PSYCHOLOGICAL REVIEW

VOL. 79, NO. 2 MARCH 1972

RECOGNITION AND RETRIEVAL PROCESSES IN FREE RECALL [1]

JOHN R. ANDERSON [2] AND GORDON H. BOWER

Stanford University

A model of free recall is described which identifies two processes in free recall: a retrieval process by which the subject accesses the words, and a recognition process by which the subject decides whether an implicitly retrieved word is a to-be-recalled word. Submodels for the recognition process and the retrieval process are described. The recognition model assumes that during the study phase, the subject associates "list markers" to the to-be-recalled words. The establishment of such associates is postulated to be an all-or-none stochastic process. In the test phase, the subject recognizes to-be-recalled words by deciding which words have relevant list markers as associates. A signal detectability model is developed for this decision process. The retrieval model is introduced as a computer program that tags associative paths between list words. In several experiments, subjects studied and were tested on a sequence of overlapping sublists sampled from a master set of common nouns. The two-process model predicts that the subject's ability to retrieve the words should increase as more overlapping sublists are studied, but his ability to differentiate the words on the most recent list should deteriorate. Experiments confirmed this predicted dissociation of recognition and retrieval. Further predictions derived from the free recall model were also supported.

This paper has several aims: First, we offer a critique of two popular "strength" theories, one which relates recall and recognition to the strength of one and the same memory trace, and another which relates only recognition memory to a similar strength measure; second, we develop a particular conceptualization about recognition memory which we believe satisfies the criticisms of the traditional strength theory; third, we illustrate how that recognition mechanism could be interfaced with a retrieval mechanism so as to yield a viable theory about multilist free recall. Along with reviewing published data relevant to these points, we also shall present new experimental evidence bearing on the second and third points.

[1] This research was supported by Grant MH-13950-04 from the National Institute of Mental Health to Gordon Bower.

[2] Requests for reprints should be sent to John Anderson, Department of Psychology, Stanford University, Stanford, California 94305.

TWO THEORIES OF RECOGNITION AND RECALL

Kintsch (1970) provides a careful review of the strength theory of recall and recognition which supposes that recall and recognition involve basically the same process, except that recognition of an item requires a lower "threshold of strength" than does recall. It is because of this lower threshold that recognition of an item is easier than recall. This strength theory has been an important learning theory for years (cf. Bahrick, 1965; McDougall, 1904; Postman, 1963). Support for this theory comes from the fact that several experimental variables affect recall and recognition in the same way. For instance, temporal variables such as study time, the retention interval,

and massing and spacing of presentations have similar effects on recognition and recall (Kintsch, 1966; Olson, 1969). Also, the two tasks display similar serial position curves with small primacy and large recency effects (Shiffrin, 1970b).

However, in contrast to these communalities, there are a number of variables that affect recall and recognition in different ways, suggesting that different processes underlie the two performances. As a first example, words of low frequency in the Thorndike-Lorge count of text are more easily recognized (Schwartz & Rouse, 1961; Shepard, 1967), but are recalled more poorly than words of high frequency (Hall, 1954). As a second example, subjects learning under intentional instructions recall better than subjects learning under incidental instructions, but this ordering of the groups is actually reversed for recognition (Dornbush & Winnich, 1964; Eagle & Leiter, 1964; Postman, Adams, & Phillips, 1955). Finally, associative and categorical relationships among items play an important role in recall, but appear to have little if any effect on recognition (Cofer, 1967; Kintsch, 1966).

An alternate theory about the relation between recognition and recall which has as prestigious a history as the strength theory is the two-process theory of recall (James, 1890; Kintsch, 1970; Müller, 1913). This asserts that while free recall involves processes similar to those occurring in recognition, recall differs qualitatively in that additional search processes are involved. Specifically, it is assumed that during recall, an item is first retrieved from memory by the search process (whatever that is), and then it is tested by the recognition process to determine if it is from the to-be-recalled list. Thus, in order for a word to be recalled it must both be successfully retrieved and recognized. Within this two-process model of recall, Kintsch was able to give a natural interpretation of the experimental variables which differentially affect recall and recognition. However, these interpretations were clearly post hoc, and therefore, as Kintsch (1970) himself admits, "little direct evidence for the two-stage theory is available at this time [p. 341]." One of the purposes of the experiments to be reported in this paper is to provide some evidence directly bearing on this issue.

STRENGTH THEORIES OF RECOGNITION

Current literature on memory appears to present a solid concensus regarding the dominant theory of recognition memory —a theory which we will criticize here and to which we will offer an alternative. The dominant theory assumes that upon presentation of a stimulus on a recognition-memory test, the subject can access directly a representation or trace of that item in memory. Stored with this representation is some single continuous (or many-valued discrete) variable that provides a measure of the subject's degree or strength of familiarity with the test item. We will call this measure the "strength" of the item, using the term formally introduced by Wickelgren and Norman (1966). However, this measure has been given many verbal guises—"familiarity" by Kintsch (1970) and Parks (1966); "amount of information" by Freund, Loftus, and Atkinson (1969); and "number of feature matches" by Bower (1967). The important feature to note about the strength theory of recognition is that it is "ahistorical"; that is, it assumes that a subject makes recognition decisions about an item not on the basis of detailed memory of the past history of occurrences of the item, but rather on the basis of a single measure which reflects to some extent its past frequency, recency, and duration of exposure. It is this ahistorical character of strength theory which is the source of all its weaknesses.

The strength conception of recognition memory is practically as old as experimental psychology (see e.g., the use of "excitatory potential" by Hull, Hovland, Ross, Hall, Perkins, & Fitch, 1940). However, a significant technical advance in strength theory came with Egan's (1958) application of statistical decision theory to recognition memory. That approach as-

sumes that there is one normal distribution of strengths for those items which have been studied recently, and a different normal distribution for those items which have not. The first distribution is called the "old" and the second the "new." The average distance between the two distributions reflects the amount of strength that was added to the studied items by their recent presentation. It is assumed that the subject chooses some criterion point, *C*, along the dimension of strength (This criterion corresponds to the former "threshold of recognition.") If the strength of a particular item exceeds *C*, the subject will decide it is an old item and thus "recognize" it. If the strength is less than *C*, the subject will decide it is a new item and reject it.

This strength theory, elaborated with the technical machinery of statistical decision theory, can handle some of the salient facts gleaned from simple recognition experiments. For example, it can handle the data taken from experiments in which the subject is asked to rate the confidence of his recognition judgment (e.g., Murdock, 1965). In particular, the theory predicts the bowed-shape memory-operating characteristic relating changes in confidence rating to changes in the proportions of old and new items assigned to the confidence-rating categories. The theory predicts how this memory-operating characteristic changes as a function of retention variables such as recency and frequency (see Kintsch & Carlson, 1967). The theory can also predict the relationship between multiple-choice and single-item recognition experiments (e.g., Green & Moses, 1966; Kintsch, 1968a).

However, two detracting points need to be emphasized regarding these successes. First, these several phenomena are not strong evidence for the theory because the results have been shown to also be quite consistent with several alternate theories of decision making. For example, Kintsch (1968a) showed that the relationship between yes–no and multiple-choice recognition could be well predicted by at least three theories other than statistical decision theory. Similarly, Lockhart and Murdock (1970) have shown that bowed memory-operating-characteristic curves are derivable from a diverse variety of underlying old and new distributions other than the presumed normal distributions. Second, and of greater consequence, these successes are really triumphs (such as they are) for the technical machinery of statistical decision theory, not for the basic strength assumptions. In fact, they provide no discriminating evidence in favor of the strength assumptions. For instance, Bernbach (1967), using a different conception of the underlying memory representation but similar decision-making machinery, produced a recognition model equally as adequate as that generated from strength theory. Similarly, we too will produce an alternate memory model that is equally capable of handling these elementary facts about recognition. The thrust of this argument is that results from simple recognition experiments do not suffice to discriminate among alternative conceptions of the memory representation.

The strength approach can explain our ability to differentiate between a positive and negative set of items on a retention test if these differ in strength; that is, if the items differ in recency, frequency, and duration of study. The approach fails, however, to distinguish theoretically among items which are equated on frequency and recency, but are nonetheless differentiable empirically on other grounds. Yet there are multiple dimensions along which to differentiate sets of items besides their frequency, recency, duration, or some composite strength measure. Example differentia in verbal learning experiments would be *where* in space the item was presented, *who* said it, *how* it was said, and other special characteristics of its physical and psychological context of presentation. In fairness, the original strength theorists never actually said that subjects could *not* perform discriminations on the basis of these dimensions. They were rather concerned with other matters. Strength of familiarity was adopted as a convenient concept for integrating with their technical

machinery. However, the evidence is now available that an undifferentiated strength of familiarity concept is not sufficiently rich to account for the subject's ability to differentiate sets of items.

Of relevance to this issue is evidence showing item differentiation in the subject's editing of his free recall. An experiment by Bower and Clark (1969) illustrates differentiation in recall of a sort that cannot be handled in simple strength terms. Their subjects learned lists of unrelated nouns by generating meaningful sentences, woven into stories around the critical words. This heuristic greatly enhanced their later recall of the critical nouns. But, for present purposes, the significant fact is that subjects almost never intruded their elaborative additions while overtly recalling (in writing) the critical nouns. In recalling, they would recite their story to themselves but write down only the critical nouns. That outcome is not an isolated incident. When asked, most subjects in free recall experiments will report thinking about many nonlist words (e.g., a category name) to help mediate recall of list words, yet the mediators are rarely intruded in recall. An experiment by G. H. Bower and P. Winchester (unpublished) showed that following a lengthy word association task, subjects had over 90% accuracy in later identifying whether a given item was one of the experimenter's stimulus words or one of their own associate responses.

The argument above claims that an elemental strength measure provides no grounds for discriminating list items from implicit associative responses during list learning. A related difficulty with a simple strength theory is that it cannot accommodate facts about list differentiation. This issue is engaged, for example, in single-trial free recall studies using multiple lists studied and recalled in succession. In the typical task, the subject is required to recall only the "most recent" list. Now, recall of the most recent list could be implemented in strength models by assuming a retrieval mechanism that outputs only those items whose strength exceeds a criterion amount (earlier items having decayed below that amount). However, as Shiffrin[1] (1970a) has demonstrated, subjects are also capable of recalling only those items from the list which just preceded the most recent list; following presentation of List n, his subjects recalled List $n - 1$, and they could do so with reasonable accuracy.

One might imagine that strength theory could also explain Shiffrin's results by invoking the idea of the decay of memory strength (see Wickelgren & Norman, 1966). Since an item's strength decays continuously between its presentation and its retention test, one might suppose that access to List $n - 1$ is provided by an "internal editor" which recalls only those items whose strengths fall in an appropriate bandwidth (see Hinrichs, 1970), above a lower bound (distinguishing List $n - 1$ from List $n - 2$) and below an upper bound (distinguishing List $n - 1$ from List n). This "bandwidth" explanation receives some support in research by Hintzman and Waters (1969, 1970) showing that identification of which of two lists an item appeared in improves with increasing temporal interval between the two lists but deterioriates with increasing interval between study of the lists and the retention test. So, temporal discriminations (implemented by whatever mechanism) are clearly important in list differentiation.

In counterargument, however, it can be shown that such temporal judgments cannot be based solely on any simple strength variable. This can be concluded from experiments which simultaneously vary frequency and recency. An item in List 1 may be presented 10 times as frequently as an item in List 2; so, according to any "strength" theory, the List 1 item will be the stronger. Yet in List 2 recall, the subject is much more likely to recall the List 2 item than the frequent List 1 item. In studies of list differentiation, Winograd (1968) found in fact that such unbalanced frequencies actually *improved* list discrimination rather than the converse as expected by strength theory. Hintzman and Block (1971) showed that correct assignment of an item to List 1 rather than List 2 increased directly with its frequency

of occurrence in List 1; a simple strength theory of list differentiation would have predicted the opposite, namely, higher probability of assigning frequent items to the more recent List 2.

Although list differentiation experiments commonly use each item in just one of several lists presented (see, however, Hintzman & Block, 1971), interesting results are also yielded by experiments in which a given item appears in several lists. We have conducted such an experiment (G. H. Bower & J. R. Anderson, in preparation) in which subjects studied a series of four overlapping lists. There are 16 (2^4) possible combinations of appearances and nonappearances for a particular word across the four lists. Several words were used to instantiate all 16 possible combinations, and subjects were later asked on a retention test to indicate in which lists each word had occurred. For present purposes, the most important result of that experiment is that subjects were much more accurate in list identifications than one could ever predict from a simple strength theory. For instance, subjects can correctly remember that one item occurred in Lists 1 and 4 and that another occurred in Lists 2 and 3, etc.; yet it is simply unimaginable how a single strength measure attached to each item could provide for such patterned discriminations. Such results show that "memory strength" alone is insufficient to account for the salient facts about list discrimination.

To pursue matters just one more countermove, a strength theorist might try to explain such embarrassing results by supplementing the strength information with a further independent measure of the temporal recency of the item. Such a "two-measure" theory might integrate the data in our "2^4" list discrimination experiment along with Winograd's results and those of Hintzman and Waters. However, we would then argue that two measures are still not enough. For instance, the proposed two measures (strength plus a recency tag) cannot explain the ability of subjects to discriminate implicit associates from test items, or their ability to identify the location or the modality (auditory or visual) in which an item was presented (D. L. Hintzman, unpublished). On these grounds it would seem reasonable to suppose that subjects can perform list discrimination on any reasonably salient dimension.

RECOGNITION VIA ASSOCIATION TO CONTEXT

For such reasons, we reject the view that recognition is mediated by a simple strength measure or any simple combination of measures. In place of this we propose that items are recognized by retrieval of certain kinds of context information originally stored along with the item in question. This contextual information includes physical characteristics of an item's presentation, implicit associations to the item, and some cognitive elements representing the list in question. We have no startling insights regarding the nature of these context elements which are presumed to vary across lists. They might include the subject's general mood or attitude, his physical posture, and his physiological state, as well as any conspicuous external cues prevailing during presentation of List n. One prominent set of cues may be provided by implicit verbalizations, especially an implicit count as when the subject says to himself "first list," "second list." Another prominent set of contextual cues for the recognition of an item could be provided by other words in the list. An interesting curiosity is how the subject identifies where List $n-1$ leaves off and List n begins; but clearly, temporal grouping and distinctive signals, instructions, or activities (like recall) interpolated between lists all combine to promote identification of beginning and end segments of that serial unit we call List n. A further curiosity, not explored here, is the issue of how the subject decides whether temporally distributed clusters of presentations constitute different presentations of the "same" list or are "different" lists.

A significant consequence of this context-retrieval theory is that the task of item recognition is not really different in

kind from that of list discrimination. In both tasks, the subject must make his decisions on the basis of contextual information retrievable from the test word. The only difference is that the recognition task involves a yes–no decision; that is, the subject must indicate whether he saw the word in a particular list context. In contrast, list discrimination involves a forced-choice task; that is, the subject must decide in which list context the word occurred.

At this point, however, the proponent of the strength of familiarity hypothesis might want to object and insist on a theoretical separation of the simple recognition paradigm from the list discrimination paradigm. His argument might go as follows: "I concede that list discrimination tasks may require something beyond a strength measure, perhaps something like your 'contextual associations' business. But that does not disprove my strength theory of pure recognition. In that task, surely, subjects use strength of familiarity because it is convenient and because it works. In other words, I claim that simple recognition and list discrimination involve fundamentally different mnemonic and decision processes."

As long as both models handle equally well the simple recognition task, there is no way to disprove this dualistic claim. However, on grounds of parsimony one would say that there is no need for a special theory for a special circumstance when a more general theory subsumes the special circumstance and many others. Besides its lack of parsimony, this attempt to give a special status to the pure recognition paradigm ignores the fact that all recognition experiments are implicitly list or set discrimination experiments. This is obvious upon closer analysis of the "pure recognition" experiments. Every subject has seen, heard, and thought thousands of words before he sees the experimental list, and he frequently sees, hears, and thinks of a great many nonlist words between study and test. We have already noted that subjects think implicitly of many words while the study list is being presented. Undoubtedly, included among the distractors in a "pure recognition" test will be words that the subject encountered before, during, and after the study list in question. Therefore, the subject's actual task is to discriminate a particular subset of words from all others on the basis of differentia such as where, when, and how the word was encountered.

For example, in a recent experiment of ours (G. H. Bower and J. R. Anderson, in preparation), the implicit discrimination involved in an alleged "pure recognition" experiment was explicitly given experimental analysis. After hearing a list of study words which they were told to remember, the subjects heard lengthy instructions describing the multiple-choice recognition test they were about to receive. The multiple choice included one word from the study list along with two distractors. Unknown to the subject, however, one of the distractors was a semantically similar word that had been mentioned two times in the instructions interpolated between the study list and the recognition retention test. In comparison to the study-list item, this "instruction distractor" should have been greatly favored in terms of a strength measure, because it enjoyed both higher frequency and shorter recency than the study-list word. However, as intuition and our model suggest, the subjects had not the slightest difficulty in selecting the study-list word of this multiple-choice set. Their ability to discriminate was also unrelated to whether the subject indicated any awareness of the instructional distractors. The instructional distractor was chosen just slightly more than the totally new distractor, a result consonant with some generalization among the "list" and "instructional" contexts associated to the words. This small demonstration just makes transparent the point we have been urging: What is, to appearances, a simple recognition memory experiment is, in actuality, a list differentiation experiment in disguise.

DETAILS OF THE MNEMONIC REPRESENTATIONS

We could represent the entire constellation of contextual stimuli prevailing

during List n presentation as a single pattern which is treated in the learning system as a single unit, subject to all-or-none association to items appearing on List n. However, for technical and theoretical reasons, it is desirable to postulate a pool of hypothetical elements which singly, separately, and independently serve to identify List n. As in the stimulus-sampling theory of Estes (1959), List n will then be represented as a population of a number of variable stimulus components. There will be one such stimulus population corresponding to each ordinal list—one for List 1, one for List 2, and so on. It would be realistic to assume that the successive stimulus populations form an ordered array of uniformly overlapping sets so that stimulus generalization, confusion, or failure of list differentiation would be most likely among temporally adjacent lists. However, the basis for that proximity metric will not be presented here, but it has been developed elsewhere (Bower, 1972b).

Our basic conception of human memory is that of a huge network of tens of thousands of nodes interconnected by associations. The nodes correspond to individual concepts the subject has, and the associations encode relations that the subject has learned between the concepts. Upon presentation of a word in study, we assume that the sensory features of that word activate the node in the network that corresponds to the word. Simultaneously, there are active in the network nodes corresponding to the various contextual stimuli that the subject is attending to. We will make two specific assumptions with respect to how the node corresponding to the word becomes connected to the contextual nodes:

1. At the time of occurrence of any word in the list, we suppose that there is simultaneously active a unique element or node in memory which will be called the "list marker" or "list tag." The purpose of this list marker is to record the context prevailing during that presentation of the word. It does this by interconnecting the set of contextual nodes active at that moment. The marker thus acts like a label for a collection or bundle of contextual elements. We assume that there is a probability θ that any particular element in the List n population of context elements is active and is associated to the list marker.

2. There is a probability α that the subject will form an association between the memory node corresponding to the word and the memory node corresponding to the list marker. This association serves to record a particular fact about the word, namely, that it occurred in a particular experimental context. Thus, we assume, in common with Mandler (1967), Kintsch (1970), Norman and Rumelhart (1970), and many others, that what is learned in a verbal learning task is not the word itself but rather information about the word (e.g., that it was presented in surprising red letters in the first part of the list while the subject was thinking about what he would buy with his wages from service in the experiment).

Figure 1 provides a schematic diagram of what we have been discussing. Three words in memory, A, B, and C are being associated to a set of contextual elements from List i and a set from List j. The figure illustrates some overlap between the sets of elements for Lists i and j because, in general, there is reason to believe that the contextual elements for a particular list will not all be unique. These overlapping elements do not serve to discriminate List

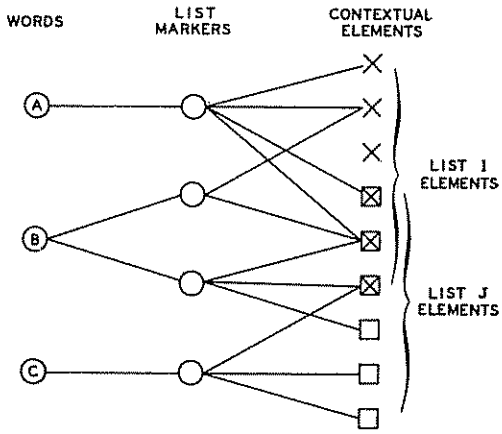

FIG 1 A hypothetical state of marking of three words A, B, and C, with respect to two lists i and j

i from List *j*, but they could help discriminate both lists from a third list. In Figure 1, Word A has been successfully associated to a List *i* marker, Word B to both a List *i* and a List *j* marker, and Word C to just a List *j* marker. The list markers serve to keep track of the lists in which a particular word occurred.

One may ask why we have inserted, in Figure 1, list markers between the contextual elements and the words. Why not simply assume that the subject associates the contextual elements directly to the words and eliminate list markers entirely? The reader should keep in mind that the point of introducing associations to context is to provide a means for keeping track of the occasions and the lists in which particular words appeared. This would be difficult to implement on the basis of direct associations between the word and the contextual elements. How would the subject sort out which contextual elements belonged to which list? But even greater difficulties arise for this alternate hypothesis when we consider the problem of how a subject would keep track of what happens within a single list. For instance, if a word appeared twice in a single list, we expect that a subject could often report that fact and describe the contexts prevailing at each of the presentations. According to the model in Figure 1, the subject would have two distinct list markers associated to the word. Each of these list markers would have associated to it a different subset of the contextual elements. Thus, when tested, the subject could retrieve two distinct list markers and therefore judge that he had seen the word in two contexts. Moreover, by means of the contextual elements associated to each of the list markers, he might be able to describe something about the two contexts. For instance, Hintzman and Block (1971) showed that the subjects had considerable accuracy in reporting both serial positions of a word that occurred twice in one list. On the other hand, consider the predicament of the subject if all he had was direct word to context-element associations. If the word was presented twice, more contextual elements would be associated to the word, so the subject might be able to judge that the word had been presented twice. However, he would have no way of sorting out which contextual elements belonged to which presentation. To reiterate, the point of introducing list markers in an association theory is to give the subject a means of keeping different events distinct in his memory.

The notion of list markers developed above is very similar to a model suggested by Hintzman and Block (1971). They also briefly considered an alternate model in which item repetition is reflected by multiple copies or replicas of each word (see also Bernbach, 1969; Bower, 1967). Within such a model, the contextual elements would be directly associated to the word without an intervening marker label, but to a different replica of the word each time it was presented. This approach may appear to eliminate the need for the list markers and the association between the words and the list markers. However, that appearance is deceiving. In such a multiple-copy model, it is necessary to postulate some means of retrieving the many copies from a single phonemic representation of the word (e.g., in order to estimate its frequency of presentation in a list). When such a model is elaborated with the necessary retrieval mechanism, it turns out to be isomorphic to ours: The prototypical phonemic representation of the word corresponds to the word in Figure 1; the different copies correspond to the different list markers; and the probability of retrieving a copy from a phonemic representation corresponds to the probability of forming an association between the word and a list marker.

FORMAL MODEL FOR RECOGNITION JUDGMENTS

The foregoing assumptions provide us with a formal model for recognition judgments that is indistinguishable from Bernbach's (1967). In deciding whether a test word was in List *n*, the subject evaluates the item with respect to how much evidence there is for the word's membership

in List n. This is done by assessing the number of List n elements associated to the list marker. During study of any word in List n, there is probability α that a List n marker is associated to the word (α would be assumed to vary with study-time and similar "strengthening" variables). If a List n marker is associated to the word, the number of List n elements associated to the node of that marker will be binomially distributed with parameters k, the number of elements in the List n population, and θ, the probability that a List n element is associated to the list marker (see our earlier assumptions). For large k, it is well known that a binomial distribution approximates to a normal distribution; so such normal approximations will be used throughout this paper. We will let $f_m(x)$ denote this probability distribution of x, the amount of evidence toward List n, for a test item that has been successfully tagged with a list marker. In the following we will refer for convenience to x as the scale of "List n evidence."

We let $f_u(x)$ denote the comparable distribution of evidence for List n for items *not* presented on List n as well as for items presented on List n but not successfully associated to a List n marker, which event occurs with probability $1 - \alpha$. One basis for $f_u(x)$ having nonzero values could be overlap or generalization between successive list contexts in which the test item occurred, and hence confusions. For such reasons, a test item may be judged falsely as occurring on List n when in fact it occurred on earlier lists.

The distributions $f_m(x)$ and $f_u(x)$, and the parameter α, are precisely those required in Bernbach's model. The probability distribution of list evidence for an item presented in the ith list is the probability mixture

$$f_i(x) = \alpha f_m(x) + (1 - \alpha) f_u(x).$$

The distribution $f_i(x)$ is definitely not normal, depending as it does on α and the means and variances of $f_m(x)$ and $f_u(x)$. Figure 2 illustrates some possible distributions for $f_i(x)$ when f_m and f_u have equal variances.

The ingredients for a "statistical decision theory" analysis of recognition memory are now at hand. The distributions $f_u(x)$ and $f_i(x)$ serve, respectively, like the noise and signal-plus-noise distributions in the theory of signal detectability. So the whole set of facts outlined previously regarding yes–no and multiple-choice recognition procedures and memory-operating characteristics are within explanatory reach of the model. The model also extends in the customary way to handle confidence ratings regarding an item's membership in List n. It is supposed that the subject assigns a rating dependent on how much evidence (number of associated List n elements) relevant to List n membership is retrieved by the test item.

The foregoing remarks develop our model for recognition and list differentiation, and it is directly relevant to such data. However, we are here concerned with the recognition process as a monitor or editor for free recall. Following input of a word list, the cue for recall initiates some search and retrieval mechanisms which begin implicitly producing plausible list candidates, items which might be on List n. These implicit candidates are then input to a recognition process to assess evidence for their List n membership. If that evidence exceeds criterion, the candidate is designated a List n item and it is given in overt recall.

RETRIEVAL PROCESS

Having completed our description of the recognition model, we will now outline our model of the retrieval process which generates implicit candidate words for possible overt recall. It is simply hopeless to imagine a random mechanism which searches haphazardly and serially (or in parallel, for that matter) through all possible memory locations, hoping to stumble across a few of the list words. Our memories are very large and not well organized for such random searches. One obviously needs a *directed* search process, and the model to be described accomplishes this by searching only those associative pathways which have been recently tagged as useful for retrieving the word set under considera-

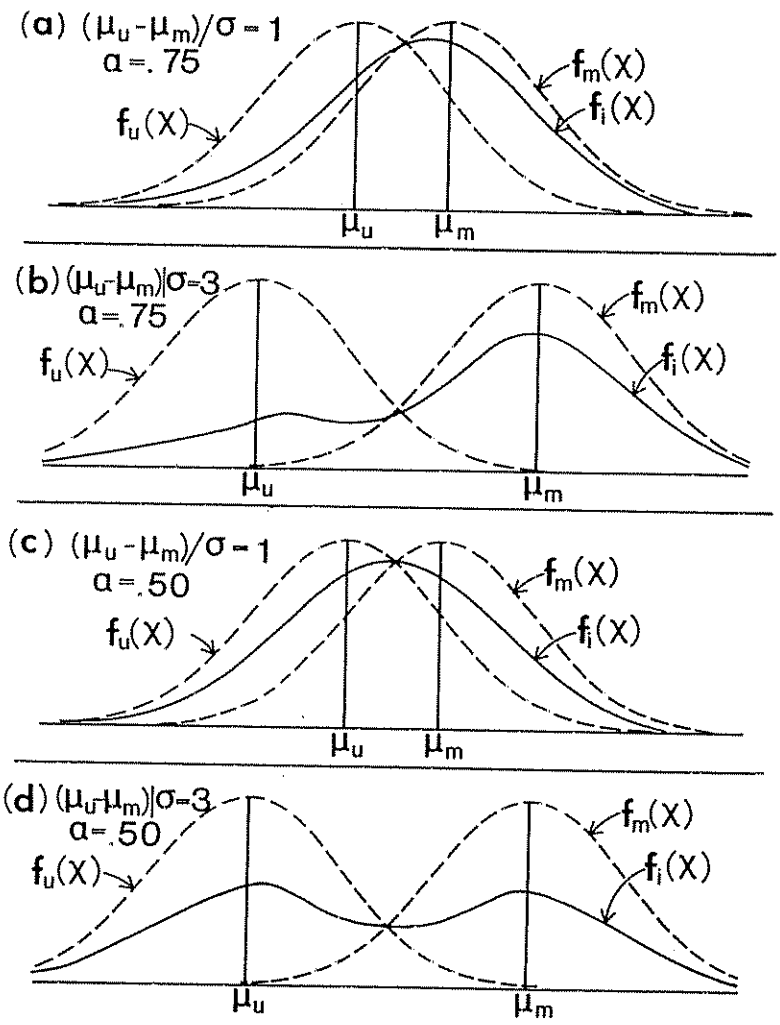

FIG. 2. Possible forms of $f_i(x)$, the distribution of List i evidence for words from List i, as a function of the probability of association to the list marker α, and the distance between the distribution for marked words, $f_m(x)$, and the distribution for unmarked words, $f_u(x)$.

tion. This model is realized in a computer simulation program described by Anderson (1972). Since its particular task environment and testing ground was free recall of word lists, it has been dubbed FRAN, an acronym for Free Recall in an Associative Network.

FRAN starts with a preexperimental network of associations among a large set of concepts (words). The experimenter selects a subset of these words to compose a free recall list; this list is then presented to FRAN one word at a time. FRAN commences to develop a way to recall the words on this presented list (call it List n) and only these. FRAN learns which words are presented on List n by tagging the memory nodes (corresponding to the words being presented) with a List n marker. This corresponds to the recognition process we have just described. In addition to this tagging of the memory node (call it A), FRAN follows one or more associative pathways radiating out from Node A, searching for other words (memory nodes) that are also on the list. This latter in-

formation is provided by whether or not the node being examined has a previous association to a List *n* marker. If another list word is found in this associative search, then FRAN attempts to tag the associative pathway from Node A to the found target node. This tagging of a prior associative pathway proceeds just like the tagging of a memory node, namely, by establishing an association to the List *n* marker. These list tags on associative pathways are what will later guide the search process; a tagged pathway essentially delivers a later directive to the search process to follow out this pathway to get from recall of Word A to recall of other list words.

There is one further process of significance occurring during the study phase of a free recall experiment and that is construction of a sublist called ENTRYSET. The list of words serve as special "starters" from which FRAN can begin her chains of recall during output. Technically, a word on ENTRYSET is simply one which receives an association from a special memory node called LIST *N*, which we may consider as a prototypical representation of List *n*. The number of "recall starters" on ENTRYSET is assumed to be very limited (in the current program, only three items). The program has a set of crude heuristics for eventually converging upon those items which are most "central" in the associative network in that they lead to recall of the largest number of list words.

In the usual free recall experiment, during output FRAN first dumps out the three to five recent list words in her short-term memory. She then uses these plus the items on ENTRYSET to start her search of long-term memory for further words to recall. This search involves following the tagged associative pathways radiating out from each of the starter nodes in turn, with each search proceeding in a "depth first" manner. During this directed associative search, FRAN considers many words, testing each for recognition (i.e., whether it has an association to a List *n* marker). If a retrieved word is recognized, it is overtly recalled.

The reader will note that context information is being used in three different ways in FRAN. First, via the list marker (collections of context elements), retrieval of context from a test node underlies correct identification of List *n* words, thus forming the basis for list discrimination. Second, via the LIST *N* node (i.e., the prototype of the List *n* context), the context also has the status of a "stimulus" initiating recall chains from the ENTRYSET words. These two uses of context are not original with us. Similar notions were suggested by Norman and Rumelhart (1970); a standard assumption of interference theory is that context cues serve as stimuli governing emission of a response (Keppel, 1968; McGovern, 1964). We have expanded these notions and provided an explicit representation and model for them. The third use of context, to tag associative paths in the memory network, is unique to our learning and retrieval model. Such second-order associations, that is, an association between an associative link and a memory node (a List *n* marker), have not been used in the more traditional associative theories. However, the second-order association is really only a labeling of a lower order association, and most current theories of memory (e.g., Quillian, 1969; Rumelhart, Lindsay, & Norman, 1972) do involve labeling of associations with such semantic relations as category membership, superordination, opposition, and the like. In FRAN, labeling preexisting associative pathways with a List *n* tag provides the mechanism for guiding FRAN's memory search during recall. Since the two kinds of tagging, of memory nodes versus associative pathways, are independent, it is the case that word recognition in FRAN is independent of the retrieval processes. This independence of the two processes will be important in the experiments that follow.

The above description of FRAN is necessarily brief and incomplete, focused on only those aspects pertinent to the two-process recall hypothesis. Papers by Anderson (1972) and Bower (1972a) should be consulted for further details of the theory, the simulation program, and a sam-

pling of results which the theory explains and those with which it has difficulties. In our opinion, it is the most viable theory of free recall currently available in terms of its range of applicability.

EXPERIMENT I

Experiment I is one of those research curiosities begun for irrelevant reasons but which turned out to provide the stimulus for this program of research and theorizing. We presented subjects 15 successive lists of 16 words, each list selected in a pseudo-random manner from a master set of 32 words. As a consequence, lists after the first had partial overlap in membership with earlier lists. After the presentation of each list, the subjects were required to recall only those words that were in the most recent list. According to our retrieval model, the ability to retrieve the full set of 32 words should improve with repeated study trials because there is more time to discover and tag associative pathways between the words.

However, the model makes a markedly different prediction about the recognition subprocess of free recall. The implicit candidates for recall generated by the retrieval component will include many or most of the master set of 32 plus some associatively related items. The task of the recognition component is to select which of these words occurred in the most recent list. This it does by checking to see which of the candidates are tagged with list markers referring to the most recent list. Since a different set of 16 words must be recalled on each trial, the subject can only succeed if he associates a different list marker with the to-be-recalled words on each trial (list). Tagging with list markers is conceived essentially as a paired-associate task in which the word functions as the stimulus term and the marker functions as the response term. Thus, the tagging of a single word appearing in successive lists exemplifies the classical A–B, A–C paradigm for negative transfer. One might assume that the amount of negative transfer accumulates as the subject learns A–B, then A–C, then A–D, etc. Similarly, we assume negative transfer grows as the subject learns associations to list markers as in A–List 1, A–List 3, A–List 5, and so forth. We have not found a cumulative negative transfer study in the literature that exactly parallels our situation, but Underwood (1945) has demonstrated that proactive inhibition increases as the same stimulus term is paired with progressively more responses. It is reasonable to assume that negative transfer would also increase as the same stimulus is paired with more responses. For instance, such a result is predicted by the assumptions about response competition proposed by Kjeldergaard (1968).

In this experiment, each word will be paired with a new list marker on every trial that it appears for study. Therefore, the probability of successfully tagging a word should decrease across lists, with a consequent decrease in the subject's ability to recognize to-be-recalled words. Thus, a dissociation of the two subprocesses is expected, with retrieval improving and recognition of implicitly retrieved words deteriorating across successive lists. Such a dissociation would emphasize the utility of conceiving of free recall in terms of the two-process model. It would be impossible for the threshold model, which relates both recognition and retrieval to the same factor of "strength," to account for such a dissociation.

Method

From a master set of 32 common concrete nouns, 15 lists of 16 words were generated. Words were assigned randomly to lists with the constraint that the overlap of each list with each other list would be eight words. Such independent, multiple overlapping of lists was done so that the subjects could not use membership of a word in one list as any indication of its membership in any other list. Therefore, decisions about membership of a word in the current list would have to be based solely on the word's rating on the scale of evidence for that list. By the sixth list, every word in the master set had occurred in at least one list. By the fifteenth list, a particular word had occurred in anywhere from 4 to 11 of the lists. The order of words within a particular list was randomly determined. The same sequence of 15 lists was presented to all of the subjects. The words were slide projected one at

a time for two seconds. Immediately after presentation of the list, the subjects wrote their recall of the most recent list of words they had just studied. They had 135 seconds for this recall. They were instructed to recall a word only if they thought it "probable" that the word came from the last list studied. The experimenter collected the recall sheets and the procedure was repeated for the next list. Including the instructions, the experimental session lasted about 50 minutes. Seventeen subjects (6 males, 11 females; 16–22 years old) were recruited through an advertisement in a local newspaper and were paid $1 75 for their services. They were tested in two groups of size 7 and 10.

Results

Figure 3 shows the average number of words recalled from the most recent list (abbreviated R words) and separately, the number of words which were intruded from earlier lists and were not on the most recent list (N words). The number of R words recalled minus the number of N words, also shown in Figure 3, provides a conservative correction for guessing in recall. In the curve for R-word recall and in the corrected-recall curve, there appears to be an initial improvement in recall followed by deterioration. Using orthogonal polynomials, an equation involving the first six powers was fitted to the corrected recall over the 15 lists. The improvements in the fit due to the linear and quadratic components of the curve were significant ($F = 8.35$, $df = 1/8$, $p < .05$, for the linear component; $F = 12.36$, $df = 1/8$, $p < .01$, for the quadratic component). However, the addition of the four higher order polynomials did not significantly improve the fit of the theoretical curve ($F = 2.03$, $df = 4/8$, $p > .10$). The smooth curve in Figure 3 describes the best fitting quadratic equation for the corrected recall. There is considerable variability of the observed points about the quadratic curve. This variability may be attributed to differences among the individual word lists. The theoretical interpretation of this overall rise then fall in recall is that the improvement in item retrievability predominated initially, but that after retrievability had reached an upper limit, the degradation of list differentiation continued and recall deteriorated as a consequence. This is because recall is monitored or edited by list differentiation decisions. This interpretation is further tested in the following experiments.

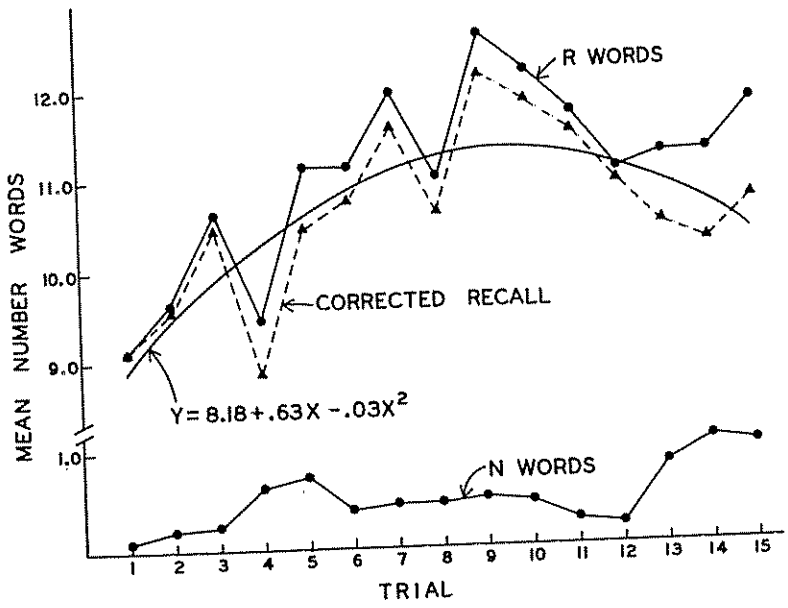

FIG. 3. Mean number of words recalled in Experiment I as a function of trials.

EXPERIMENT II

A simple challenge to this interpretation of the results is to question whether the results have anything to do with the particular experimental manipulation employed. Quite possibly the initial improvement may simply reflect "learning to learn" and the later decrement may only reflect progressive fatigue or loss of motivation of the subjects. The obvious control, then, is to repeat the last experiment but with a completely different list (i.e., no words repeated) on each trial. There should be negligible across-trial change in retrievability because the subject is learning new items on every list. Also because new words are being studied on each trial, both the stimuli and the responses for the hypothetical paired-associate task underlying the list marking are different on each trial. Therefore, there should be little, if any, negative transfer in identifying the most recent list. Hence, if the results of Experiment I were really due to the pseudo-random repetition of words across lists, then neither the initial increase nor the later decrease in recall should be observed in this second experiment.

Method

Twenty lists of 16 words were created by sampling randomly without replacement from a master set of 320 common concrete nouns. The same 20 lists were used in the same order for all subjects. The procedure for testing was identical to that in Experiment I. This experiment, including instruction, lasted about 70 minutes. Twenty-one subjects (10 males and 11 females; 18–22 years old) served in this experiment as partial fulfillment of a requirement for the introductory psychology course at Stanford. They were tested in two groups of size 10 and 11.

Results

Figure 4 gives the mean number of words recalled in Experiment II as a function of the trial number. Strictly speaking, only the first 15 trials are relevant to a comparison with Experiment I. Over these trials there is neither a significant linear nor a significant quadratic trend. However, if all 20 trials are considered, the linear trend becomes significant ($F = 8.04$, $df = 1/17$, $p < .05$) although the quadratic trend is still not significant ($F = 1.87$, $df = 1/17$). The smooth curve in Figure 4 describes the best fitting quadratic equation to the data over the 20 trials. The nonsignificant quadratic trend in Figure 4 is exactly opposite to that of Figure 3—that is, recall is worse in the middle trials of the experiment. However, the only significant effect is the linear trend which indicates something of a warm-up effect with recall improving slightly toward the end of the experiment. While the improvement across trials is not very substantial, it contrasts in a minor way with the results of Murdock (1960) who found no improvement across unrelated lists in single-trial free recall.

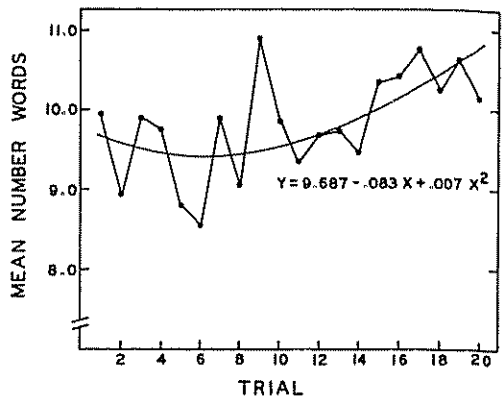

FIG. 4. Mean number of words recalled in Experiment II as a function of trials.

EXPERIMENT III

Although the preceding experiments confirmed expectations, the hypothesized mechanisms would be more credible if we could observe separately the decay in recognition and the increase in retrievability rather than viewing only their combined effect on recall. One would then be able to determine whether improvement in retrievability first dominated but was later overcome by deterioration in list recognition. The third experiment provides separate measures of retrieval and list recognition.

Method

The subjects in this experiment studied the identical sequence of 15 lists as used in Experiment I. However, they were required to try to recall the

entire master set of 32 words on each trial, not just the 16 presented in the most recent list. The subjects were further required to rate on a 6-point confidence scale whether or not they thought each word they recalled came from the most recent list they had studied. The rating scale ranged from "1 = confident the recalled word was on the most recent list" up to "6 = confident the recalled word was not on the most recent list." Hence, to use the terminology of Experiment I, a low confidence rating indicated that the subject thought the word was an R word, whereas a high confidence rating indicated that he thought it was an N word. By examining recall of R words without regard to their rating, changes in retrievability may be monitored; by examining the confidence ratings, changes in the subject's ability to recognize words can be monitored.

The subjects were given 165 seconds to write their recall and to record confidence ratings beside each of the words recalled. Detailed instructions and illustrations on use of the confidence scale were given since a pilot study had indicated that many subjects would misunderstand the instructions and use low numbers just for those words which appeared for the first time in the most recent list. Including the instructions, the experimental session lasted about 75 minutes. Twenty-four subjects (10 males, 14 females; 16–23 years old) were recruited through an advertisement in a local newspaper. In a two-hour experimental session they participated in this and another unrelated experiment which followed, and they were paid $3.50 for the total session. They were run in three groups of sizes 6, 8, and 10. Six subjects were excluded from the analysis because of failure to use the recognition scale properly on some of the trials.

Results

An initial question is whether the earlier results of Figure 3 have been duplicated under these altered conditions. In this experiment, a rating of 1 indicated that the subject was certain that he had seen the word in the list just studied, and a rating of 2 indicated he thought it probable that the word came from that list. Since the subjects were instructed in Experiment I to recall only the words they thought "probably had occurred" in the last list, the words recalled and rated 1 or 2 in Experiment III should be comparable to those recalled in Experiment I. Figure 5 shows the frequency with which R words (in the most recent list) were recalled and rated 1 or 2, and the frequency with which N words (not in most recent list) were recalled and rated 1 or 2. Figure 5 also shows recall corrected for guessing by subtracting the mean for the N words from the mean for the R words. The data represented in

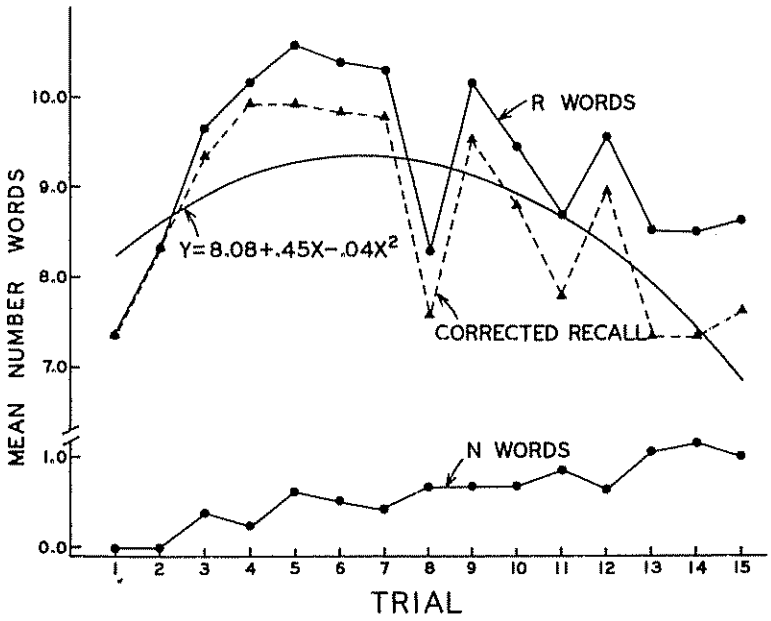

FIG. 5 Mean number of words recalled and rated "1" or "2" in Experiment III as a function of trials

Figure 3 and the data represented in Figure 5 are quite similar. As in Experiment I, an equation involving the first six powers was fitted by means of orthogonal polynomials to the corrected recall over trials. The improvement in the fit due to the linear component was marginally significant ($F = 5.06$, $df = 1/8$, $p < .10$); the improvement in fit due to the quadratic component as quite significant ($F = 9.54$, $df = 1/8$, $p < .025$); the four higher order polynomials did not significantly improve the fit ($F = .90$, $df = 4/8$). The smooth curve in Figure 5 describes the best fitting quadratic equation to the corrected recall. That quadratic equation confirms an initial rise and subsequent fall in the corrected recall. Thus, we may conclude that the procedural changes in this experiment have not altered the basic processes that were occurring in the first experiment.

We can now determine whether, as hypothesized, retrievability improves while list recognition deteriorates across trials. Figure 6 presents the mean number of words recalled, while Figure 7 shows the mean confidence rating of the words recalled. In these figures, the data are classified into four groups according to whether or not the word was in the most recent list and according to the number of times the word had appeared before that list. The eight R words that had been presented fewer times than the mean number for those in the most recent list were classified as LR (less frequent, recent), and the other eight presented more than the mean, MR (more frequent, recent). When frequencies were tied, the words assigned to the LR group were those for which the lag between the current presentation and the next most recent was the longest. On a similar basis, the N words were subclassified as LN and MN.

In Figure 6 it is clear that recall of R words increases with negative acceleration to an asymptote, confirming the hypothesis about the improvement in retrievability. On the other hand, Figure 7 shows that the mean difference in recognition ratings between R words and N words steadily decreases as a function of trials, confirming the hypothesis about the degradation in list recognition. Thus, the interpretation of the results of Experiment I as arising through the dissociation of retrieval and recognition has been confirmed.

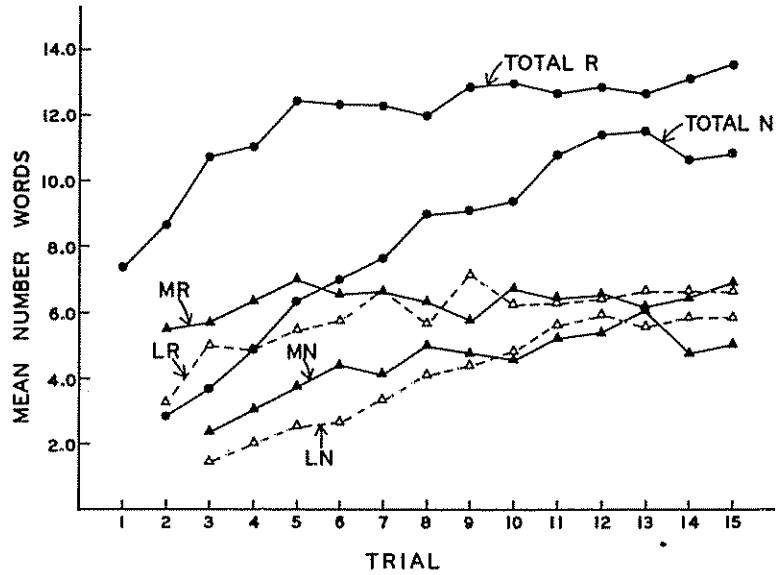

FIG. 6. Mean number of words recalled in Experiment III as a function of trials

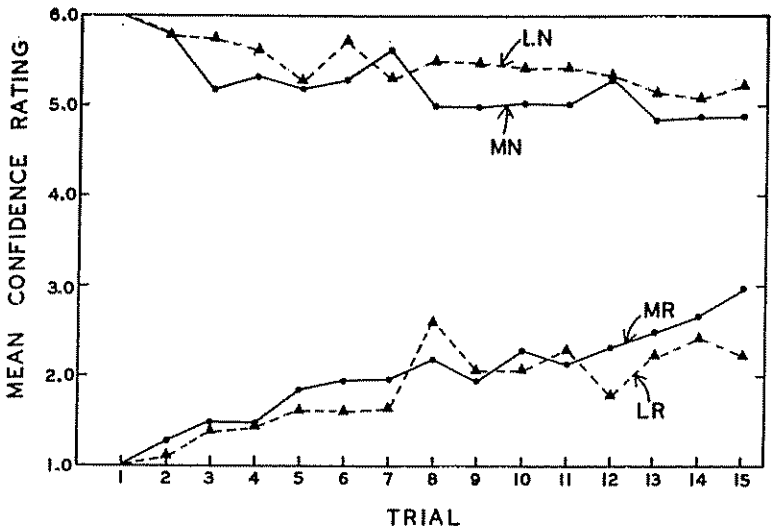

FIG 7. Mean rating of the words recalled in Experiment III as a function of trials.

Several auxiliary hypotheses arising from our theoretical position may also be examined. First, our model relates improvements in retrieval over trials to the fact that the subject has more opportunities to locate and learn associative paths to access the words. Therefore, those words best recalled should be those that have appeared in the most prior lists—a simple frequency effect. Inspection of Figure 6 confirms that the more frequently occurring words (represented in the curves labeled MR and MN) were initially better recalled than the less frequently occurring words (curves LR and LN). As retrievability approached its asymptotic level, the differences between more and less frequently occurring words diminished, as is to be expected. These conclusions from a visual examination of Figure 6 are fully substantiated by statistical analyses.

Second, on the hypothesis that negative transfer increases as the same word is paired with more list markers, one would predict that the mean confidence rating should be lower (i.e., more correct) for those R words that occurred less frequently. The fewer prior list markers associated to a word, the more "novel" it is, the less the negative transfer in establishing the new "word → List i" association. Figure 7 shows 14 trials over which a comparison of confidence ratings can be made between MR and LR words. Of these 14 comparisons, 11 trials exhibit the predicted inequality. To confirm the reliability of this difference, the mean ratings across these 14 trials for the MR and LR words were computed for each of the 18 subjects. The difference in average ratings between the two types of words is highly significant ($t = 4.41$, $df = 17$, $p < .001$). This finding is all the more impressive when one realizes that just the opposite ordering of the MR and LR curves would be predicted on the basis of a strength model which relates recognition to frequency, recency, and duration of exposure.

Considering the data at the top of Figure 7, on 12 of the 13 relevant trials, the less frequent N words were better differentiated (have higher scores) than the more frequent N words. In other words, more frequently presented N words were judged as more likely to have been on the most recent list than were less frequently presented N words. A t test confirms the significance of this difference ($t = 3.31$, $df = 17$, $p < .005$). It is not clear how to interpret this result. Perhaps it could be explained by the mechanism of generalization of contextual elements between lists.

The more prior lists in which an item has appeared, the greater the probability would be that one of the list markers associated would be mistaken as a list marker for the current list. The problem with the generalization mechanism is that it would predict the opposite ordering of the MR and LR curves. To explain why the LR curve is below the MR we would have to make the added and unmotivated assumption that the negative transfer in list tagging outweighed the list generalization factor. In any event, the conclusion to be drawn from Figures 6 and 7 is that, in this paradigm, retrieval is directly and recognition is inversely related to frequency of exposure.

Model Testing

In the introduction, we stated that the distribution of values for nonpresented N words on the scale of List i evidence is the normal $f_u(x)$, but the distribution for the R words is the nonnormal $f_i(x) = \alpha f_m(x) + (1 - \alpha) f_u(x)$. A relatively simple test of this mathematical model is possible. The probability that a particular N word is rated with a confidence exceeding j provides an estimate of the standard normal deviate corresponding to the jth criterion point. In this way, the data from the N words provide estimates of the intervals between the five criterion points. A grid search can then find the probability of tagging, α, and of the mean of the distribution of tagged words, μ_m, that will yield the best fitting distribution $f_i(x)$ for the R words. The best fitting parameters are those that yield predicted frequencies of the confidence ratings that deviate minimally from the observed frequencies as measured by a chi-square test.

One could test the statistical significance of the deviations between observed and predicted frequencies as a goodness-of-fit statistic. However, a few large chi-squares may be expected even if the model were essentially correct because the estimates of the confidence intervals are only approximate and are subject to random error. A fairer test of the model would find the minimum chi-square estimates of seven free parameters, namely, the five criterion points (for the confidence ratings), α, and μ_m. Unfortunately, the computing cost to find seven parameters is many orders of magnitude greater than the cost to estimate two parameters. So, instead of the best absolute test of the model, we will present a comparative test of the model. We will compare our model with a plausible alternative theory, the traditional recognition model that presumes that one normal distribution underlies words from the list and another underlies words not from the list (e.g., Parks, 1966; Wickelgren & Norman, 1966). Since that model assumes that the two distributions have the same variance, the grid search needs only to estimate the mean of the likelihood distribution for R words, d'. The value of d' would characterize a memory-operating characteristic in the traditional signal-detectability analysis.

The test proposed is impossible for Trial 1 because there are no N words recalled; also not enough N words were recalled on Trial 2 to provide reliable estimates of the intervals between the criterion points. Therefore, comparisons of our model with the traditional one will be confined to Trials 3 through 15. Table 1 presents for these trials the estimated values of the criterion points (c_1,c_2,c_3,c_4,c_5), α, μ_m, and d' as well as the chi-square measures for deviations of our model and of the traditional signal-detectability model. For every trial, the predictions of the traditional signal-detectability model lead to large chi-squares, while the assumptions of our model results in smaller chi-squares. Only 4 of the 13 chi-square totals for our model are significant at the .05 level. Notice that as predicted by the principle of negative transfer, there is a decrease across trials in α, the probability of successfully tagging (associating) a word with a list marker. Using the weights suggested by Abelson and Tukey (1963) to test for monotonic trend, the decrease in α across trials is found to be highly significant ($F = 24.32$, $df = 1/11$, $p < .001$). Although less obvious, there is also a marginally significant decrease in the value of

TABLE 1

PARAMETER ESTIMATIONS AND CHI-SQUARE DEVIATIONS FOR THREE MATHEMATICAL MODELS

Trial	Criterian points					Traditional model		Proposed model			Tradition model with variable σ_i		
	c_1	c_2	c_3	c_4	c_5	d	($df = 4$) χ^2	α	μ_m	($df = 3$) χ^2	d'	σ_i	($df = 3$) χ^2
3	1.88	1.34	1.18	1.14	.82	2.60	25.39	.91	3.42	.50	4.44	2.38	.80
4	2.33	1.70	1.37	1.08	.61	2.85	54.13	.93	3.63	2.54	4.75	2.40	4.59
5	1.60	1.25	1.13	.90	.51	2.20	32.80	.85	3.08	5.70	3.58	2.34	3.72
6	1.60	1.48	1.37	1.23	.76	2.16	20.99	.95	2.38	16.98	2.68	1.48	15.26
7	1.88	1.55	1.51	1.23	.69	2.34	29.03	.89	2.98	4.79	3.14	1.83	4.26
8	1.88	1.44	1.23	1.08	.47	1.86	36.75	.79	2.54	4.54	2.34	1.82	1.15
9	1.70	1.44	1.10	.93	.37	2.05	46.68	.83	2.81	15.49	3.09	2.14	10.43
10	1.88	1.48	1.10	.86	.40	1.95	62.44	.77	2.91	6.16	2.77	2.17	2.30
11	1.64	1.44	1.18	.82	.47	1.67	36.44	.78	2.37	8.50	2.05	1.64	9.76
12	2.17	1.60	1.20	.92	.45	2.09	103.64	.79	2.99	22.50	2.93	2.31	12.57
13	1.60	1.34	1.04	.64	.11	1.62	35.66	.77	2.34	1.76	2.02	1.72	.66
14	1.44	1.23	.90	.63	.35	1.45	29.34	.70	2.27	5.88	1.79	1.69	4.15
15	1.64	1.34	.97	.66	.40	1.51	52.76	.65	2.69	7.55	1.97	2.15	5.60

μ_m ($F = 6.14$, $df = 1/11$, $p < .05$). This decrease in the distance between the distributions $f_m(x)$ and $f_u(x)$ may reflect increasing confusion between the list marker from the most recent trial and list markers from earlier trials because of the generalization of contextual elements.

Perhaps a more appropriate alternative model would be the traditional signal-detectability theory with the variance of the distribution for R words as a free parameter to be estimated from the data. This model has the advantage of equaling ours in the number of estimated parameters. The fit of this expanded model is shown in the last columns of Table 1, showing the estimates of d' and σ_i, the mean and standard deviation of the $f_i(x)$ distributions. In terms of goodness of fit, there is little basis to choose between this model and ours. This outcome was not entirely unexpected. Inspection of Figure 2 shows that the likelihood distribution $f_i(x)$ is sometimes sufficiently normal to make difficult the discrimination between it and the true normal distribution predicted by the traditional model for the R words. Furthermore, we are dealing with an average of likelihood distributions. There is a different distribution $f_i(x)$ for every word because each word has appeared in a different set of lists and therefore has a different amount of negative transfer associated with it as a stimulus. The average of these many individual distributions will be considerably more normal than any of its constituents.

It is pertinent to examine the changes in the parameters σ_i and d' across trials for this version of the traditional signal-detectability recognition model. The estimates of d', the distance between the means of the distributions, show a fairly consistent decrease across trials. One difficulty with this traditional signal-detectability model is that it provides no theoretical base from which one might predict this decrease in d'. Also, the parameter σ_i is fluctuating erratically from trial to trial. These fluctuations are probably a consequence of the minimum chi-square estimation procedure. Occasionally, events with low theoretical probabilities will have frequencies that are considerably deviant. The estimation procedure gives considerable weight to such deviant low-probability events. In an attempt to minimize deviations of predicted from observed frequency for such an event, the predicted distribution can be markedly altered from one trial to the next.

Discussion of Experiment III

The results of Experiment III showing an improvement in recall alongside a de-

cline in recognition, bring into sharp focus the failing of the "threshold" theory which relates recognition and free recall to a single theoretical construct, such as strength or familiarity. We have dissociated two behavioral measures which that theory claims are coupled by the nature of the system. Such data, along with other arguments advanced in the introduction, render the threshold theory untenable as an account of multilist free recall.

In a similar vein, our results increase the implausibility of a simple strength of familiarity theory of list recognition. The basic problem with such theories is that the subject's judgment of the list membership of an item is not wholly predictable from that item's "composite strength." Both in Experiment I and Experiment III, all 32 words of the master set rapidly became quite "familiar" to the subject. And yet, as Figures 3 and 5 demonstrate, the subjects still showed considerable ability to recognize (discriminate) R words from N words. This discrimination varied only moderately with the frequency of an item's presentation prior to the current list being discriminated.

One might suppose that this R versus N discrimination could be effected by a rapidly decaying "short-term" strength attached to items, such that only items in the most recent list would have strengths exceeding a specific criterion. But this account then finds inexplicable the results of Experiment III in which subjects not only recalled but correctly labeled the N words that were not presented on the most recent list. How are such N items being discriminated from the vast pool of known common nouns in the subject's long-term memory which are not being used at all in the experiment? Such questions, and the several others posed in the introduction, overstrain and discredit the simple strength or familiarity assumptions about list recognition. In place of simple strength, we propose that subjects use something like "list markers" for indexing the lists in which an item has appeared.

One piece of evidence that appears to upset the picture established by the first three experiments comes from a free recall study reported by Ehrlich (1970). He gave his subjects 10 trials of free recall learning on a 20-word list. By the tenth list, recall had reached a near maximum of 19 out of 20. Then the subjects were switched to a series of 10 single-trial free recall trials on semirandom subsets of the original set of 20, much in the manner of Experiment I. Since retrievability for the set of 20 had apparently asymptoted, one might expect to see a continuous decay in the number of words recalled as a consequence of the build-up in negative transfer for list identification. However, the level of recall remained constant over the 10 sublists in Ehrlich's experiment. It is possible that over the initial 10 trials on the whole list in his experiment, negative transfer may have reached its asymptotic level, and hence there would be no more deterioration in the list-tagging process over the last 10 trials on the part lists. By the fifteenth trial in our Experiment I, subjects had seen each word in a mean of 7 lists; by the first part-list trial of his experiment, subjects had seen each word in 10 lists.

The point is that we cannot expect the decay in recognition to continue forever. If our experiment had been extended for more trials, recognition performance would surely decay eventually to some asymptotic level. The two facts that both retrievability increases to an asymptotic level and that recognition also decreases to an asymptote implies that the phenomenon of a rise and fall in recall found in Experiments I and III is in reality a rather delicate matter. It depends on retrievability increasing faster than recognition deteriorates, but asymptoting sooner. In a pilot attempt to replicate Experiment III with altered procedures (more time for recall, different rating scale, different subject population), we replicated the increase in retrieval displayed by Figure 6 and the deterioration of recognition displayed by Figure 7. However, the quadratic trend in Figure 5 (the rise and fall in recall) was relatively slight and in fact nonsignificant statistically.

EXPERIMENT IV

Our model assumes that the processes underlying recognition in free recall are

identical to those processes which underlie performance in a pure recognition task. However, evidence has not yet been presented to support this assumption. The next experiment does this by examining pure recognition performance with the same 15 lists that were used in Experiments I and III. If our negative-transfer assumption is correct, we should find the same deterioration in recognition with frequency of presentation as was obtained in Experiment III.

Method

The identical sequence of 15 lists were presented as in Experiments I and III, but with two different testing methods. For the first test method, immediately following the presentation of the list on any trial, the subject had to add mentally 15 single-digit numbers presented at a two-second rate. This manipulation was intended to prevent the subject from recognizing any test word from his short-term memory (see Anderson, 1972). The list marker theory of recognition only applies to items not in short-term memory at the time of testing. In FRAN, any item still in short-term memory at the time of testing is, of course, assured of perfect recognition. After this summation task, the subjects were given a sheet listing the master set of 32 words and they were to rate on a 1 to 8 scale the subjective likelihood or confidence that each word had been on the most recent list. In this experiment, a high rating indicated that the subject thought the word very probably was on the most recent list, and a low number indicated that it very probably was not on the most recent list.

For the second test method, the subject was shown the test items one at a time slide-projected at a five-second rate. The subject rated the items on the same 8-point scale as in the first test method. These two different methods of test were used to insure the generality of the results. The second method corresponds more to how we postulate that recognition happens in free recall—that is, the subject must judge one word at a time. However, there is no compelling reason to expect differences between the two methods. Thirty-one subjects (15 males and 16 females; 18–32 years old) participated in this experiment as partial fulfillment of a requirement for the introductory psychology course at Stanford. They were run in groups ranging in size from 5 to 10. Twenty subjects were tested by the first test method, and 11 by the second test method. Two subjects tested under the first test method were eliminated from the analysis because of incorrect use of the recognition scale.

Results

No systematic differences appeared in the confidence ratings as a function of the method of test, so only the pooled results will be presented. Figure 8 presents the mean confidence ratings for the different types of words across trials. Note that the confidence scale in this experiment was reversed from Experiment III in that in this experiment high numbers mean the subject thought the word was on the most recent list. Also an 8- rather than a 6-point confidence scale was used. It does appear that the mean difference in rating between

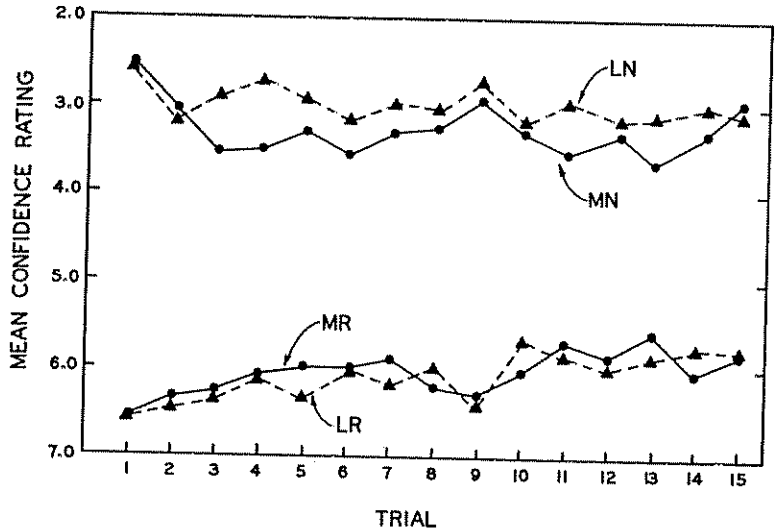

FIG. 8. Mean rating of the words in Experiment IV as a function of trials.

the R words and the N words decreases across trials. On Trials 1–5, the mean difference is 3.31; on Trials 6–10, it is 2.92; on Trials 11–15, it is 2.65. A test for a monotonic decreasing trend in the difference between the R and N words is highly significant ($t = 6.10$, $df = 13$, $p < .001$). However, the deterioration in recognition displayed in Figure 8 is not as dramatic as that in Figure 7 from Experiment III. As in Experiment III, those words presented less frequently (represented in the curves LR and LM) are the more accurately identified. Thus we have replicated the findings of Experiment III with respect to recognition.

From the data of this experiment, we can discover what free recall would be like in an experiment in which retrieval was perfect from the start but recognition was still subject to deterioration through negative transfer in the list-tagging process. Retrieval is perfect because the experimenter provides the subject with all the words; the subject needs only to decide which came from the most recent list. Figure 9 indicates the sort of free recall data that would obtain under the circumstance of perfect retrieval. Here we have the number of R words rated 7 or 8 and the number of N words rated similarly. These may be taken to represent the number recalled and the number intruded in the hypothetical circumstance. We have also plotted in Figure 9 the corrected recall obtained by deducting intrusions from recalls. Figure 9 is to be compared with Figure 3 from Experiment I and Figure 5 from Experiment III. As we would predict, since "retrieval" is asymptotic from the start, there is a rather dramatic decline in "recall" as negative transfer builds up and impairs the recognition component.

Thus, it would appear that we may

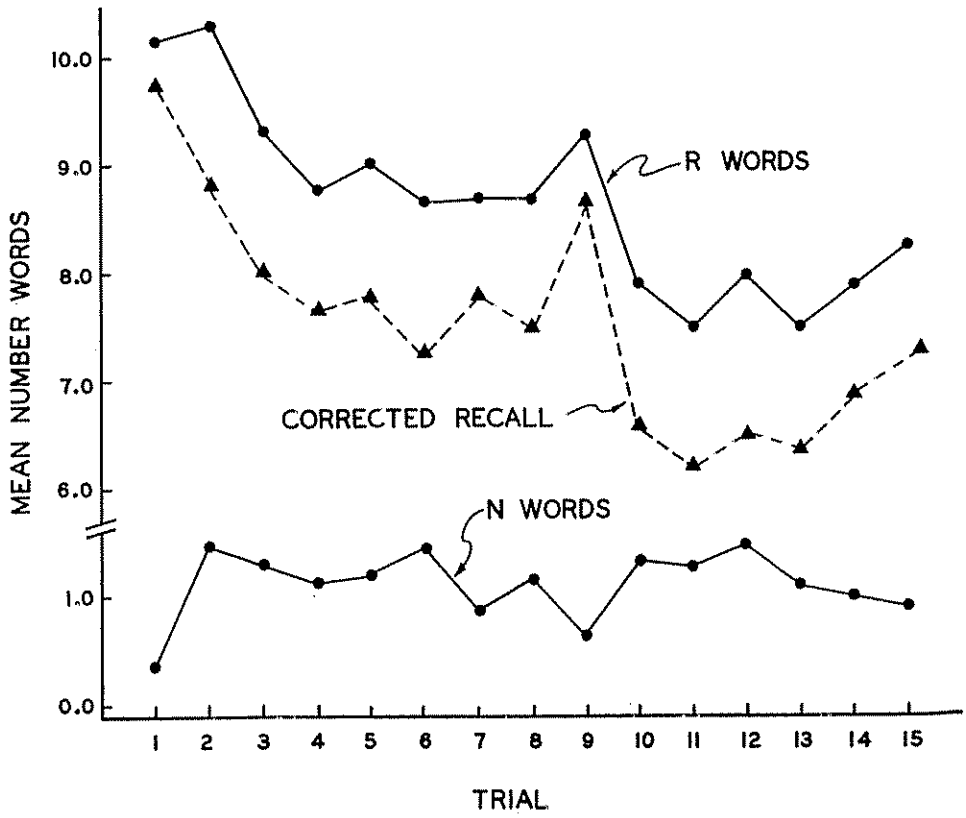

FIG. 9. Mean number of words rated "7" or "8" in Experiment IV as a function of trials.

profitably conceive of recognition and list discrimination experiments in paired-associate terms in which each word serves as the nominal "stimulus" and the List n marker as the nominal response. This functional identification leads to implications regarding negative transfer and retroactive and proactive inhibitions. This experiment has clearly shown negative transfer may be obtained in a list discrimination task. In G. H. Bower and J. R. Anderson (in preparation), further data will be reported to indicate that retroactive and proactive inhibition can also be obtained in a similar design.

GENERAL DISCUSSION

We will mention several conceptual puzzles regarding recognition memory which may have a viable solution within the framework of our theory. It should be stressed, however, that the theory was not constructed to account for these phenomena; they are rather in the form of afterthoughts. They will be given a more complete analysis in a forthcoming paper of G. H. Bower and J. R. Anderson.

Part–Whole Negative Transfer

The notion we have developed of negative transfer in the tagging of items in multiple lists may help to explain the puzzling phenomenon of negative transfer in part-to-whole or whole-to-part studies of free recall (see Tulving, 1966; Tulving & Osler, 1967). A subject pretrained with part of a free recall list will subsequently learn the whole list more slowly than a control subject pretrained on an irrelevant list before receiving the whole list. The difficulty is largely localized in the very poor improvement in recall of old (part-list) items (see Bower & Lesgold, 1969). This outcome would be predicted if there were negative transfer in associating a List 2 marker to an item previously associated to a number of List 1 markers, and if whole-list recall were monitored and edited for a List 2 tag associated to the candidate items before they were overtly recalled. Thus, part-list items previously associated to a List 1 tag would acquire List 2 tags more slowly and would thus be frequently edited out from free recall. This outcome hinges critically on the experimental subject not being aware that all part-list items are contained in the whole list. If he were to be informed of this fact, then there would be no list discrimination problem, and the monitor would recall any candidate item retrieved having either a List 1 or a List 2 tag associated to it. Thus, informed subjects should give only positive part-to-whole transfer. This is indeed the case, as has been found by E. Tulving (personal communication, 1971).

According to this analysis, the negative transfer in part–whole experiments is occurring in the recognition phase of free recall and not the retrieval phase. In fact, the theory expects negative transfer if the subject went from study of a whole list to more study on the selfsame whole list—provided he was led to believe that the first and second lists contained some different items. This result is precisely what has been found by R. M. Schwartz and M. S. Humphreys.[3]

Associatively Related False Alarms

The theoretical location of $f_u(x)$ relative to $f_m(x)$ in Figure 2 may provide a means for rationalizing the effects on recognition memory of similar or associatively related distractors. It is well known that such related distractors elicit more false-positive recognition judgments than do unrelated distractors. We may conceive of this as mediated recognition. Although the test item may not be directly marked, it may call to mind an associated list word which was marked, and on the basis of that evidence the subject may infer that the test item was on the list.

This indirect evidence, from mediated recognition, is available not only for related distractors but also for unmarked list items. Recall that during list study, it is presumed that the subject is searching out and marking associative pathways linking list items.

[3] R. M. Schwartz and M. S. Humphreys. List differentiation in part–whole free recall. Unpublished manuscript, 1971.

Hence, a list item may have several associative paths marked to other list items without itself being marked. Because mediated retrieval of a list marker can occur for either unmarked list words or associatively related distractors, it is of only partial reliability as evidence for list membership of the test item.

What the subject does with such indirect evidence will probably depend on the testing situation. For example, in multiple-choice tests, the subject would doubtless choose a directly marked test word in preference to a mediately marked test word; but he would also choose the latter over distractors having neither direct nor mediate list markings. For single stimulus or yes–no tests, allowing for mediated retrieval of list markers is equivalent mathematically to adding a constant C to the "List i evidence" scores for all items that have marked associations to items that can be independently identified as list members. This constant would be added to marked list items, unmarked list items, and associatively related distractors.

The effect of this allowance on performance depends jointly on the nature of the old study items and the new distractor items. Let us consider the base control condition to be recognition performance on a list of unrelated words tested against essentially unrelated distractors. Compared to that control, recognition would be higher for a list of highly interassociated items (e.g., members of one taxonomic category) when tested against unrelated distractors. This corresponds to what Kintsch (1968b) called "category recognition" and it is implemented in the model by increasing μ_i, the mean of the $f_i(x)$ distribution, by a constant C without at the same time shifting the distribution $f_u(x)$ for the unrelated distractors.

On the other hand, consider the case when all the distractors are in the same category (or categories) as the study items; then $f_i(x)$ and $f_u(x)$ would both be shifted up the scale by C. This is because in either case an unmarked test word of the category is likely to elicit highly associated words of that category which are marked. The result, then, would be no net change in d' or the overall recognition performance, as Kintsch (1968b) reported. In this sense, associative or categorical "organization" of the study list will affect recall but not recognition against semantically related distractors.

If the recognition test series contains a mixture of semantically related and unrelated items, then the subject is essentially dealing with three distributions having respective means of $\mu_I + C$ for old items, $\mu_u + C$ for related distractors, and μ_u for unrelated distractors. Since the decision model supposes that the subject selects a single, item-independent criterion for making his yes–no recognition judgments, the expected outcome is a higher false-alarm probability for related than for unrelated distractors.

Returning to the earlier discussion of differentiation from unrelated distractors, the importance of list-mediated recognition will depend on the average probability of retrieving indirect list evidence for an unmarked list item during testing. That probability will depend in turn on the average degree of interitem associations established by the list-studying conditions. For example, if free classification of a set of items into more tightly packed categories improves free recall (Mandler, 1967), then the number of categories should also have a significant (though smaller) effect on list recognition against unrelated distractors. Mandler, Pearlstone, and Koopmans (1969) have reported such results; following free classification, later recognition performance (mainly hit rate) improved linearly with the number of categories into which the subjects had sorted the items.

Mandler et al. proposed a "retrieval check" hypothesis to explain their correlation between recognition and the number of categories. That hypothesis assumes that an item not recognized immediately may be recognized indirectly because it is recallable from one or another retrieval cue for the list. Our account is similar: A test item not associated directly to a list marker may be recognized because it elicits associates which do have associated

list markers. Mandler et al. would seem to emphasize the associations *to* the test item *from* other list items or retrieval cues, whereas our "mediated recognition" hypothesis emphasizes the reverse associations. The two views would appear difficult to distinguish in practice.

Distinguishing Implicit Responses

An earlier criticism of familiarity theories of recognition is that they provide no basis for editing out implicit responses that subjects make to the list items, since at the time of test the two sets of items would be equated in terms of their frequency and recency. The present approach could handle such matters by supposing that the subject has control over whether or not he will try to associate an item to a list marker. Items presented by the experimenter during that temporal block denoted as "List i" activate processes which associate the items to List i markers. But implicit responses are discriminated at the time and do not activate any processes designed to associate them to List i markers. An alternative and more satisfying mechanism which achieves the same end assumes that the subject associates with the item, as one of the prevailing contextual elements, its source (the experimenter or the subject). On this latter view, the subject might be able to recall an item that helps mediate recall of other list items and have available the further information that (*a*) this word occurred in the context of List i, and (*b*) it was an implicit associate of the subject, not provided by the experimenter.

Frequency Estimates

Although the marker model was developed for handling list differentiation, it would appear also to give a usable account of frequency estimates, that is, judgments from memory of how many times a test item appeared in an earlier context. A prototypical experiment is one by Underwood, Zimmerman, and Freund (1971) in which the subject studied a long list composed of words that individually appeared either one, two, four, or six times scattered throughout the input list. During a later test series, they judged from memory how frequently a given word had appeared. This is obviously an extended recognition memory experiment, since a "zero frequency" judgment is equivalent to nonrecognition of the item, whereas a nonzero frequency judgment corresponds to recognition of the item.

Our model can keep track of item frequency by counting the different list markers associated to the item. Each time an item is presented in the context of List i, there is probability α that it becomes associated with a List i marker. Let $M_{i,n}$ denote the number of List i markers associated to an item presented n times in List i. Then $M_{i,n}$ has the binomial distribution given by

$$Pr\{M_{i,n} = x\} = \binom{n}{x} \alpha^x (1-\alpha)^{n-x} \quad [1]$$

The mean of $M_{i,n}$ is $n\alpha$ and the variance is $n\alpha(1 - \alpha)$, both increasing linearly with the number of presentations of an item.

A plausible hypothesis is that the subject judges the frequency of occurrence of a test item in List i by some transformation of $M_{i,n}$, the number of test markers associated to that item. The simplest mapping from $M_{i,n}$ to a frequency judgment $F_{i,n}$ is a linear transformation, namely

$$F_{i,n} = a + b\, M_{i,n}. \quad [2]$$

This relation would predict a linear function relating the mean frequency estimate to the number of presentations, a result reported by Underwood et al. (1971). Furthermore, the variance of the frequency estimates should increase linearly with the number of presentations; this prediction is approximately borne out in the Underwood et al. data.

This model may be elaborated to predict an effect of forgetting on frequency estimates. By whatever mechanism one adopts, forgetting surely produces a loss of discrimination between items presented varying numbers of times. A simple realization of this in the mathematical model is to assume that each association, established during study between an item

and a list marker, has a probability $1 - f(t)$ of being retained over an interval of duration t. Therefore, as time increases, the distance between $M_{i,0}$ and $M_{i,n}$ would shrink. This could appear in the data as a decrease in the slope of the line relating mean frequency judgments to actual frequency. This prediction is upheld by the Underwood et al. data plotting the average judged versus the actual frequencies at retention intervals of a few seconds, one day, or seven days after study of the list.

REFERENCES

ABELSON, R. P., & TUKEY, J. W. Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. *Annals of Mathematical Statistics*, 1963, 34, 1347–1369.

ANDERSON, J. R. FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation*. Vol. 5. New York: Academic Press, 1972, in press.

BAHRICK, H. P. The ebb of retention. *Psychological Review*, 1965, 72, 60–73.

BERNBACH, H. A. Decision processes in memory. *Psychological Review*, 1967, 74, 462–480.

BERNBACH, H. A. Replication processes in human memory and learning. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 3. New York: Academic Press, 1969.

BOWER, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

BOWER, G. H. A selective review of organizational factors in memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press, 1972, in press. (a)

BOWER, G. H. Stimulus-sampling theory of encoding variability. In E. Martin & A. Melton (Eds.), *Coding theory in learning and memory*. New York: Academic Press, 1972, in press. (b)

BOWER, G. H., & CLARK, M. C. Narrative stories as mediators for serial learning. *Psychonomic Science*, 1969, 14, 181–182.

BOWER, G. H., & LESGOLD, A. M. Organization as a determinant of part-to-whole transfer in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 501–506.

COFER, C. N. Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory*. New York: Wiley, 1967.

DORNBUSH, R. L., & WINNICH, W. A. Short-term intentional and incidental learning. *Journal of Experimental Psychology*, 1964, 68, 58–63.

EAGLE, M., & LEITER, E. Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology*, 1964, 68, 105–111.

EGAN, J. P. *Recognition memory and the operating characteristic*. (Tech. Note AFCRC-TN-58-51) Bloomington: Indiana University, Hearing and Communication Laboratory, 1958.

EHRLICH, S. Structuration and destructuration of responses in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 282–286.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.) *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959.

FREUND, R. D., LOFTUS, G. R., & ATKINSON, R. C. Applications of multiprocess models for memory to continuous recognition tasks. *Journal of Mathematical Psychology*, 1969, 6, 576–594.

GREEN, D. M., & MOSES, F. L. On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin*, 1966, 66, 228–234.

HALL, J. F. Learning as a function of word frequency. *American Journal of Psychology*, 1954, 67, 138–140.

HINRICHS, J. V. A two-process memory-strength theory for judgment of recency. *Psychological Review*, 1970, 77, 223–233.

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

HINTZMAN, D. L., & WAIERS, R. M. Interlist and retention intervals in list discrimination. *Psychonomic Science*, 1969, 17, 357–358.

HINTZMAN, D. L., & WAIERS, R. M. Recency and frequency as factors in list discrimination. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 218–221.

HULL, C. L., HOVLAND, C. J., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. *Mathematico-deductive theory of rote learning*. New Haven: Yale University Press, 1940.

JAMES, W. *Principles of psychology*. Vol. 1. New York: Holt, 1890.

KEPPEL, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

KINTSCH, W. Recognition learning as a function of the retention interval and changes in the retention interval. *Journal of Mathematical Psychology*, 1966, 3, 413–433.

KINTSCH, W. An experimental analysis of single stimulus tests and multiple-choice tests of recognition memory. *Journal of Experimental Psychology*, 1968, 76, 1–6. (a)

KINTSCH, W. Recognition and free recall of organized lists. *Journal of Experimental Psychology*, 1968, 78, 481–487. (b)

KINTSCH, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

KINTSCH, W., & CARLSON, W. J. Changes in the memory operating characteristic during recognition learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 891–896.

KJELDERGAARD, P. M. Transfer and mediation in verbal learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

LOCKHART, R. S., & MURDOCK, B. B., JR. Memory and the theory of signal detection. *Psychological Bulletin*, 1970, 74, 100–109.

MANDLER, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

MANDLER, G., PEARLSTONE, A., & KOOPMANS, H. S. Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 410–423.

MCDOUGALL, R. Recognition and recall. *Journal of Philosophical and Scientific Methods*, 1904, 1, 229–233.

MCGOVERN, J. B. Extinction of associations in four transfer paradigms, *Psychological Monographs*, 1964, 78(16, Whole No. 593).

MÜLLER, G. E. Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufes. *Zeitschrift fur Psychologie*, Erganzungsband 8, 1913.

MURDOCK, B. B., JR. The immediate retention of unrelated words. *Journal of Experimental Psychology*, 1960, 60, 222–234.

MURDOCK, B. B., JR. Signal detection and short-term memory. *Journal of Experimental Psychology*, 1965, 70, 443–447.

NORMAN, D. A., & RUMELHART, D. E. A system for perception and memory. In D. A. Norman (Ed.), *Models of human memory* New York: Academic Press, 1970.

OLSON, G. M. Learning and retention in a continuous recognition task. *Journal of Experimental Psychology*, 1969, 81, 381–384.

PARKS, T. E. Signal-detectability theory of recognition-memory performance. *Psychological Review*, 1966, 73, 44–58.

POSTMAN, L. One-trial learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963.

POSTMAN, L., ADAMS, P. A., & PHILLIPS, L. W. Studies in incidental learning: II. The effects of association value and method of testing. *Journal of Experimental Psychology*, 1955, 49, 1–10.

QUILLIAN, M. R. The teachable language comprehender. *Communications of the Association for Computing Machinery*, 1969, 12, 459–476.

RUMELHART, D. E., LINDSAY, P. H., & NORMAN, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press, 1972, in press.

SCHWARTZ, F., & ROUSE, R. D. The activation and recovery of associations. *Psychological Issues*, 1961, 3 (Whole No. 9).

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 156–163.

SHIFFRIN, R. M. Forgetting: Trace erosion or retrieval failure? *Science*, 1970, 168, 1601–1603. (a)

SHIFFRIN, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970. (b)

TULVING, E. Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 193–197.

TULVING, E., & OSLER, S. Transfer effects in whole-part transfer. *Canadian Journal of Psychology*, 1967, 21, 253–262.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs*, 1945, 59(3, Whole No. 273).

UNDERWOOD, B. J., ZIMMERMAN, J., & FREUND, J. S. Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, 1971, 87, 149–162.

WICKELGREN, W. A., & NORMAN, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, 3, 316–347.

WINOGRAD, E. List differentiation as a function of frequency and retention interval. *Journal of Experimental Psychology*, 1968, 76 (2, Pt. 2).

(Received June 14, 1971)