using the echo-probe conversion algorithm from Hintzman (1986b). The feature activation (C_j) values returned in the primary echo triggered by the probe were normalized into the -1 to $+1$ range by dividing each by $\max(|C_j|)$. The resulting vector, including both item and context segments, was then used as a secondary probe. The intensity of the secondary echo was then added to the intensity of the first echo to represent the activation due to the probe and intertrace resonance, combined. Weighting the intensities of the primary and secondary echoes equally in this way is arbitrary, of course; what is important is not the relative weightings but the basic point that the addition of intertrace resonance yields higher echo intensities in the same-context condition than in the different-context condition. This was the outcome, as can be seen in the main panel of Figure 13.

Judgments of Global- and Element-Level Frequency

A recent study by Hock, Malcus, and Hasher (1986) investigated whether subjects could judge the frequencies of letters independently of the frequencies of the four-letter strings in which the letters appeared. Letter frequencies and string frequencies were manipulated orthogonally (letters appeared either 6 or 12 times and strings either 3 or 6 times). Following presentation of the list of letter strings, subjects were asked to judge either letter frequency or string frequency. Both sets of judgments were reasonably accurate; however, string frequency was found to affect letter-frequency judgments—for example, letters that had appeared six times always in the same string were given higher judgments than were letters that had appeared six times in different strings. A reliable converse effect, of letter frequency on string-frequency judgments, was not obtained.

The effect of string frequency on letter-frequency judgments is similar to the context effect described in the preceding section. It is reasonable to suppose, therefore, that the effect can be simulated by MINERVA 2 if the intertrace resonance process is used. However, the failure of letter frequencies to affect string-frequency judgments is more problematic. Each probe string should activate not only its own traces, but also, albeit weakly, traces of all strings with which any letter is shared. Thus, judgments of string frequency should be affected somewhat by the frequencies of their component letters.

Simulations were done to confirm these intuitions and also to answer a question concerning the relative magnitudes of the two effects. The question is whether an experiment with enough power to detect the influence of string frequency on letter-frequency judgments (Hock et al., 1986) could nevertheless fail to detect the converse effect. To some extent, the answer should depend on what proportion of the string is taken up by a letter. A single letter comprised one fourth of one of the strings used by Hock et al. Here, three simulations were done, using letter segments that made up one eight, one fourth, or one half of each encoded event.

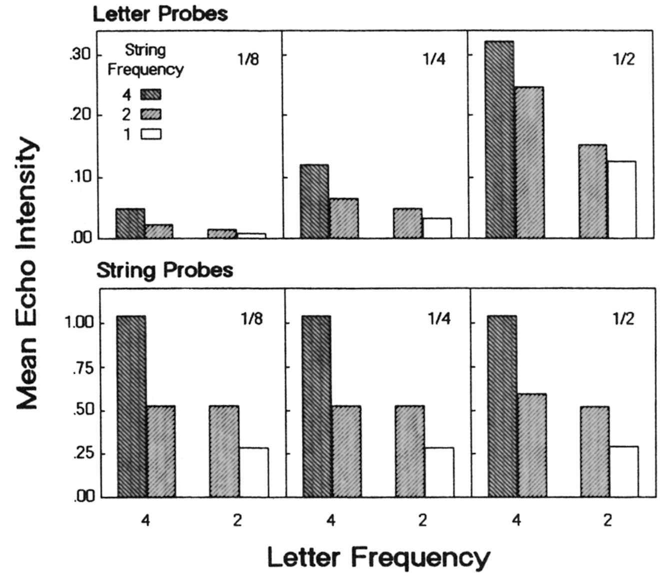

Figure 14. Effects of letter frequency and string frequency on summed echo intensity in the intertrace resonance model. (Fractions indicate the size of a letter relative to the whole 32-feature string.)

A 32-element vector was used to represent each complete string. The subset of features representing what was considered a letter probe was the first 4, 8, or 16 of these features (conditions ⅛, ¼, and ½, respectively). The initial letter segment and the remaining subset of the features of each event were generated randomly, as in preceding simulation runs. The encoded events represented four conditions: 4–4, 4–2, 2–2, and 2–1, in which the first digit stands for letter frequency and the second for string frequency. In all, 12 events are required to fully represent these four conditions. For each simulated subject, 2 replications of the design, or a total of 24 traces, were encoded in memory using a learning rate of $L = .75$. Testing was carried out first using the initial letter segment and second using the entire string for each of the original events. As in the previous simulation, the primary echo was used as a probe to produce a secondary echo, and the intensities of the two echoes were then summed. In each simulation, there were 200 simulated subjects, yielding 800 observations per condition. Echo intensities obtained with letter probes are depicted in the top three panels of Figure 14, and those obtained with string probes are shown in the bottom three panels.

Echo intensity was smallest when the probe consisted of 4 features (letter probes in the ⅛ condition) and greatest when it consisted of 32 features (string probes in all conditions), for the obvious reason that the larger probes were more complete. If we assume that frequency-judgment criteria can somehow be adjusted for the range of intensities likely with a given probe size, then these differences can be ignored.

The remaining results of the three simulations can be described as follows: (a) The primary determinant of echo intensity, for both letter and string probes, was the frequency of the probe. (b) String frequency influenced echo intensity when letter probes were used. When the letters comprised one eight of the string, this effect was roughly as large as the effect of letter frequency, and when they comprised one half of the string, it was roughly 25% the size of the effect of letter frequency. (c) Letter frequency influenced echo intensity when string probes were used, but this effect ranged from a virtually undetectable 0.7% in the one-eighth simulation up to 13.7% in the one-half simulation. Notice that when a letter comprised one fourth of each string, as was the case in Hock et al. (1986), the effect of word frequency with letter probes was much larger than the effect of letter frequency with string probes. The outcome of the simulation thus appears entirely consistent with the Hock et al. results. It is, however, contrary to their conclusion that global-level and element-level units are independently represented in memory because exactly the same memory traces produced the echo intensities obtained with letter probes and string probes shown in Figure 14.

Intertrace Resonance and Earlier Results

A question that arises when a new process such as intertrace resonance is added to a model is whether conclusions based on the simpler model are valid for the new one. Two simulations allowing a performance comparison between the basic model and the modified model are reassuring on this point. The first was a simulation of the standard frequency-discrimination task, using the same parameters that were used in generating the data shown in Figure 5. This task was chosen because variations on its structure were the bases for all of the other simulations reported earlier. The comparison showed no effect of intertrace resonance on frequency-discrimination performance. The resonance process increased the differences among echo-intensity means, but it increased variances to a degree that balanced the effect on the means, so that condition-by-condition comparisons showed no differences in the performance of the two models.

The second simulation was of the list-specific frequency-judgment task (the one shown in Figure 10). Of the several problems the basic model was applied to earlier, this one seemed most likely to show an effect of intertrace resonance on a priori grounds because resonance tends to change the subset of strongly activated traces. The spread of activation to traces not specified by the original probe (e.g., traces from the nontarget list) should cause the discrimination between List 1 frequencies and List 2 frequencies to decline. Except for the addition of intertrace resonance, the simulation was in all respects the same as the one that generated the data in Figure 10. The obtained DI measure was .093—about twice as high as the one obtained without resonance and about the same as that for the human data of Hintzman and Block (1971). Recall that DI ranges from 0 to 1, with 0 indicating perfect discrimination. Thus, although intertrace resonance had the expected effect of increasing generalization across lists, the effect was small.

Other Phenomena

In this section I discuss work done on several phenomena that are not duplicated by either the basic model or the intertrace resonance model in their unmodified forms. These problems are worthy of attention because they suggest ways in which the model can be improved or ways in which it may be fundamentally wrong.

Frequency-Discrimination Reaction Times

Response latencies in frequency discrimination behave much like those in psychophysical discrimination tasks: Both latencies and errors decrease as the difference increases between the two quantities being judged (Hintzman & Gold, 1983; Hintzman et al., 1981). Hintzman and Gold also obtained a kind of congruity effect, in which reaction times differed systematically depending on whether the subjects had been told to choose the item with the larger or the smaller frequency. This finding appeared to entail a subtle speed–accuracy trade-off that Hintzman and Gold were able to capture in a simulation model in which a decision is made by two simultaneous, negatively correlated random walks.

An important question is whether the same type of decision process can be integrated easily into MINERVA 2. Although it will not be presented in detail here, some progress has been made in developing a version of the model that predicts frequency-discrimination reaction times. The key added assumption is that echo-intensity information is not retrieved from secondary memory instantaneously, but accumulates over time, driving a random-walk decision mechanism in the process (Link, 1975; Link & Heath, 1975; Ratcliff, 1978). There are several ways in which this might be realized. The one that has been adopted assumes that subsets of features become available to the decision mechanism sequentially, as would happen if the stimulus is inherently sequential (e.g., auditory), if it must be scanned, or—in the case of words—if visual features become available first, followed by phonemic features, and then by semantic features (which are not present in the nominal stimulus and therefore must be retrieved).

In applying this general idea to MINERVA 2, the list of features constituting the probe has been broken down into several segments that are successively used as probes until a decision process reaches a criterion. On each retrieval cycle, echo intensities are obtained for the corresponding segments of both test alternatives (A and B). The signed differences between the corresponding echo intensities ($I_A - I_B$) are summed, segment after segment, until either a positive (A) or negative (B) criterion is reached. The decision yields a correct response or an error, and the number of cycles to completion is interpreted as a reaction time. The error-rate pattern produced by this model is very similar to that shown for the basic model (and for Hintzman & Gold's, 1983, subjects) in Figure 5. The reaction-time data of Hintzman and Gold are less stable than would be ideal for comparison with a model because of an insufficient number of observations. Nevertheless, the simulated and human data again show the same general pattern: Simulated decision times (cycles to completion) and human reaction times both decrease with an increasing difference between the two frequencies being compared. For a more complete description of the model, see Hintzman (1986a). This version of the model represents a promising approach to explaining both error and latency data in frequency discrimination. A different kind of decision mechanism would have to be added to account for the latencies of absolute frequency judgments (Hockley, 1984).

Recognition Memory for Rare and Common Words

On tests of recognition memory, people usually score better on words that are rare but familiar than on common words. This phenomenon has attracted the attention of theorists because it appears somewhat paradoxical and is the opposite of what is typically found using recall. In relating a multiple-trace model like MINERVA 2 to effects of preexperimental frequency, one's first inclination might be to assume that preexperimental traces are very slightly activated by the test words, and that this activation interferes with recognition-memory judgments. Indeed, simulations with MINERVA 2 show that to the extent that the activation of extraexperimental traces is not suppressed (e.g., by the inclusion of many contextual features in the probe), it increases the variance of the echo-intensity distributions and so depresses recognition performance. The more such preexperimental traces of the test items there are, the poorer the required discrimination between old and new test items becomes.

Unfortunately, this straightforward sort of account is inconsistent with a number of findings. The most severe problem is that activation from preexperimental traces raises not only the variance of the echo-intensity distribution but also the mean; thus, the model predicts that a common word is more likely than a rare word to be identified as *old,* not only when both words are new but also when both words are old. The usual finding is that common words are called *old* more often than are rare words when both are new, but *less* often when both are old. In terms of signal-detection theory, both false alarms and misses are higher for common words than for rare words. This has been called the *mirror effect* (Glanzer & Adams, 1985) because the difference in tendencies to respond affirmatively to a common word and a rare word when they are both old is the reverse of what it is when they are both new. Frequency in the language may be just one variable associated with the mirror effect (Glanzer & Adams, 1985).

A finding by Glanzer and Bowles (1976; see also Bowles & Glanzer, 1983) puts significant constraints on explanations of the mirror effect. A forced-choice recognition test was used in which old–old pairs and new–new pairs were mixed in with the usual old–new pairs. The probability of choosing a common new word over a rare new word was .66, whereas the probability of choosing a common old word over a rare old word was only .36. In neither case is there a correct answer, of course, and in both cases chance performance would be .50. As was pointed out by Glanzer and Bowles (1976), an account of their results in signal-detection terms, assuming a single familiarity dimension, would require that the underlying distributions be ordered: RARE NEW < COMMON NEW < COMMON OLD < RARE OLD.

One way of altering MINERVA 2 to predict the mirror effect would be to follow Gillund and Shiffrin (1984) in their development of the model SAM. To explain the effect of language frequency on recognition, Gillund and Shiffrin (p. 34) assumed that a subject adopts two different criteria, one for common words and the other for rare words, and that forced-choice recognition decisions are based not on raw familiarity but on the distances of the two test words' familiarities from their respective criteria. Each distance is essentially a z score, scaled with respect to an estimate of the standard deviation of the distractor distribution for the appropriate word frequency. The standard deviation is assumed to be estimated during original presentation of the list (Gillund & Shiffrin, 1984, p. 46).

This account seems unattractive for three reasons: First, to explain how the subject estimates the appropriate standard deviation, it is presupposed that the subject knows that each item in the presentation list is new—one of the things the model should explain. The calibration of a subject who just assumed that all list items were new could be thrown off simply by repeating items in the list. Second, the strategy can come into play only well into a list—after the subject has discovered that there are two frequency categories and has assigned enough distractors to each category to estimate the two standard deviations. Third, if different criteria can be set up for common and rare words, different ones could just as easily be established for nouns, verbs, and adjectives, or for long words and short words, or for names of trees, vehicles, mammals, and body parts (cf. Gillund & Shiffrin, 1984, p. 55). Any proliferation of criterion settings detracts considerably from the appeal of the signal-detection analysis of recognition memory; indeed, in applications of signal-detection theory to the forced-choice task, the notion of a criterion is usually not invoked.

Another approach to explaining the differential recognition of rare and common words is to assume that the critical variable is not language frequency per se, but rather one or more correlates of language frequency. This approach was taken in developing a version of MINERVA 2 that reproduces the basic findings of Glanzer and Bowles (1976). Some ad hoc assumptions had to be made about parameters controlling similarity and learning rate, but no changes were made in the model's structure. The underlying assumptions were (a) that certain words are particularly salient—that is, they stand out from other words because of unusual spellings, emotional connotations, and so forth, (b) that a greater percentage of rare than common words have these salient properties, and (c) that salient words are encoded into secondary memory better (with higher L) than are nonsalient words. With the right choice of parameter values, the first and second assumptions together have the effect of ordering rare new words lower than common new words on the echo-intensity scale because common new words tend to share more features with presented words (both common and rare) than do rare new words. And again, with the right choice of parameters, the second and third assumptions together allow rare words to leapfrog the old words during encoding, so that rare old words have higher average intensity than do common old words. Thus, the model produces the ordering of distributions specified by Glanzer and Bowles—RARE NEW < COMMON NEW < COMMON OLD < RARE OLD—and accounts for their data. Details of the simulation can be found in Hintzman (1986a).

All three assumptions can be given some independent justification; for example, (a) Landauer and Streeter (1973) reported that rare words are less orthographically regular than are common words; (b) there is both correlational (Rubin, 1980) and experimental (Groninger, 1976) evidence that intense emotional connotations enhance a word's recognition memory; (c) certain words that one would expect to be salient, such as one's own name and the name of one's home town, are well recognized even though they are highly familiar (Brown, Lewis, & Monk, 1977), and (d) there is suggestive evidence that subjects pay more attention to rare than to common words (Rao & Proctor, 1984). Nevertheless, this particular version of the model is sufficiently ad hoc to discourage further development until favorable experimental evidence of a more direct nature is in hand.

Spacing Effects

Effects of the spacing of repetitions are found in recognition and frequency judgments as in other memory tasks. In a typical experiment involving memory for individual words or pictures, judged frequency increases and recognition memory improves as spacing increases up to about 15-s offset-to-onset time, where it appears to asymptote (Hintzman, 1974). (The parameters of the function may differ for different kinds of materials.) Spacing effects have been the focus of a number of theoretical accounts (see reviews by Hintzman, 1974, 1976). As presented so far, the basic MINERVA 2 model cannot explain spacing effects because the encoding of an item is not influenced by previous events, and no consolidation-like changes are postulated in a trace as a function of its age.

A simple addition to MINERVA 2 enables the model to predict an improvement in memory with spacing. This is the assumption that whenever a feature is encoded into memory it enters a temporary refractory state that prevents its being encoded again until recovery has occurred. Unfortunately, there is experimental evidence against this sort of habituation–recovery account of the spacing effect (Hintzman, 1976; Hintzman, Summers, & Block, 1975). Furthermore, if the habituation process acts at the level of features, then another problem arises: If a similar item is substituted for the second presentation of the same item, an effect of smaller magnitude but in the same direction as the basic spacing function should be obtained. Spacing effects are often found using similar items, but they sometimes are the inverse of that found using repetitions. If the second event is the same word presented in a different modality, then the usual, undiminished, positive spacing effect appears (Hintzman et al., 1973). If the second event is the same word translated into another language, an attenuated, positive spacing effect is obtained (Glanzer & Duarte, 1971), as the habituation and recovery hypothesis would predict. However, recognition memory for synonyms *decreases* with increasing spacing between presentations (Stern & Hintzman, 1979). Thus, the function relating spacing and similarity to retention is complex.

The generality of these relations needs to be further explored. It is not clear whether the inverse spacing effect obtained with synonyms and the positive spacing effect obtained with repetitions are unrelated phenomena, or whether both outcomes need to be explained by the same mechanism. Taking the latter view, a version of MINERVA 2 has been developed that accounts for both results. The basic new assumption is that encoding efficiency is raised from baseline when events in a sequence are of moderate similarity and lowered from baseline when they are highly redundant. The justification for this assumption is that successive events of moderate similarity are especially likely to provide useful information about the causal structure of the environment. Effects tend to immediately follow their causes, and they tend to resemble their causes—for example, the locations, intensities, and temporal patterns of causes and effects tend to be alike (Testa, 1975). To capitalize on these low-level characteristics of the flow of events in time, the learning mechanism modulates encoding efficiency. It compares each new event with traces from the very recent past, suppressing the encoding of sameness (which is uninformative) and enhancing the encoding of potentially interesting relations.

For purposes of simulation, a nonmonotonic function relating learning rate to similarity was chosen, such that $L = 0$ if the current item was identical to, $L = 1$ if it was moderately similar to, and $L = .5$ if it was unrelated to the most recent few events. In principle, all past events affected L, but the function was very heavily weighted by recency. This model did a fairly good job of accounting for both the positive and the negative spacing functions obtained by Stern and Hintzman (1979). Again, the model is ad hoc and lacks independent support, although the result was accomplished within the model's existing structure by manipulating parameters relating to similarity and learning. Details of this simulation can also be found in Hintzman (1986a).

Recognition Memory and General Context

If a list is studied in one context and tested in another, recall can be disrupted even though recognition performance is not. This has been the outcome of experimental manipulations of drug state (J. E. Eich, 1980) and environmental context (e.g., Godden & Baddeley, 1980; Smith, Glenberg, & Bjork, 1978). Recall in MINERVA 2 is highly sensitive to contextual change (Hintzman, 1986b), but previous work with the model shows that recall and recognition can behave quite differently even though they are based on the same memory traces (Hintzman, 1987). Accordingly, several simulations were done to assess the effects of changed context on recognition performance in MINERVA 2. The conclusions can be illustrated by two of these simulation runs.

The basic frequency-discrimination program that produced the data in Figure 5 was modified by adding the same two contextual features, -1 and $+1$, in the first two positions of each event vector. These features were meant to represent the environmental context in which the list was learned. Testing was carried out with probes in which these features were set either to -1, $+1$ (same context) or to $+1$, -1 (changed context). All other features were generated as before. In all, 50 subjects were simulated, with 20 features and $L = .50$. Testing was carried out using all frequency pairings (1–0 through 5–4). Performance on 1–0 (recognition) pairs was 75% correct using same-context probes and 66% using changed-context probes. This difference also held for higher frequency pairings; means were 84% and 74% in the same and changed conditions, respectively, for all pairings combined. Runs using other parameter values and schemes for generating context features yielded essentially the same basic result.

This outcome may at first seem puzzling because changing the contextual features between presentation and test reduces all probe–trace matches to the same degree. A same-context probe will match every list trace on the two context features, and a changed-context probe will mismatch every trace on the same two features, so the difference in degree of match is two features, and this holds for both the more frequent and the less frequent member of the test pair. One might therefore expect a change in contextual features to have no effect.

The failure of this line of reasoning can be traced to Equation 2, which transforms degree of match into intensity (see Figure 2). Because the transformation is nonlinear, the same target-lure difference in degree of match yields smaller intensity differences as the overall degree of match goes down. Intensities of same-context test pairs are easier to discriminate than are those of changed-context pairs because they are higher on the positive limb of the curve. This interpretation was confirmed by a simulation run in which a linear activation function was used—the exponent of Equation 2 was set to 1. In order to compensate for the lower signal-to-noise ratio, L was raised to .80 and the number of features was raised to 30. There were 75 simulated subjects. Performance on 1–0 pairs was 60% in the same-context condition and 61% in the changed-context condition. The corresponding means for all frequency pairings combined were 75% and 74%.

Thus, the decrease in recognition performance when general context features of the trace and the probe do not correspond is caused by the exponent in Equation 2. To eliminate the exponent and its attendant nonlinear generalization gradients would eliminate the difference and simplify the model, but it would almost certainly decrease the range of phenomena the model can handle. For example, nonlinearity seems necessary to simulate the differential forgetting result in the schema-abstraction task (Hintzman, 1986b).

One way around the difficulty is to assume that the instructions, rather than the general contextual features prevailing during the test, determine the contextual features of the probe. This is consistent with the approach to list-specific frequency judgments described earlier, which assumed that the instructions cue the system to retrieve the contextual features that differentiate lists and that the appropriate features are then included in list-specific probes. In the present case, the assumption would be that the instructions cue the retrieval of features of the general experimental context, leading to construction of a context-specific probe. One might object that it should be easier to reinstate the original context when the test context is the same than when it is different; but all that is really required is that the *operative* features of the earlier context be reinstated—that is, those contextual features salient enough to have been incorporated into most secondary memory traces during presentation. This proposition is less difficult to defend (see Bain & Humphreys, in press).

Recent evidence indicates that the sensitivity of human recognition memory to changes in environmental context may depend on study instructions. Smith (1986) found large and reliable context-shift effects following incidental learning, but none when learning was intentional. Thus, the basic model may be seen as predicting the incidental-learning, but not the intentional-learning results. The model offers no definite insights into why orienting instructions should have this effect.

General Discussion

We have seen how the multiple-trace memory model, MINERVA 2, simulates a number of phenomena that have been uncovered in experiments on memory for frequency and recognition memory. The basic model consists of a simple set of assumptions concerning similarity, repetition, and retrieval. Recognition and frequency judgments are based on the intensity of the echo, a composite response of all traces in secondary memory to a retrieval probe. Several results reported here fall out of the model because of its structure, requiring no special assumptions. Examples are the pattern of correct response rates in frequency discrimination and the effect of relatedness of test pairs on performance on a forced-choice recognition test. Other results, such as the effect of orienting task on frequency judgments and the ability to give list-specific frequency judgments, require minimal additional assumptions. Addition of the intertrace resonance process allows the model to explain effects of context on frequency judgments, but does not otherwise change the model in any fundamental way. Addition of a random-walk decision mechanism promises to extend the model to response latencies in forced-choice recognition and frequency discrimination. For certain other results, such as the mirror effect and spacing effects, the model fails without the addition of special, ad hoc assumptions. Despite these problems, the overall applicability of the model to frequency-judgment and recognition-memory data seems good.

It is worth noting that the generality of the MINERVA 2 model goes beyond these tasks. Using the content of the echo as the basis for associative recall, the model has been applied to paired-associate learning (Hintzman, 1984) and to the classification learning or schema-abstraction task (Hintzman, 1986b). The intertrace resonance model used in two of the present simulations was first developed for classification learning (Hintzman, 1986b). It bears emphasizing that this range of tasks is quite broad. Schema abstraction is widely regarded as an experimental analog of the acquisition of concepts or generic memories, whereas recognition memory and frequency judgments are clearly episodic memory tasks. The success of MINERVA 2 in both domains suggests that the distinction between episodic and generic memories, however intuitively compelling, may not reflect different underlying systems obeying different laws.

Because the basic model can perform both associative recall and recognition, it can be used to explore the relation between the two measures. Hintzman (1987) tested MINERVA 2 in the recognition-failure paradigm, in which subjects first study pairs of cue and target words, then are tested for recognition of the targets, and still later attempt to recall the targets when presented with the cues (Tulving & Wiseman, 1975). Results show a surprising degree of independence between the two measures, with subjects often recalling targets they did not recognize (Flexser & Tulving, 1978; Tulving & Wiseman, 1975). Recognition failure of recallable words also characterizes the performance of MINERVA 2—even to the point of apparent independence—despite the fact that the same traces underlie performance in the two tasks (Hintzman, 1987).

Limitations of the Model

The simplicity of the present model is a strength that necessarily entails a number of weaknesses. The failure of the model to explain mirror and spacing effects without ad hoc assumptions or to deal adequately with effects of context on recognition suggests that the model needs further development.

Two other limitations mentioned earlier concern the model's presupposing certain retrieval processes that one would like it to explain. In simulating effects of orienting task on memory for frequency, it was assumed that semantic features of a word are encoded into memory during a semantic orienting task, even though the nominal (visual or auditory) stimulus contains no semantic features. What is missing is an explanation of how the nominal probe retrieves the semantic features to be incorporated into the new trace. Likewise, in the simulation of list-specific frequency judgments, it was assumed that the context features appropriate to the target list are retrieved from memory, based on the instruction to restrict the frequency estimate to a particular list, and are then incorporated into the current probe. Again, the mechanism that carries out this initial retrieval was not explained. Such explanatory gaps are common in present-day memory theory, but it should be recognized that the vague processes filling the gaps are memory processes. As such, they are processes that a more complete theory of memory should explain.

A more subtle problem arises when explanatory gaps are filled with probabilistic processes, as has been done here with regard to stimulus generation and learning. The variability produced by such random processes can be associated roughly with that due to uncontrolled determinants of performance such as subject differences, item differences, subject–item interactions, fluctuations in motivation and attention, and other uncontrolled determinants of experimental performance. The random processes should be seen as oversimplifications, not explanations. They are not meant to imply that the underlying processes are nondeterministic, and the associated parameters and their properties—such as their constancy from trial to trial or item to item—are not to be taken seriously. These disclaimers reinforce the earlier argument against evaluating the model through quantitative data fitting.

A case in point concerns the increase in echo-intensity variance with frequency (Figure 4), which plays a key role in modeling frequency discrimination and numerical judgments of frequency (Figures 5 and 6). The probability L is applied independently to the encoding of each feature of each trace, so that all traces matching the probe contribute to echo intensity as stochastically independent and additive sources of variance. Realistically, there are several reasons why one would not expect learning rates to be uncorrelated over repetitions. An item that is relatively easy to encode on one presentation should also be easy to encode on the next, a subject may decide to attend to poorly learned items at the expense of well-learned ones, and so on. Thus, the fact that different sources of variance are being added is crucial for interpretation of the model's performance, but the fact that the sources are stochastically independent is not.

Another weakness of the model concerns the problem of representation. Event vectors consisting of the values -1, $+1$, and 0 facilitate stimulus generation and the computation of a measure of similarity, but are not likely to be able to capture the complex structure that characterizes the memory code. The model demonstrates how activation of multiple memory representations as a function of similarity could produce several phenomena of human learning and memory, but it should not be construed as a serious account of the basis of that similarity or of the structure of the representations themselves. One might hope that a multiple-trace theory of episodic memory would say precisely what sort of thing the *event* or *episode* is that is encoded in an individual trace. In a typical memory experiment, of course, the flow of experience is nicely segmented into presentations of individual pictures or words; but in other situations, the break points are often unclear, and it may be more difficult to decide on the most appropriate descriptive level to identify as an event. This question might be ignored temporarily on the view that the complex processes underlying everyday memory reduce to a simpler, multiple-trace mechanism under the constraints of the typical memory experiment. Ultimately, though, the issue needs to be addressed.

Still another shortcoming, shared with a number of other models of recognition memory, is that the model does not explain how criteria are set on the echo-intensity scale in order to perform yes–no recognition, confidence rating, and absolute frequency-judgment tasks. Whatever the merits of the idea that subjects base criterion settings on estimates of statistics of the distractor distributions derived during list presentation (Gillund & Shiffrin, 1984) where recognition memory is concerned, this scheme would be totally unworkable for frequency judgments. An adequate account of criterion settings may also have to allow for the rescaling of criteria when factors that substantially affect the intensity of the echo change. Frequency judgments given after a 24-hr interval are less accurate than those given immediately after list presentation, but they are not uniformly lower (Leicht, 1968). Furthermore, judged frequencies of letters are not lower than those for the strings in which the letters appeared (compare Hock et al., 1986, with Figure 13). A nagging problem with models using signal-detection analyses of recognition memory is that, to date, no one seems to have come up with a good answer to the question of how criteria are set.

The recognition-memory literature contains scattered hints of another problem arising from the fact that in MINERVA 2, false recognition and correct recognition are always positively correlated. Let X be an old item, stored in secondary memory, and X' be an new probe item similar to X. Although the intensity of the echo to probe X' will be weaker than that to X, it will be affected similarly by factors that influence the echo intensity to X. Contrary to such behavior, certain findings suggest that under conditions of optimal retrieval of X—in which false recognition of X' is predicted to be highest—subjects can often decisively reject X' as *not* being from the presentation list. Consequently, false recognition of X' bears a nonmonotonic relation to recognition of X.

MacLeod and Nelson (1976), for example, varied test lag in a continuous word-recognition task and found a nonmonotonic function: The false recognition rate for X' was lowest at lag 0, considerably higher at lag 5, and then dropped gradually over lags longer than 5. Other studies have manipulated the frequency of presented items. Hall and Kozloff (1970) varied frequency of X from 1 to 7 and found that false recognition of X' first increased and then decreased, peaking at a presentation frequency of 3. Later, the same investigators showed that if one increases the number of *different* related words (as in the present Experiments 1 and 2), rather than the number of occurrences of the *same* related word, false recognition does not drop, but continues to rise (Hall & Kozloff, 1973). Finally, Tulving (1983, p. 305) presented two groups of subjects with a set of test words that were associatively related to words in a preceding list. Some subjects judged whether the test words were from the list on a recognition test; others were asked to recall the related words from the list, using the test words as cues. Considering only the test words that were new, the ones that were most likely to be falsely recognized by the first group of subjects proved to be the least effective cues for the recall of list associates by the second group. This suggests that subjects in the first group were most susceptible to false recognition when they were unaware that there was a related word in the list. When they knew that the test cue, X', was reminding them of a related list word, X, they were able to reject X' as not being from the list.

Such findings suggest that recognition-memory performance sometimes involves an element of recall (see Gillund & Shiffrin, 1984; Tulving, 1983). In terms of the present model, this implies the involvement of echo content, and not just echo intensity, in recognition memory. It should not be difficult to build such a process into the model because the content and the intensity of the echo are both automatically available following a probe. An echo with some content features that clearly clashed with those of the test probe could be a signal for rejecting the probe as not being from the list, and evidence for a clash would likely be strongest when the associate of the probe appeared relatively recently or frequently in the list.

Allowing a role for echo content in recognition raises the question of whether the separate familiarity signal, based on echo intensity, could be abandoned altogether. Other models have based recognition judgments on the rough equivalent of echo content (e.g., J. M. Eich, 1985; McDowd & Murdock, 1986; Pike, 1984), and some simplification of the present model might be achieved by using this property of the echo for both recognition and recall. More specifically, frequency and recognition judgments could be based on the dot product of the echo content and the probe. Preliminary simulations have been done in which absolute and relative frequency judgments were based on

$$\mathrm{D} = \sum_{j=1}^{N} P_j \mathrm{C}_j / N, \qquad (5)$$

instead of on Equation 3. The simulations show a fairly high correlation between D and I, primarily because A_i values figure prominently in the calculation of both; and the behavior of the model using Equation 5 closely mirrors that of the model using Equation 3, although its performance is marginally worse. With regard to recognition and frequency judgments, therefore, it appears that little besides computational simplicity would be sacrificed by abandoning I in favor of D. It is important to note that the use of D will not in itself explain the decisive rejection of X' when X is very frequent or very recent. For that, a new mechanism would have to be added (e.g., one that detects clashes on specific features), regardless of whether one uses Equation 3 or Equation 5.

There is another consideration that might motivate one to maintain a difference between the familiarity signal and echo content. Study of the amnesic syndrome has produced several findings that suggest a dissociation between *explicit* and *implicit* memory (e.g., Graf & Schacter, 1985). Explicit memory involves recall with awareness, and implicit memory involves recall without it. Graf and Schacter demonstrated that amnesics can learn new associations to words, but the learning was revealed when the instructions simply required giving the first word that came to mind (implicit memory), and not when the instructions referred to words on the previous list (explicit memory). This dissociation was interpreted as evidence of "two qualitatively distinct representational consequences of a learning episode, only one of which occurs normally in amnesic patients" (Graf & Schacter, 1985, p. 516), which is consistent with the currently popular hypothesis that amnesics lack the capacity for new declarative learning, although procedural learning is spared (e.g., Graf, Squire, & Mandler, 1984).

An alternative view is that the same representations underlie both explicit and implicit memory, but that newly encoded traces cannot produce a familiarity signal. Thus, an amnesic might recall the content of a recent event but, because the echo intensity accompanying the recall is low, fails to realize that the content originated in memory. Under implicit memory instructions, patients would produce the content, but under explicit recall instructions they would withhold it, interpreting the absence of a familiarity signal as an indication that the answer is not known. This explanation of the amnesic syndrome gives echo intensity the function of Hart's (1965) "memory monitor," and is based on a conjecture by Hart that memory monitoring may be defective in amnesics (J. T. Hart, personal communication, 1965).

Representation of Frequency

The improvement of memory with repetition has been explained in two ways: Strength theories assume that when an item or class of events is repeated, a single underlying representation is made more potent or more complete. Multiple-trace theories assume that each individual event leaves behind its own representation in memory—thus repetition increases redundancy rather than strength. Apparently, this basic distinction was first recognized by Ward (1893), who favored the strength approach. Although, historically, most memory theorists have shared Ward's bias, several multiple-trace theories have been proposed (Bernbach, 1970; Bower, 1967; Cook, 1946; Hintzman, 1976; Koffka, 1935; Köhler, 1929, 1938; Landauer, 1975; Semon, 1909/1923).

A multiple-trace approach to repetition was adopted in the present model because of its support by the following four basic empirical findings:

1. The multiple-trace assumption suggests that different rep-

etitions may be remembered as independent events. Experiments in which subjects are asked to recall multiple serial positions of a repeated word confirm that the same item can simultaneously have more than one remembered recency or *time tag* (Flexser & Bower, 1974; Hintzman, 1976; Hintzman & Block, 1971, Experiment 2).

2. A corollary of the claim that multiple recencies can be remembered is that frequency judgments should be relative to the *time window* that the instructions specify. In agreement with this prediction, subjects can give largely independent estimates of the frequencies of the same items in two separate lists (Hintzman & Block, 1971, Experiment 3).

3. A multiple-trace representation of frequency should be obligatory—that is, it should be available regardless of whether the subject intended to encode information about frequency per se. Several studies have failed to find superior memory for frequency following an instruction to learn frequency, as compared with the instruction to study items for a later memory test (Flexser & Bower, 1975; Harris, Begg, & Mitterer, 1980; Hasher & Chromiak, 1977; Howell, 1973; Kausler, Lichty, & Hakami, 1984). Frequency judgments also appear to be largely insensitive to feedback and practice (Hasher & Chromiak, 1977; but see Hockley, 1984), and most surprisingly, they show little or no change over an age range from kindergarten to college (Hasher & Chromiak, 1977; Hasher & Zacks, 1979).

4. A multiple-trace representation of frequency would also be in a privileged, analog form that is not shared with other kinds of information, which may be propositionally or associatively learned. Consistent with this property, frequency judgments appear immune to intrusions from deliberately learned numerical information such as digit associates (Hintzman, Nozawa, & Irmscher, 1982) and spatial numerosity (Hintzman, 1982). It is interesting that frequency information can intrude into the recall of digit associates and spatial numerosities, but only if the recall attempt is preceded by a frequency-judgment test—presumably because judging an item's frequency entails translation from the privileged, analog format into a general-purpose, propositional code.

The present work has focused entirely on frequency judgments and recognition memory, and so there has been no demonstration that the model can account for the first finding, although it appears that with the use of contextual or temporal features it would be an easy thing to do. The simulation of the second finding on list-specific frequency judgments was already reported. Clearly, MINERVA 2 has the properties described in the third and fourth findings: Information on the frequency of an item is available if the item's presentations were encoded (an intention to learn frequency is not needed), and this obligatory frequency information has a unique, nonassociative or nonpropositional form.

The strengthlike character of echo intensity may suggest that it is superfluous to keep the traces of episodes separate and that some kind of strength model would explain frequency judgments as well as does the present model. Clearly, such a model would have to allow for some variation among encoded features from presentation to presentation in order to account for the first and second findings, but invariant information could be strengthened with repetition. Associative matrix models (e.g., Anderson, Silverstein, Ritz, & Jones, 1977; Pike, 1984) display this sort of duality. The viability of a such a hybrid approach to repetition effects was acknowledged by Hintzman and Block (1971, p. 304), and it is still an open question whether the range of data on memory for frequency can be explained in this way.

The third finding relates to the argument of Hasher and Zacks (1979, 1984) that the encoding of frequency is automatic. There have been several attacks on this claim, most involving demonstrations that memory for the frequency of an item is diminished if extraneous cognitive demands are made on the subject at the time the item is learned. For example, judged frequency of words is poor if the subject must search for digits while the word list is being studied (Fisk & Schneider, 1984) or if the subject is compelled to rehearse the words aloud repeatedly as a secondary, distractor activity while trying to concentrate on a short-term digit-recall task (Greene, 1984). Similarly, memory for frequency is worse after shallow than after deep processing of the words (Fisk & Schneider, 1984; Rose & Rowe, 1976). The idea behind these attacks on automaticity seems to be that automatic encoding implies that information on an item's frequency will be equally good, regardless of the quality of the traces the item left behind.

Whether or not such findings really undermine the intended position of Hasher and Zacks (1979), it should be apparent that the term *automatic encoding* can be interpreted in more than one way. In a multiple-trace model like MINERVA 2, if an item was effectively encoded on each presentation, then it inevitably has left behind some evidence of its frequency. Obviously, manipulations that degrade the encoding or retrievability of the item's traces will impair the ability to judge the frequency with which the item occurred, as in the levels-of-processing simulations reported earlier. This does not alter the fact that the information on which the frequency estimate is based was acquired automatically (i.e., with no attention directed to the item's frequency per se). Of course, frequency information is not really encoded in MINERVA 2; that is, it is not represented explicitly, in a particular trace. Rather, it is inferred at retrieval from an inevitable, analog property of the memory record (cf. Begg, Maxwell, Mitterer, & Harris, 1986). This distinction is important, as there is evidence that the explicit, propositional encoding of frequency information—which humans obviously can use—is not automatic and is inefficient by comparison with the obligatory, analog representation of frequency postulated in MINERVA 2 (Hintzman, 1982; Hintzman et al., 1982).

The point is that the model appears consistent with the evidence on both sides of the automaticity question. Cognitive development, practice, and instructions to remember frequency do not help much, if at all, because the optional, deliberate encoding of frequency information in a propositional format cannot improve significantly on the privileged, multiple-trace representation that inevitably follows from attention to repetitions. Nevertheless, it does matter whether and how one attends to the repetitions, because that determines the retrievability of the stored information and frequency estimates depend crucially on the effectiveness of retrieval.

A basic assumption of the present model is that recognition memory is a special case of memory for frequency—that is, the information and processes that mediate performance of both tasks is the same (see also Underwood, Zimmerman, & Freund, 1971). Proctor (1977) argued against this view after finding better discrimination between frequencies of 0 and 1 under frequency-judgment than under recognition-memory instructions, but this result has repeatedly failed to replicate (Begg et al., 1986; Harris et al., 1980; Malmi, 1977), and so will not be considered here.

Another argument against the continuity of recency and frequency information was given by Wells (1974, Experiment 1). She reasoned that if frequency judgments of 0, 1, and 2 were based on distributions on a single dimension partitioned by two criteria, c_1 and c_2, then manipulations affecting decisions based on one criterion should similarly affect decisions based on the other. To test this prediction, Wells varied both frequency and recency in a running frequency-judgment task. A scatter diagram of Pr{judgment > 0} (recognition) versus Pr{judgment = 2}, with one point for each frequency and recency combination, revealed functions that were different for frequency = 1 and 2. When the presentation frequency was 2, the two values were monotonically related in the predicted way, but when presentation frequency was 1, Pr{judgment = 2} was constant at about .10, whereas Pr{judgment > 0} varied over a range of .80 to nearly 1.00. It is as though subjects were reluctant to give a judgment of 2 to an item having frequency = 1, no matter how strong the familiarity of the item might be.

Wells's (1974) results are not what the present model would predict, but there are two reasons to be cautious about accepting her conclusion. First, in a running frequency-judgment task, propositional representations regarding frequency will be encoded during presentation and may play some role in determining subsequent judgments, especially at short lags. Second, recency is a confounding factor in Wells's results. That is, values of Pr{judgment > 0} for items having frequencies of 1 and 2 were equated only when the former were tested at much shorter lags than the latter. Subjects are quite good at discriminating recencies over intervals of seconds to minutes (e.g., Hinrichs & Buschke, 1968). If Wells's subjects discriminated the recencies of items that exceeded c_1, they could have adjusted c_2 as seemed appropriate for each item's recency. This possible explanation again underscores the need for an accounting of how criteria are set.

Comparison With Other Models

Two other multiple-trace models of memory for frequency have been proposed. Estes (1976; Whitlow & Estes, 1979) postulated a limited-capacity memory in which the total number of traces is constrained by the number of contextual elements to which they can become attached. Frequency discrimination is done by searching this memory for both alternatives, A and B, and responding with the one that is discovered first. As it has been developed, the model does not apply to numerical frequency judgments, and so it is limited in scope. Moreover, it appears to be inconsistent with several findings from the frequency-discrimination task (Hintzman et al., 1981): Subjects' performance often exceeds the strict maxima that the model predicts; data do not show as much retroactive interference as the limited-capacity assumption implies; and the primary determinant of response latencies is the difference between the frequencies of A and B, not their absolute frequencies, as the proposed search mechanism would lead one to expect.

A multiple-trace model for both absolute judgments of frequency and frequency discrimination was proposed by Schmidt (1978). Much as in MINERVA 2, frequency estimates are based on the number of traces that the test item retrieves. The model assumes all-or-none encoding and all-or-none retrieval of traces, however, and it differs further from the present model in assuming that only traces of the test item are retrieved. The primary source of error is in discriminating the relevant traces of the target item formed during study of the list, from the irrelevant traces formed on other occasions. By contrast, the assumption here has been that the activation of extraexperimental traces is negligible, but even list traces of low similarity to a probe will be activated by the probe to some degree, so a major source of difficulty is in determining whether the retrieved intensity reflects activation of target-item traces or only of nontarget-item traces from the list (cf. Gillund & Shiffrin, 1984; Ratcliff, 1978).

A model developed by Ratcliff (1978) has been applied only to recognition and not to memory for frequency, but it bears similarities to MINERVA 2. Repetition is assumed to give rise to multiple traces, and a memory probe contacts all traces in parallel, with each resonating according to its relatedness or similarity to the probe. Decision times are predicted by a continuous diffusion process, similar to the discrete, random-walk mechanism described earlier (the primary focus of Ratcliff's article is on recognition decision times). One difference between the two models is that relatedness is a primitive construct in Ratcliff's model, although it derives from overlapping features in MINERVA 2. Another is that Ratcliff assumes that the resonance of each trace drives its own individual decision process, whereas in the present model resonances or activation values are pooled.

In the current literature, the model that handles recognition most similarly to MINERVA 2 appears to be the Gillund and Shiffrin (1984) version of SAM. Development of the SAM model has proceeded in different directions than the present work, dealing with free and cued recall as well as with recognition memory. To date, SAM has not been applied to schema abstraction or memory for frequency, although the latter has been suggested as a natural extension of the recognition model (Gillund & Shiffrin, 1984). In SAM, as in MINERVA 2, traces are activated according to similarity to the retrieving probe, and activation of each trace is a positively accelerated function of the degree of overlap of the trace and the probe. MINERVA 2 and SAM are alike in summing the activation of all traces to make what Gillund and Shiffrin call a *global recognition decision,* but SAM retrieves and processes individual images in order to do recall. Despite the similarity in the two models' assumptions regarding the basic recognition process, there are some differences in specific applications. The difference in approaches to the mirror effect were noted earlier, and in SAM, forgetting is caused by retroactive interference rather than by information loss. SAM's recognition performance is not affected by a shift in context, because a matching context simply multiplies the activation of target traces and nontarget traces by the same amount, so that signal-to-noise ratios remain the same. This was the result noted earlier for the present model when the exponent is deleted from Equation 2.

McDowd and Murdock (1986) recently compared MINERVA 2 with Murdock's model, TODAM, in their fit to data from an experiment by Avant and Bevan (1968). The experiment concerned the effect on recognition memory of variation among training stimuli. There were 20 categories of nonsense shapes, and each subject saw four stimuli from each category. For one group of subjects, all four stimuli were the category prototype; other groups saw the prototype 3, 2, or 1 time(s), supplemented by from 1 to 3 distortions of the prototype to bring the total to 4. Following presentation of the patterns, which the subjects were simply told to learn, a recognition-memory test was given that included the 20 prototypes and 20 new distractors. Reported hit rates showed a complex pattern in which the group that had seen each prototype 4 times performed worst (71%), and the group that had seen each prototype 3 times and a distortion of each prototype 1 time performed best (84%). The other two groups fell in between.

Although neither model correctly predicted that the group that had studied only the prototypes would perform most poorly, TODAM fit the data better than did MINERVA 2. Indeed, TODAM predicted little variation among the conditions, whereas MINERVA 2 predicted a monotonic increase in performance with the number of times the category prototypes had been seen. As McDowd and Murdock (1986) pointed out, one reason for TODAM's better fit to the data is that a closed-loop version of the model was used. The closed-loop model abhors redundancy, adding information to memory only when performance (in this case, the ability to recognize the presented item) shows the need. In the prototypes-only condition, by the fourth time exactly the same patterns occurred, TODAM probably had little remaining to learn.

It is doubtful in any case that the Avant and Bevan (1968) findings would generalize very far. The lack of data on false alarm rates makes interpretation of their hit rates problematic. More important, because of the between-subjects manipulation of stimulus variation, the task of looking at the patterns must have become less interesting for some groups than for others. Indeed, Avant and Bevan offered exactly this sort of motivational explanation for their results. An experiment using similar materials but a within-subjects manipulation of stimulus variation would be a better test of MINERVA 2.

In general, it is difficult to see how a closed-loop model (e.g., the version of TODAM in question or a connectionist model based on the *delta rule*) could handle experiments on memory for frequency, such as those that have been of central concern here. Judged frequency continues to increase with repetition, in an apparently open-ended fashion. People are adept even at discriminating the frequencies of letters in the language (Attneave, 1953). Moreover, as we have seen, identical repetitions yield higher judgments of frequency than do varied repetitions (Hintzman & Stern, 1978), which is the opposite of what a closed-loop model would lead one to expect. Certainly, subjects may ignore information that is already highly familiar, so that it fails to register in memory (Fisk & Schneider, 1984), but they have the option of attending to it and encoding it if they wish. The literature on memory for frequency appears anomalous from the view that learning is a purely adaptive process that ceases once mastery is reached. A primary reason for taking a multiple-trace approach to effects of repetition has been to explain this otherwise anomalous set of results.

References

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

Anisfeld, M., & Knapp, M. (1968). Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology, 77,* 171–179.

Attneave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology, 46,* 81–86.

Avant, L. L., & Bevan, W. (1968). Recognition of a stimulus class member after training with varied numbers of cases per class. *Journal of General Psychology, 78,* 241–246.

Bain, J. D., & Humphreys, M. S. (in press). Relational context: Independent cues, meanings, or configurations? In G. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory.* London: Wiley.

Banks, W. P. (1970). Signal-detection theory and human memory. *Psychological Bulletin, 74,* 81–99.

Begg, I., Maxwell, D., Mitterer, J. O., & Harris, G. (1986). Estimates of frequency: Attribute or attribution? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 496–508.

Bernbach, H. (1970). A multiple-copy model for post-perceptual memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 103–116). New York: Academic Press.

Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence, J. T. Spence, & N. H. Anderson (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 1, pp. 229–325). New York: Academic Press.

Bowles, N. L., & Glanzer, M. (1983). An analysis of interference in recognition memory. *Memory & Cognition, 11,* 307–315.

Brown, J. (1986). Similarity relations in recognition: A model to explain Tulving's data. *Journal of Memory and Language, 25,* 197–201.

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency, and negative recognition. *Quarterly Journal of Experimental Psychology, 29,* 461–473.

Cermak, L. S., & Craik, F. I. M. (Eds.). (1979). *Levels of processing in human memory.* Hillsdale, NJ: Erlbaum.

Cook, T. W. (1946). Repetition and learning: 3. Memory and thought. *Psychological Review, 53,* 214–224.

Craik, F. I. M. (1979). Levels of processing: Overview and closing comments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 447–461). Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Eich, J. E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition, 8,* 157–173.

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

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 397–417.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review, 83,* 37–64.

Fisk, A. D., & Schneider, W. (1984). Memory as a function of attention, level of processing, and automatization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 181–197.

Flexser, A. J., & Bower, G. H. (1974). How frequency affects recency judgments: A model for recency discrimination. *Journal of Experimental Psychology, 103*, 706–716.

Flexser, A. J., & Bower, G. H. (1975). Further evidence regarding instructional effects on frequency judgments. *Bulletin of the Psychonomic Society, 6*, 321–324.

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

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

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13*, 8–20.

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

Glanzer, M., & Duarte, A. (1971). Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior, 10*, 625–630.

Godden, D., & Baddeley, A. (1980). When does context influence recognition memory? *British Journal of Psychology, 71*, 99–104.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 501–518.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 164–178.

Greene, R. L. (1984). Incidental learning of event frequency. *Memory & Cognition, 12*, 90–95.

Groninger, L. D. (1976). Predicting recognition during storage: The capacity of the memory system to evaluate itself. *Bulletin of the Psychonomic Society, 7*, 425–428.

Hall, J. F. (1979). Recognition as a function of word frequency. *American Journal of Psychology, 92*, 497–505.

Hall, J. F., & Kozloff, E. E. (1970). False recognitions as a function of number of presentations. *American Journal of Psychology, 83*, 272–279.

Hall, J. F., & Kozloff, E. E. (1973). False recognitions of associates of converging versus repeated words. *American Journal of Psychology, 86*, 133–139.

Harris, G., Begg, I., & Mitterer, J. (1980). On the relation between frequency estimates and recognition memory. *Memory & Cognition, 8*, 99–104.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208–216.

Hasher, L., & Chromiak, W. (1977). The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning and Verbal Behavior, 16*, 173–184.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*, 1372–1388.

Hinrichs, J. V., & Buschke, H. (1968). Judgment of recency under steady-state conditions. *Journal of Experimental Psychology, 78*, 574–579.

Hintzman, D. L. (1969). Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology, 80*, 139–145.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Potomac, MD: Erlbaum.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10, pp. 47–91). New York: Academic Press.

Hintzman, D. L. (1978). *The psychology of learning and memory.* San Francisco: Freeman.

Hintzman, D. L. (1982). Are presentation frequency and spatial numerosity distinct attributes of memory? *Bulletin of the Psychonomic Society, 20*, 196–198.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, and Computers, 16*, 96–101.

Hintzman, D. L. (1986a). *Judgments of frequency and recognition memory in a multiple-trace memory model* (Tech. Rep. No. 86-11). Eugene: University of Oregon Cognitive Science Program.

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

Hintzman, D. L. (1987). Recognition and recall in MINERVA 2: Analysis of the "recognition failure" paradigm. In P. Morris (Ed.), *Modelling cognition: Proceedings of the international workshop on modelling cognition* (pp. 215–229). London: Wiley.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology, 88*, 297–306.

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Contextual associations and memory for serial position. *Journal of Experimental Psychology, 97*, 220–229.

Hintzman, D. L., & Gold, E. (1983). A congruity effect in the discrimination of presentation frequencies: Some data and a model. *Bulletin of the Psychonomic Society, 21*, 11–14.

Hintzman, D. L., Grandy, C. A., & Gold, E. (1981). Memory for frequency: A comparison of two multiple-trace theories. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 231–240.

Hintzman, D. L., Nozawa, G., & Irmscher, M. (1982). Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 127–141.

Hintzman, D. L., & Stern, L. D. (1978). Contextual variability and memory for frequency. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 539–549.

Hintzman, D. L., & Stern, L. D. (1984). A comparison of forgetting rates in frequency discrimination and recognition. *Bulletin of the Psychonomic Society, 22*, 409–412.

Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). What causes the spacing effect? Some effects of repetition, duration, and spacing on memory for pictures. *Memory & Cognition, 3*, 287–294.

Hock, H. S., Malcus, L., & Hasher, L. (1986). Frequency discrimination: Assessing global-level and element-level units in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 232–240.

Hockley, W. E. (1984). Retrieval of item frequency information in a continuous memory task. *Memory & Cognition, 12*, 229–242.

Howell, W. C. (1973). Storage of events and event frequencies: A comparison of two paradigms in memory. *Journal of Experimental Psychology, 98*, 260–263.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 1–47). New York: Academic Press.

Johnson, M. K., Raye, C. L., Wang, A. Y., & Taylor, T. H. (1979). Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 229–240.

Kausler, D. H., Lichty, W., & Hakami, M. K. (1984). Frequency judgments for distractor items in a short-term memory task: Instructional variation and adult age differences. *Journal of Verbal Learning and Verbal Behavior, 23*, 660–668.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt Brace Javanovich.

Köhler, W. (1929). *Gestalt psychology.* New York: Liveright.

Köhler, W. (1938). *The place of value in a world of facts.* New York: Liveright.

Landauer, T. K. (1975). Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology, 7,* 495–531.

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumption for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12,* 119–131.

Legge, G. E., Grosmann, C., & Pieper, C. M. (1984). Learning unfamiliar voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 298–303.

Leicht, K. L. (1968). Recall and judged frequency of implicitly occurring words. *Journal of Verbal Learning and Verbal Behavior, 7,* 918–923.

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology, 12,* 114–135.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40,* 77–105.

Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 397–406.

MacLeod, C. M., & Nelson, T. O. (1976). A nonmonotonic lag function for false alarms to associates. *American Journal of Psychology, 89,* 127–135.

Malmi, R. A. (1977). Context effects in recognition memory: The frequency attribute. *Memory & Cognition, 5,* 123–130.

McDowd, J. M., & Murdock, B. B. (1986). Mathematical models and the problem of stimulus variation: A comparison of MINERVA 2 and TODAM. *Acta Psychologica, 62,* 177–188.

McKoon, G., Ratcliff, R., & Dell, G. S. (1986). A critical evaluation of the semantic–episodic distinction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 295–306.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing vs. transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review, 91,* 281–294.

Proctor, R. W. (1977). The relationship of frequency judgments to recognition. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 679–689.

Rao, K. V., & Proctor, R. W. (1984). Study-phase processing and the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 386–394.

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

Ratcliff, R., & McKoon, G. (1986). More on the distinction between episodic and semantic memories. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 312–313.

Rose, R. J., & Rowe, E. J. (1976). Effects of orienting task and spacing of repetitions on frequency judgments. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 142–152.

Rowe, E. J. (1974). Depth of processing in a frequency judgment task. *Journal of Verbal Learning and Verbal Behavior, 13,* 638–643.

Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior, 19,* 736–755.

Schmidt, R. (1978). Frequency estimation in verbal learning. *Acta Psychologica, 42,* 39–58.

Semon, R. (1923). *Mnemic psychology* (B. Duffy, Trans.). London: George Allen & Unwin. (Original work published 1909)

Shepard, R. N. (1961). Application of a trace model to the retention of information in a recognition task. *Psychometrika, 26,* 185–203.

Smith, S. M. (1986). Environmental context-dependent recognition memory using a short-term memory task for input. *Memory & Cognition, 14,* 347–354.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6,* 342–353.

Stern, L. D., & Hintzman, D. L. (1979). Spacing and the retention of synonyms. *Bulletin of the Psychonomic Society, 13,* 363–366.

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review, 19,* 447–462.

Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99,* 181–198.

Testa, T. J. (1975). Effects of similarity of location and temporal intensity pattern of conditioned and unconditioned stimuli on the acquisition of conditioned suppression in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 1,* 114–121.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior, 20,* 479–496.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

Tulving, E. (1986). What kind of hypothesis is the distinction between episodic and semantic memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 307–311.

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

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Underwood, B. J. (1954). Speed of learning and amount retained: A consideration of methodology. *Psychological Bulletin, 51,* 276–282.

Underwood, B. J. (1964). Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior, 3,* 112–129.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70,* 122–129.

Underwood, B. J., Zimmerman, J., & Freund, J. S. (1971). Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology, 87,* 149–162.

Ward, J. (1893). Assimilation and association. *Mind, 2,* 347–362.

Wells, J. E. (1974). Strength theory and judgments of recency and frequency. *Journal of Verbal Learning and Verbal Behavior, 13,* 378–392.

Whitlow, J. W., & Estes, W. K. (1979). Judgments of relative frequency in relation to shifts in event frequencies: Evidence for a limited-capacity model. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 395–408.

Received September 2, 1986
Revision received December 4, 1987
Accepted December 17, 1987 ■