SAM: A THEORY OF PROBABILISTIC SEARCH OF ASSOCIATIVE MEMORY

Jeroen G. W. Raaijmakers

UNIVERSITY OF NIJMEGEN, NIJMEGEN, THE NETHERLANDS

Richard M. Shiffrin

INDIANA UNIVERSITY, BLOOMINGTON, INDIANA

I. A Search Theory for Retrieval from Associative Memory 208
  A. The Structure of Long-Term Store (LTS) 208
  B. Retrieval from Long-Term Store 211
  C. Quantitative Sampling and Recovery Rules 214
  D. Short-Term Store and Long-Term Store 216
  E. Long-Term Storage and Learning 216
  F. Long-Term Forgetting as Retrieval Failure 217
II. A Model for Free and Cued Recall 218
  A. Storage Assumptions 218
  B. Retrieval Assumptions 219
  C. Parameters 222
III. Applications of the Theory 222
  A. Free Recall: Serial Position, List Length, Presentation Time 222
  B. Extended Recall, Repeated Recall, Hypermnesia, and Interresponse Times 229
  C. Categorized Free Recall: Cuing, Output Interference, Test Order 235
  D. Free Recall of Pictures and Words 245
  E. Part-List Cuing 247
  F. Paired-Associate Paradigms 252
IV. General Discussion and Final Comments 257
  References 260

We shall introduce in this article a theory of retrieval from long-term memory, and present a number of applications to data from paradigms involving free recall, categorized free recall, and paired-associate recall. The theory combines elements of probabilistic search theory (e.g., Shiffrin, 1970) and associative network theory (e.g., Anderson, 1972). It posits cue-dependent probabilistic search of an associative long-term

THE PSYCHOLOGY OF LEARNING
AND MOTIVATION, VOL. 14

ISBN 0-12-543314-X

memory network, and is denoted SAM, for Search of Associative Memory.

Our goals, given the length of this article, are limited in scope. The general theory is surveyed briefly, but the reader is referred to Raaijmakers and Shiffrin (1981) for a detailed discussion of the underlying basis of the theory. Then a quantitative simulation model of SAM will be described. This model is used, in essentially intact form, in many cases with no changes in parameters, to fit data from a variety of memory paradigms. To reduce the article's length these paradigms will be restricted to free and cued recall tasks for lists of singly presented items to be remembered.

I. A Search Theory for Retrieval from Associative Memory

A. THE STRUCTURE OF LONG-TERM STORE (LTS)

Long-term store (LTS) is held to be a richly interconnected network, with numerous levels, stratifications, categories, and trees, containing varieties of relationships, schemata, frames, and associations. Roughly speaking, all elements of memory are connected to all others, directly or indirectly (though perhaps quite weakly).

The "objects" of memory are defined by the task and the level of analysis pursued by the investigator. The boundary of a memory object is seldom clearly defined. For example, a "word," a "letter," and a "story" may be memory objects in different tasks; each consists of a complex bundle of informational elements, associations, and relations. In the present article the level of analysis is chosen so that the "word image," or some other similarly complex and distinct entity (such as a picture), is the basic object of memory. Even though a memory object has no clear boundaries, it can make sense to distinguish such objects from each other, in the sense that interconnections between elements and features will be stronger and more numerous within one object than between objects. Thus a memory object tends to be a relatively unitized entity.

We propose that memory structure at a given level of analysis be summarized in a retrieval structure. This structure contains retrieval strengths between the possible probe cues and the objects in memory. These strengths represent an average associative relationship between probe cues and memory objects, ignoring details such as the kind of relationships involved. As we shall indicate below, these strengths are used in a

simple ratio rule to determine the probability that a particular object will be elicited from memory when long-term memory is probed with a given set of cues. The retrieval structure is designed to capture those aspects of the memory structure that are important for retrieval. The only restriction on these strengths is that they be positive numbers. Such a structure is a rich enough representation for our retrieval model to predict many results from a variety of paradigms.

In the tasks treated in this article, the memory objects will typically be combinations of word features and contextual features, called "images." The important role played by temporal and contextual information is understandable in light of the tasks, requiring memory that a word was presented during a particular list. It would not do to let the memory image consist of word information without temporal context, since the strong preexperimental strengths between such images would mask the relatively small increments in strength that would occur due to presentation in a single list. Temporal-context, separate from word information, may be used as a cue to probe such a memory structure, or combinations of context with words may be used as cue sets to probe memory. Presumably, context alone is used as a cue when no words are available, as might be the case at the start of free recall.

Although SAM does not require a particular memory representation, it is useful to give one simplified representation to illustrate our main points. Figure 1 schematizes associations that might be formed after study of a five-item list. The item information, and the context information associated to the item information, are enclosed by solid lines. The strength of association of context to an image, when context is used as a cue, is given by the solid arrows. The solid arrows point to the item information since in many tasks the "name" of the sampled image is required. The dashed regions enclosing both context and item features indicate images of an item within the present context, i.e., the memory objects. These images are associated to other such images and the strength and direction of these interitem associations are indicated by the dashed arrows. (Associations between features are complex and are not shown in this figure; also not shown are residual associations between items not rehearsed together.)

In general, we prefer to treat multiple representations of the "same" item as separate images, each with its particular temporal-contextual elements. However, these images may be closely associated due to their pool of shared features, so that a type-token (e.g., Anderson & Bower, 1973) or episodic-semantic (e.g., Tulving, 1972) distinction may still be maintained. The set of common semantic features in many images can be considered the "type" or semantic image.

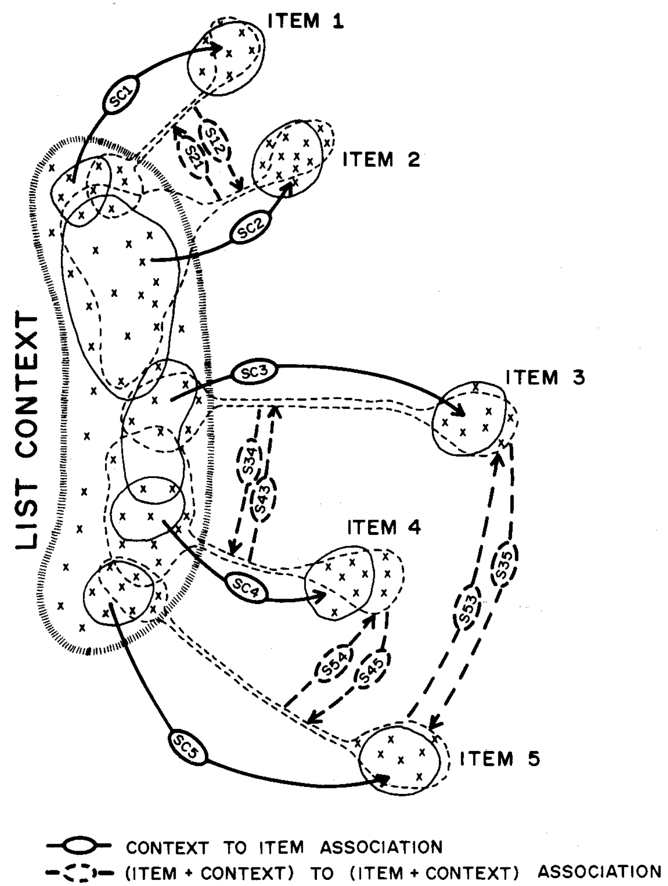

Fig. 1. A schematic depiction of the associative network in long-term store after study of a five-item list. Shown are the directional associations (dashed arrows) between images (each consisting of item plus context); also shown are the associations between context elements when used as cues, and the various images (solid arrows). Residual associations between images of items not rehearsed together are not shown. Xs refer to various features.

The final point to be emphasized is the permanence of long-term memory. We assume that information may be entered in, and added to, the long-term store, but not deleted from, or subtracted from the store. Forgetting is thus assumed to occur as a result of retrieval failure; factors governing retrieval failure will be discussed later.

B. RETRIEVAL FROM LONG-TERM STORE

The most important feature of the retrieval system is cue-dependence (see Tulving, 1974). Probe cues, whether consciously selected or not, govern each stage of the memory search. The degree to which an image in memory is associated to the set of probe cues, in comparison with the degrees to which other images are associated to the set of probe cues, determines the probability that that item will be selected at that moment in the memory search.

It is assumed that the retrieval system is noisy and inherently probabilistic; for a given memory structure and set of probe cues, the image selected from memory is a random variable. It is easy to misinterpret such a statement and ascribe more randomness to the retrieval system than is, in fact, present. The strength may be such that one image is far more likely to be selected than any other. Furthermore, the subject can control the search by changing the probe cues as needed. Nevertheless, the inherently random nature of the search has important consequences; for example, images that are sampled at one point in the search may be resampled later, especially if the probe cues are not changed.

The retrieval system as a whole is an extension of that proposed by Shiffrin (1970). It envisions retrieval as a memory search proceeding in a series of discrete steps, each step involving a selection, or sample, of an image from long-term store. The substages within any one step are depicted in Fig. 2. Retrieval begins with some question the subject needs to answer regarding the contents of long-term store. This may be as simple as "what is another word on the list most recently presented?" In the most general case, a retrieval plan will next be generated to guide the search for the answer. Initially, the plan may be somewhat vague by intention, in the hope that later phases of the search will be guided by information located in earlier phases. The plan includes such things as an initial decision whether to search long-term store, how to search (for instance, in a temporal order, or by an alphabetic strategy), how to choose probe cues (for instance, should recalled information be used as probe cues?), what combinations of probe cues should be employed, with what weights, whether to employ the same probe cues on successive loops of the search or whether to alter the cues, whether to search first for preliminary cues to guide later search, and how long to search (i.e., how many loops of the search process are expected). Of course, the plan itself is constructed on the basis of the information in the test query, the information currently available in short-term memory, and information retrieved from long-term memory; the long-term information may be concerned

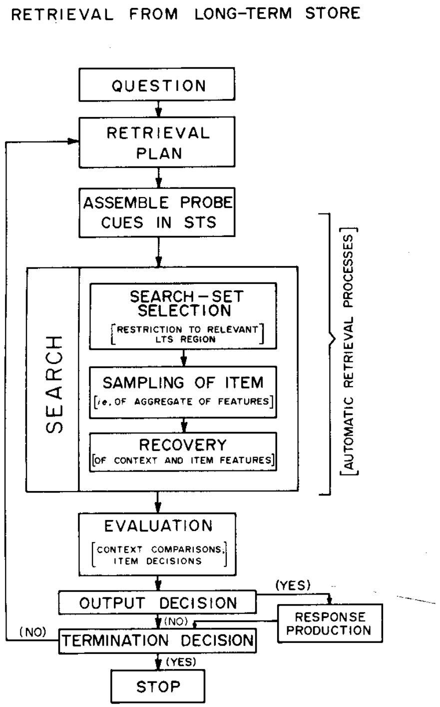

Fig. 2. A generalized depiction of the various phases of retrieval in the theory.

with search plans, previous successful plans in similar situations, and so forth (see Williams, 1977, for a discussion of retrieval plans).

Next, on the basis of the retrieval plan, the subject assembles probe cues to be used in retrieval. Generally, these cues will include: (1) information the subject has about the context at the time of study, (2) context representative of the moment of test (although these cues may not be useful or desired), (3) information from the test question, (4) information retrieved earlier in the search, and (5) information generated during construction of the retrieval plan. It is almost certainly the case, however, that there will be limitations on the amount of information that may be combined effectively into a set of probe cues. Perhaps the number of probe cues that may be used has an upper limit, or perhaps the various cues are weighted in importance (and effect), the sum of the weights being limited.

We argue that long-term memory images and probe cues are quite

distinct. For example, suppose "horse" has been placed in long-term memory at the time of study, and then a recognition test is given with "horse" as the test item. The image will consist of "horse-at-study plus study context" and the test cues will consist of "horse-at-test" along with "test context" (the encodings of horse may differ in the two instances). Thus if the memory image is sampled, it may be evaluated alone or compared with the probe cues. However, these two entities will usually be strongly associated due to their large pool of common information; it is this fact that makes it likely that the cue "horse" will cause the image "horse" to be sampled. We shall not deal with recognition in this article, but even in recall tasks note that the image corresponding to a word cue may often be sampled (though such a sample will not be useful).

The next phases of the retrieval process concern sampling and recovery. As opposed to the other stages, these stages are largely automatic and not under direct control of the subject. They determine what image is sampled and how much of the information in (or perhaps near) the sampled image becomes available to the subject for evaluation and decision making.

An image has a probability of being sampled that is determined by the associative strength relating the set of probe cues to the image, in comparison with the strengths relating all other images to the set of probe cues. (This rule will be quantified shortly.) In fact, almost all images in long-term memory will have such low strengths of association to the cues, that their sampling probabilities will be vanishingly small. The relatively small set of images with nonnegligible sampling probabilities is denoted the "search-set." It is therefore convenient (especially when incorporating the model in a computer simulation) to separate the sampling phase into two parts: first, a restriction to the search-set; second, an appropriate probabilistic choice from the search-set. The choice of search-set is generally determined by task considerations. For example, if a subject is asked to recall a just-presented list, the search-set might be assumed to consist of the images of all the words in that list (or perhaps of all the words in the session, if it is necessary to predict intrusions).

When an image is sampled, its features will tend to become activated. It is assumed that the stronger is the association between the selected image and the probe cues, the larger will be the proportion of image elements that will be activated and made available to the subject's evaluation and decision-making mechanisms. This process is termed "recovery." It may well be that the particular elements recovered from a given image for a given set of cues are fixed (at least for the short-run), so that the same elements will be recovered if the same image is sampled several times in succession.

Once a given set of informational elements has been recovered, the subject carries out evaluations and makes appropriate decisions. Such evaluations include deciding what is the verbal "name" of the sampled image, whether the sampled image was indeed on the list being tested, whether the sampled image matches the test cue (in a recognition test), etc. The subject also decides whether he has succeeded in his search, whether a response should be output, and whether the search should be continued. If the search is continued, the process loops back to the retrieval plan to start the next step in the retrieval process.

C. QUANTITATIVE SAMPLING AND RECOVERY RULES

Let us begin by positing an $N + 1$ by N matrix, with every possible memory image in the search set (N of them) given horizontally, and every possible individual cue ($N + 1$ of them, including context) given vertically. The cues, excepting the context cue, correspond to the stored images in a one-to-one fashion. Thus each image represents an item that could be used as a cue. Let the matrix (Fig. 3) contain a strength (of association) between each cue and each image. Let $S_T(Q_i, I_j)$ denote this strength between Q_i and image I_j. (The T indicates that these are the strengths that apply at Test.) Call this a "retrieval structure."

The sampling assumption may now be stated as follows:

$$P_S(I_i|Q_1, Q_2 \ldots Q_M) = \frac{\prod_{j=1}^{M} \{S_T(Q_j, I_i)\}^{W_j}}{\sum_{k=1}^{N} \prod_{j=1}^{M} \{S_T(Q_j, I_k)\}^{W_j}} \tag{1}$$

The term on the left indicates the probability of sampling image I_i given cues $Q_1, \ldots, Q_M$ are used in combination as a probe set. The W_j in the right-hand expression are weights assigned to the different cues representing their relative saliency, or importance (or overlap, or similarity). (In the applications in this article it was unnecessary to assume unequal weights, and the W_j were all set to 1.0.)

The key to the present approach is the method used to combine cues: the strengths to the different cues are multiplied (perhaps in weighted fashion), and the ratio rule (Luce, 1959) applied to the products. This multiplicative feature has the useful and important consequence that it allows focusing of the search. The images with the highest probability of being sampled are those with the highest product of strengths, and hence those that tend to be strongly associated to *all* of the cues. The sampled image tends to come from the intersection of the sets of images strongly

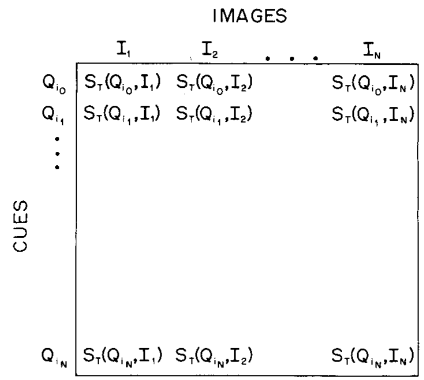

Fig. 3. The "test matrix": the matrix of strengths which determine the probabilities of selection and recovery of list images (horizontal margin) when different cues (vertical margin) are used in the cue set. Entries in the cells are strengths from individual cues to individual images; when multiple cues are used in the cue-set, then the strengths are combined according to Eqs. (1) and (2) in the text. Q_{i_0} refers to the context cue; Q_{i_m} refers to the word whose image is I_m.

associated to each cue separately. By contrast, an additive combination rule could give a high probability of sampling an image if only one cue strength is high (even if all the others are zero).

Consider next the recovery process. We give here the recovery rule when the subject's task is to generate the "name" of the word encoded in the selected image:

$$P_R(I_i|Q_1, Q_2 \ldots Q_M) = 1 - \exp\{-\sum_{j=1}^{M} W_j S_T(Q_j, I_i)\} \tag{2}$$

The expression on the left represents the probability of recovering enough information to correctly give the "name" encoded in image I_i, which has just been selected using probe cues Q_1 to Q_M. The right-hand expression is somewhat arbitrary mathematically, though it does capture a number of features we consider desirable for a recovery rule in this case. First, the stronger the strength to any one cue and the stronger the summed strengths to all cues, the more likely is recovery. Second, the larger is a cue weight, the more the strength to that cue will affect recovery. Note that recovery obeys an additive rather than multiplicative rule, so that recovery probability will be high if even one weighted strength is high. Third, the probabilities will range from 0 to 1 as the sum of the strengths ranges from 0 to ∞. Note that both these sampling and recovery rules are natural elaborations of the Shiffrin (1970) rules.

D. SHORT-TERM STORE AND LONG-TERM STORE

The description of structure and retrieval given above is to a large degree independent of the theoretical assumption that memory is a two-phase system. Nevertheless, we find it useful for many reasons to place our retrieval system within a memory theory organized around a short-term store (STS) and long-term store (LTS). Atkinson and Shiffrin (1968) provide a prototype of such a system, but Shiffrin (1975) gives a more contemporary treatment. We shall review this system very briefly.

STS is postulated to be a temporarily activated subset of the information (and structure) in LTS, the permanent storage system. Sensory information presented to the system is analyzed automatically in a series of stages along many parallel paths. This analysis results in activation of information in LTS, and activation is equivalent to entry in STS. Alternatively, information is activated from LTS and placed in STS on the basis of internally generated probe cues, as described in earlier sections. Inevitably, both types of LTS activation occur together, so that sensory input (bottom-up processing) and information previously retrieved from LTS and presently still in STS (top-down processing) will jointly act to determine subsequent activation. In general, the activated information decays (becomes inactive) very rapidly, though small amounts of information may remain active in the absence of new input, or may be maintained in an active state for a long time through control operations like coding and rehearsal.

A major role of STS is its use as a working space for control processes of all sorts, including plans, coding, rehearsal, decisions, and so forth. The most important characteristic of STS is its limited capacity (see Shiffrin, 1976). There are limitations upon the rate of retrieval and examination of the contents of STS, upon the duration of residence in STS, upon the amount of information active in STS, upon the ability to focus and divide attention, and upon the rate of encoding of new information, among others.

These STS limitations affect retrieval in a number of ways. The impermanence and capacity limitations of STS limit the amount of information that may be sampled from the search set and maintained in an active mode. The limited rate of examination leads to sequential examination of one image at a time. The limitations on total STS load put bounds on the number of cues that may be used simultaneously (or on the sum of the weights).

E. LONG-TERM STORAGE AND LEARNING

Learning in our system consists of the formation of new associations, relations, and structures, in LTS, between elements and images already present in LTS, but concurrently active in STS. Although contiguity in STS may produce some storage in a passive mode, the strongest associations are formed when the subject gives attention to the material, and applies control processes like rehearsal, coding, evaluation, and relating the new material to already stored structures.

In most learning situations, storage will result from a combination of LTS retrieval and STS control operations. Indeed, since the coding of sensory input is also a form of LTS retrieval, all storage can be conceived as retrieval of a variety of LTS structures followed by the formation of new associative relationships between the retrieved structures.

The idea that the information simultaneously active in STS tends to be stored together is an extremely important aspect of the theory. In particular, it explains the prominence of temporal-contextual features (i.e., episodic memory—see Tulving, 1972) in memory images. Such temporal-contextual features include "incidental" information from the sensory environment and the subject's long-term store that happens to be present in STS at the time of a storage event. They might include the location, the temperature, the time of day, recent events, the subject's physical state, feelings, emotions, and recent thoughts. Each and every storage event will contain such temporal-contextual information to some degree, and this temporal context plays a prominent role in our retrieval theory and explanations of forgetting. In all retrieval situations, context will play a role as one of the probe cues, either by intent or accident. Presumably, the subject can, through attention, vary the weight assigned to this context cue, but such information will always be present in STS and will always play at least a small role as a retrieval cue. Whenever possible, of course, a knowledgeable subject will try to reinstate in STS as far as possible the contextual cues that had been present at the time that the to-be-recalled image had been stored.

F. LONG-TERM FORGETTING AS RETRIEVAL FAILURE

There are two basic reasons why an image may be retrieved better at time A than at time B. First, the cues utilized at time A may be more strongly associated to the image than those used at time B. Second, the strength or number of *other* images associated to the cues (even if the cues are the same) may be greater at time B than at time A. Everything else being equal, an increase of cue to image strength will increase both sampling and recovery probabilities [see Eqs. (1) and (2)]. On the other hand, for fixed cue to image strength, an increase in the strengths of cue to *other* images will reduce the sampling probabilities (though probably leaving recovery unaffected).

The increase in the strengths of association of cues to other images tends to be an inevitable consequence of new learning. This new learning will not necessarily lead to forgetting, however. The new information might be organized together or integrated with the old image so strongly that the retrieval of either set of information will then lead at once to retrieval of the other set. This integration could be conceptualized either as resulting in a single, new larger image, or as resulting in two tightly associated images. In the latter case, retrieval of one of the images could result in that image being used as a cue, and thereby eliciting the other image. This possibility is an example of a general principle: forgetting due to new learning occurs when the *same* cue is utilized in an attempt to locate one image among an increasing number of other images. On the other hand, the cues may be changed during the search so that each cue is related to a subset of the increasing number of images; in this event forgetting may be ameliorated or even reversed.

The decrease in the strengths of association of cues to image can be the result of several factors, chief of which is the change of context over time (see Estes, 1955; Bower, 1972). The context at the time of storage makes the best retrieval cue, but at the time of test, the context cue used may consist largely of the context information at the time of test, which will usually differ from the storage context by a greater amount as time between storage and test increases. Similar considerations apply to noncontextual cues, the general rule being that sampling and recovery will be worse as the retrieval cues chosen are less effectively associated to the desired image.

II. A Model for Free and Cued Recall

We develop the theory initially for the paradigm of free verbal recall. A list of N "unrelated" words is presented, one at a time. The presentation is sometimes followed by an arithmetic task to clear STS and restrict retrieval to LTS. The test involves recall of as many list words as possible, in any order. Usually, enough recall time is provided that the subject decides to cease retrieval, in the belief that memory is exhausted, before the recall period ends.

A. STORAGE ASSUMPTIONS

On the basis of coding and rehearsal operations that operate during list presentation, an LTS structure is generated and stored in LTS. The strength of associations of the cues at test to the LTS images is based on this structure.

Although many storage models are possible, we find it easiest to adopt the now traditional buffer rehearsal process of Atkinson and Shiffrin (1968). The buffer size is r. New words enter the buffer until it is full; then each new word replaces a randomly chosen word already in the buffer. The retrieval structure contains associative strengths between a general context cue and the images on the list, and between word cues and those images. It is assumed that these associative strengths grow linearly as a function of the total time that a word or a pair of words is rehearsed in the buffer. If we let t_i and t_{ij} be the times spent in the buffer respectively by I_i, and by I_i and I_j together then we assume: $S_T(C, I_i) = at_i$; $S_T(I_i, I_j) = S_T(I_j, I_i) = bt_{ij}$, $t_{ij} \neq 0$; $S_T(I_i, I_i) = ct_i$. Finally, even if two words on the list are not rehearsed together, they share context and are therefore assumed to have a nonnegligible residual retrieval strength, d: $S_T(I_i, I_j) = S_T(I_j, I_i) = d$, $t_{ij} = 0$. Thus the four parameters, a, b, c, and d, along with the buffer size r, completely determine the test matrix *at the start of retrieval.*

One additional storage process needs to be discussed. We assume that additional storage may take place during the course of retrieval itself. During retrieval, it will sometimes happen that a word is sampled, recovered, and recalled when a particular combination of cues is utilized as a probe set. Whenever this happens, but only when the recalled word is actually output, we assume that the strengths of the cues to the sampled image are incremented (i.e., increased). In addition, we assume that the self-association strength of the sampled image is also incremented in each such case. Thus we assume: $S'_T(C, I_i) = S_T(C, I_i) + e$; $S'_T(I_i, I_j) = S'_T(I_j, I_i) = S_T(I_i, I_j) + f$; $S'_T(I_i, I_i) = S_T(I_i, I_i) + g$, where the primes indicate the strengths after incrementing, and e, f, and g are the parameters giving the amount of the increment in each case. As we shall see in the retrieval model, context will always be a cue, so the context strength and the self-association strength will always be incremented after a recall. The word–word strength will be incremented only in those cases in which a word was one of the cues.

B. RETRIEVAL ASSUMPTIONS

The heart of the retrieval model is Eqs. (1) and (2), giving sampling and recovery probabilities. In the present applications the weights W_i are set equal to 1.0.

At test time, any words still remaining in STS are output. Then retrieval from LTS begins. Figure 4 gives a flowchart corresponding to the first, main, phase of the model (which was written as a computer simulation).

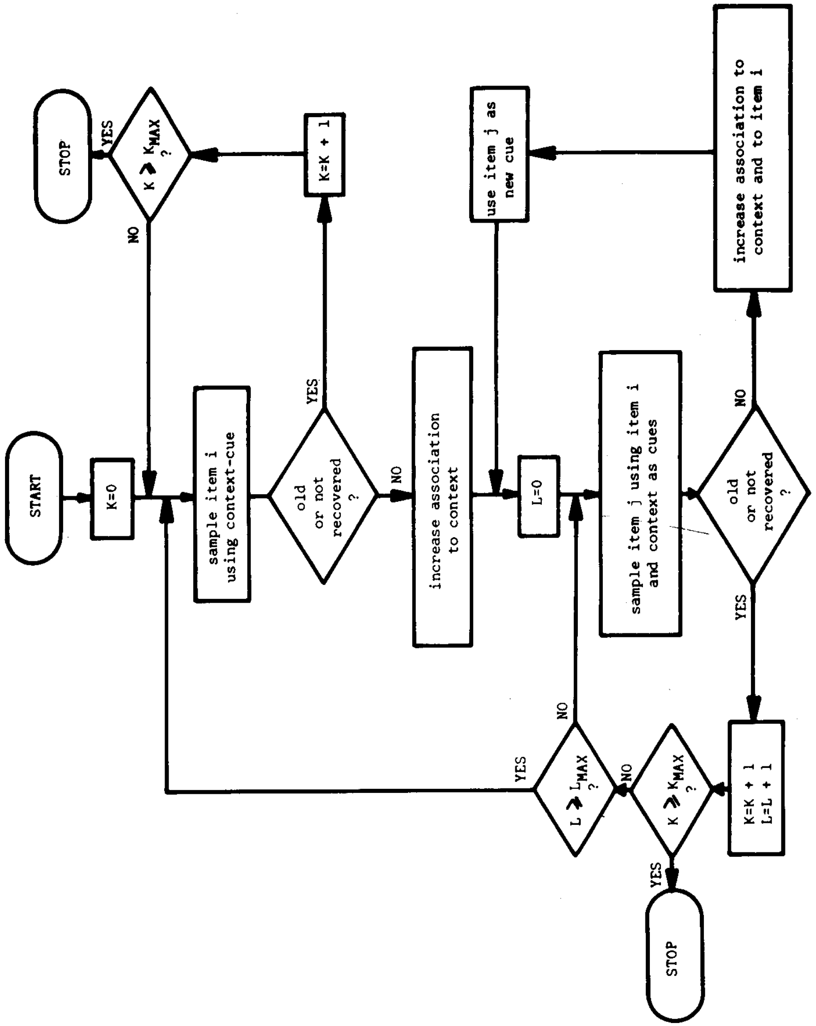

Fig. 4. A flowchart for phase one of the retrieval process in the computer simulation of SAM developed for free recall.

Search goes on until a criterion of K_{MAX} total failures is reached. A failure is every retrieval attempt that does not lead to recall of a new word. Thus, at the start of the recall from LTS the total failure counter K is set equal to 0. Next, the subject uses the general context cue, representing the context during study of that particular list, to sample from the images that are associated to that cue [Eq. (1) is used]. Suppose that image I_i has indeed been sampled; this is a failure if the image is "old" or if the image is new and cannot be recovered. An image is considered "old" if it has already been recalled, or if each of the present cues has previously been used to sample this image (unsuccessfully). This may be justified by the assumption that a given, fixed set of probe cues will always lead to recovery (i.e., activation) of the same set of features from a given sampled image. If the image is not "old," the recovery rule of Eq. (2) is applied. If recovery fails, then a failure is counted, and context sampling is tried again. If recovery succeeds, then the strength $S_T(C, I_i)$ is incremented by e, and the self-association $S_T(I_i, I_i)$ is incremented by g.

After a successful recovery, the recalled word, I_i, is used as a retrieval cue along with context for the next sample [Eq. (1) is used]. Suppose image I_j is sampled. Then, as before, this is a failure if I_j is "old" or is new and cannot be recovered. (Note that in this case, the image may have been sampled unsuccessfully before, but the image will be considered "new" as long as the retrieval route $I_i - I_j$ is "new.") If I_j is new, then the recovery probability is determined by Eq. (2). If recovery fails, then a failure is counted, and L and K are increased by 1. If L_{MAX} and K_{MAX} are not reached, then the same cue combination is used again. If L_{MAX} is reached, then only context is used in the next cue set.

If recovery succeeds, then all relevant strengths are incremented: $S_T(C, I_j)$ is incremented by e, $S_T(I_j, I_j)$ is incremented by g, and $S_T(I_i, I_j)$ is incremented by f [as is $S_T(I_j, I_i)$ since we assume bidirectionality].

If I_j has been recovered, then this word is used as a cue, along with context, in the next cue set. This entire process continues until K_{MAX} total failures are reached.

In summary, extensive use is made of interitem associative routes: whenever a new word is recalled it is used as a cue either until L_{MAX} failures accumulate or until a new word is recalled, in which case the new word is used as a cue. Of course, it could be argued that all interitem routes have not been fully explored, since a switch to a new word cue may occur before search with the previous word cue has been exhausted.

For this reason a final "rechecking" process is incorporated in the model after the K_{MAX} criterion has been reached. Every word that has been recalled (presumably they are written down and hence available) is used as a cue, along with context. L_{MAX} samples are made with each such cue combination. Any new words recalled during this period are also "rechecked." With this rechecking process added, it may be argued that the subject feels all retrieval routes have been tried and exhausted.

It is not too difficult to see that this retrieval model is at least potentially able to explain a variety of data in free recall, especially if one recognizes that the model combines features of two powerful models, namely, the model of Shiffrin (1970) and the FRAN model of Anderson (1972).

C. PARAMETERS

The model described above is ready to be applied to the data from free recall studies. The parameters are *a* (context to image strength), *b* (image to image strength), *c* (image to self strength), *d* (residual strength), *e* (context to image increment), *f* (image to image increment), *g* (self increment), K_{MAX} (total failure stopping criterion), L_{MAX} (stopping criterion for a word cue), and *r* (buffer size).

At first glance, 10 parameters seems quite a high number, even though we shall fit a great deal of data from a variety of paradigms. For example, Shiffrin (1970) fit a great deal of free recall data with just three parameters. This objection is ameliorated by the following factors. We can show that most of the present parameters, and their precise values, are *not* essential for the fit of the model to most of the data. The parameters are listed above for generality, even though some are never varied and others are equated before fits to the data are begun. Some of the parameters are given nonzero values and included in the fit merely to demonstrate that the presence of the processes they represent will not harm the ability of the model to predict the data. In fact, we have set many of these parameters to zero, and no harm to the model's predictions results. However, each of these parameters represents processes that we feel are needed on logical grounds, or needed to deal with data from at least one of the studies to be discussed in this paper. The roles played by the various parameters have been extensively explored by simulation means, as have certain process assumptions, and the results of these explorations will be summarized briefly or reported in detail in the remainder of the article.

III. Applications of the Theory

A. FREE RECALL: SERIAL POSITION, LIST LENGTH, PRESENTATION TIME

Primacy and recency effects are predicted by our model as a consequence of the buffer assumption. These effects are therefore easy to

predict but they are not very informative concerning the LTS retrieval process. A more interesting result is that the model is able to describe the serial position curves for different list lengths and presentation times with the same set of parameter values. Figures 5, 6, and 7 show the serial position curves obtained by Murdock (1962) in a task including STS recall. In Murdock's experiment six groups of subjects each had a different combination of list length and presentation rate. The six conditions were 10–2, 20–1, 15–2, 30–1, 20–2, and 40–1, where the first number refers to the list length and second number indicates the number of seconds that an item was presented.

The parameters of our model were very roughly estimated by a Monte Carlo simulation technique from the data of conditions 10–2, 20–1, 20–2, and 40–1. These parameter estimates should not be regarded as optimal since only a limited search of the parameter space was feasible. Moreover, the parameter space is quite shallow, so that many other combinations of parameter values will give a fit about equally good.

Many of the parameters were set arbitrarily, rather than estimated. The buffer size, *r*, was set equal to 4, and K_{MAX} set equal to 30, on the basis of previous work (Shiffrin, 1970). L_{MAX} was set equal to 3 (a value that later simulations showed produced near maximum recall). The values of *e, f,* and *g* (incrementing) were set equal, *d* (the residual) was set to one fifth of *b,* and then *a, b,* and *e* were estimated. The values that gave a "best" fit, roughly, were $a = .055$; $b = .02$; $e = .6$. The resulting

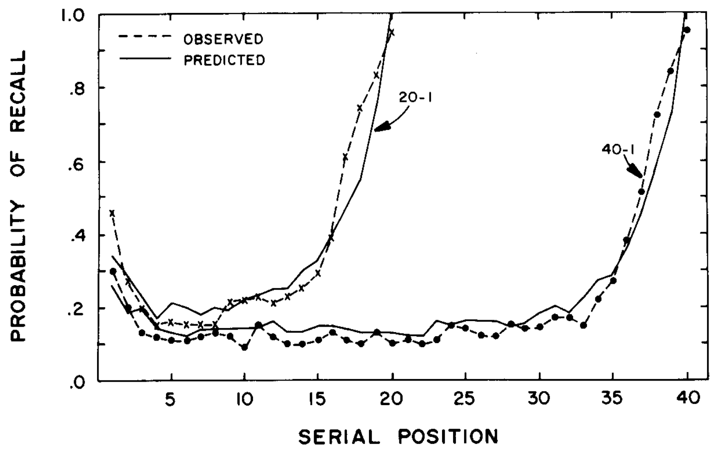

Fig. 5. Serial position results (Murdock, 1962) for 20- and 40-word lists at 1 sec/word. Predictions from the SAM model with $r=4$, $K_{MAX}=30$, $a=c=.055$, $b=.02$, $d=.004$, $e=f=g=.6$. These parameters are also used to derive the predictions in Figs. 6 and 7.

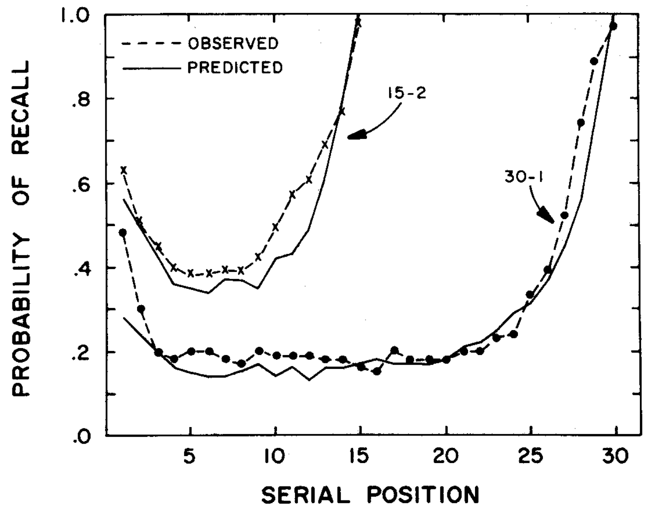

Fig. 6. Serial position results and SAM's predictions for a 30-word list at 1 sec/word and a 15-word list at 2 sec/word. (Data from Murdock, 1962.)

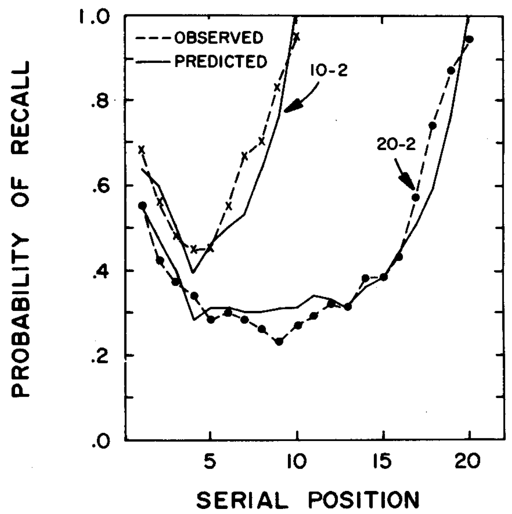

Fig. 7. Serial position results and SAM's predictions for 10- and 20-word lists at 2 sec/word. (Data from Murdock, 1962.)

predictions are shown in Figs. 5, 6, and 7. Clearly, the fit is quite adequate.

Note that both the predictions and the data show a list-length effect: the probability of recall is a decreasing function of list length. This list-length effect is predicted by the model because the search termination criterion is exceeded sooner for the longer lists, relative to the list length, i.e., *relatively* fewer samples are made from a longer list than from a shorter list. The probability of sampling an item is therefore lower for an item from a longer list. This effect is predicted by the model even when the criterion is set very high ($K_{MAX} = 100$) or when a stop-rule of K_{MAX} *consecutive* failures is used. Thus, this prediction is a consequence of the basic structure of the model: a sampling-with-replacement retrieval process coupled with a fixed termination criterion (i.e., the criterion does not vary with list length).

We should note that many of the process assumptions and parameter values are not essential for predicting the Murdock data. If rechecking is eliminated, a *very* slight adjustment in the *a, b,* and *e* values will produce an equivalent fit. If the residual association, *d,* is removed (set to 0), an equivalent fit is obtainable by changing *a* to .065 and *b* to .015. If the stopping rule is changed to K_{MAX} *consecutive* failures, an equivalent fit is obtained without changing any parameters, but letting $K_{MAX} = 15$ (including the three failures in the last search with a word cue).

Although the fit to Murdock's data is quite good, list length and presentation time per item were not varied over a very wide range. Roberts (1972) reported the results of a large, well controlled study, where four list lengths (10, 20, 30, or 40 items) and five presentation rates (.5, 1, 2, 4, or 8 sec per item) were varied in a factorial design. His results are shown in Figs. 8 and 9 in the top panels, where we have averaged the data for the auditory and the visual presentation modes. Note that these results include recall from short-term store since no interpolated task was given. These results show that the mean number of words recalled is not a linear function of the total presentation time (Murdock, 1960) but a negatively accelerated function as found by Waugh (1967). They also clearly show that the total-time hypothesis (Murdock, 1960) is incorrect: equal total presentation times do not yield equal levels of recall.

Figures 8 and 9 also show the predictions derived from the present model. These predictions include of course the recall from the STS-buffer. Parameters were estimated as in the case of Murdock's data: *a, b,* and *e* were estimated. The best fitting values were $a = .10$, $b = .10$, $e = .70$. The quality of the fit to the data is seen most easily in Fig. 10, which compares probabilities of predicted and observed recall for each of the 20 points in Fig. 8 (or 9). Obviously the fit of the model is quite

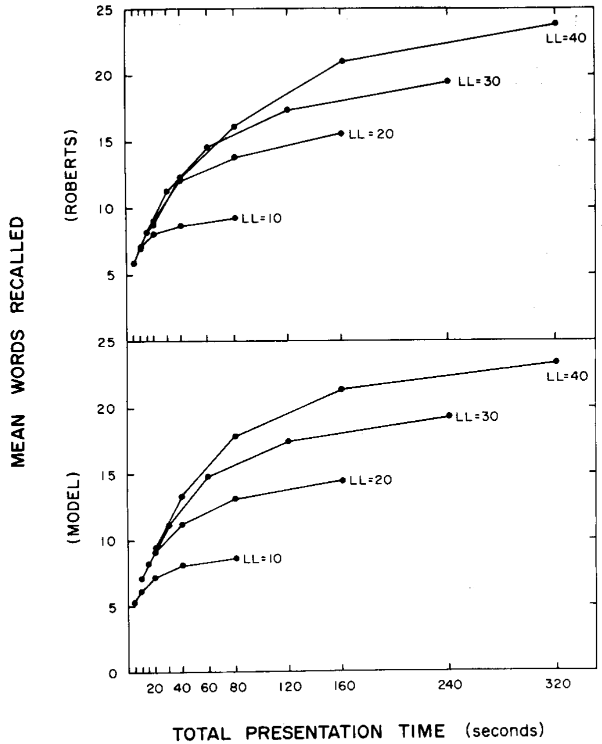

Fig. 8. Observed data (Roberts, 1972—top panel) and predictions of SAM (lower panel), for mean words free-recalled as a function of presentation time and list length (LL). The parameters of the model (see text): $r=4$; $K_{\mathrm{MAX}}=30$; $L_{\mathrm{MAX}}=3$; $a=c=.10$; $b=.10$; $d=.02$; $e=f=g=.70$.

satisfactory. Note that in our model presentation time per item has its effect mostly on the probability of recovery, not on the probability of sampling [see Eqs. (1) and (2)]. There is only a small effect on the probability of sampling due to the fact that the increment upon successful recovery is the same constant in all conditions and thus relatively higher in the case of a lower presentation time per item. As with Murdock's data, a list-length effect is evident in Fig. 8. Our model predicts such effects because relatively fewer samples are made from a longer list. Thus, the list-length effect is predicted to be a retrieval effect, not a storage effect. Of course, the subjects do not know how long the list is going to be.

As with Murdock's fit, certain processing assumptions and parameter

values are not essential in this case. Rechecking makes little difference. Setting $d = 0$ can be compensated for by raising a to .12, with equally good results. Changing to a K_{MAX} consecutive failure rule, with $K_{\mathrm{MAX}} = 11$, and the other parameters unchanged, gives a fairly good fit, but with the predicted points in Fig. 10 lying along a line with slightly greater slope than the observed points. This is easily fixed, however, by slight changes in the values of the other parameters. All in all, it seems clear that the predictions of the list-length effects and the effects of presentation time are the result of the basic structure of the model and not of the particular parameter values or assumptions used.

In the applications both to Murdock's data and Robert's data, no mention has been made of the effects of the interword-association retrieval routes. The reason is simple. The word cue searches are not needed to

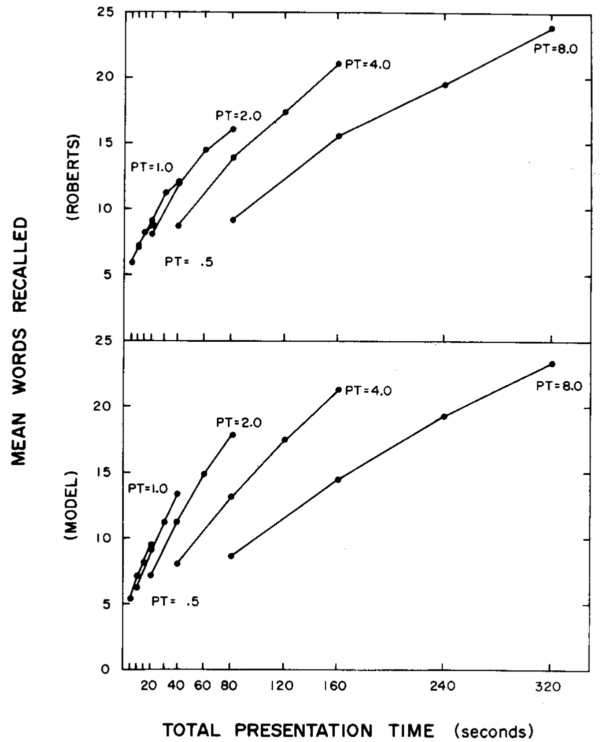

Fig. 9. Same as Fig. 8, except parameter of each curve is presentation time per item.

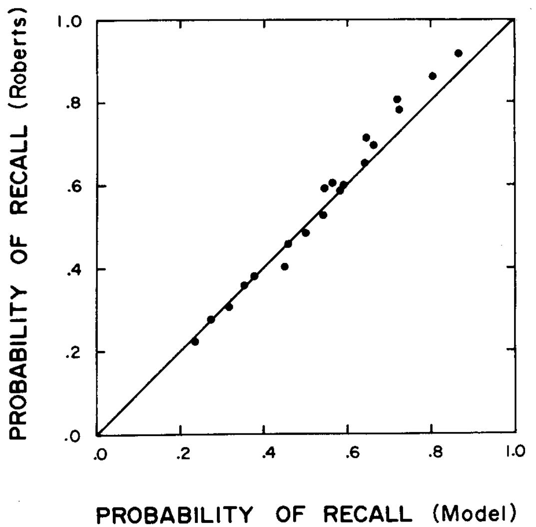

Fig. 10. Comparison of the predicted and observed probabilities of recall, corresponding to each of the 20 points in Fig. 8 (or 9).

predict these data (see Shiffrin, 1970). However, the word-cue searches are crucial for many of the applications to be covered later, and it is noteworthy that the inclusion of such a process does not affect the predictions. Interestingly enough, we have even found that virtually identical predictions, for the same parameter values, can be obtained if it is assumed that the *two* most recently recalled words are used as cues along with context (until L_{MAX} failures accumulate). Such results suggest that a subject may not have a uniformly optimal strategy of cue selection. They further suggest that different subjects, or the same subject at different times, may use different numbers of item cues without much affecting total recall.

Finally, consider the strictness of the search stopping criterion (the value of K_{MAX}). The fact that K_{MAX} was not estimated and yet a good fit was obtained suggests that the criterion may be shifted, one or two other parameter values changed slightly, and equally good predictions obtained. This is, in fact, the case. More important, it is one of the great successes of the model that empirical manipulations designed to change the stopping rule produce results that are extremely well predicted by the model, with the only alterations occurring in the value of K_{MAX}. We turn next to such studies and predictions.

B. EXTENDED RECALL, REPEATED RECALL, HYPERMNESIA, AND INTERRESPONSE TIMES

A model for free recall should exhibit several closely related properties: (1) the criterion for cessation of search should be such that a reasonable subject could be expected to "give up" at that point; (2) the temporal point at which time-unlimited search stops should be a point at which few new items are being recalled (for all conditions); (3) predicted cumulative recall functions, at times before subjects cease searching, should grow at a rate similar to that seen in the data; (4) if subjects are induced to extend their search beyond the point of normal cessation, the number of additional words recalled should be predictable by shifting the stopping criterion, or by otherwise altering the model's stopping strategy in a manner consistent with the instructions and task demands.

Consider first the stopping rule. When a *total failure* rule is used, even with a high value of K_{MAX}, it is not obvious that new recalls will be occurring at a slow rate just before search ceases. In fact, however, even with $K_{MAX} = 30$, the output rate is quite low most of the time when search stops. This is supported by the observation that in the models for the Murdock and Roberts data, *consecutive failure* rules of 15 and 11, respectively, give predictions virtually equivalent to those for the total failure rule. Certainly it seems reasonable that a subject should cease recall after such a long string of failures (ignore for the moment the fact that we assume rechecking to take place *after* this criterion is reached).

In order to gain a clearer picture of the output rates, we give in Fig. 11

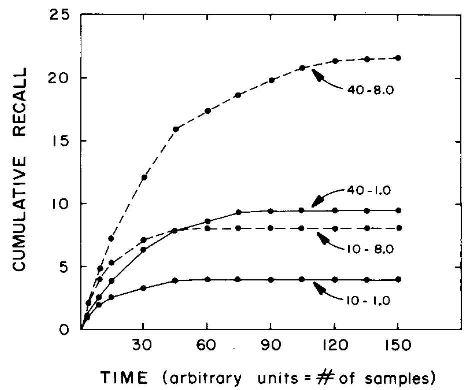

Fig. 11. Predicted cumulative output functions for list lengths of 10 and 40 and presentation times of 1.0 and 8.0 sec per item. SAM's parameter values are the same as for Fig. 8, but note that STS retrieval has been deleted.

predicted cumulative output functions for four of Robert's conditions, with Robert's parameters. Note that a stop rule followed by rechecking is assumed to apply to each "subject"; the cumulative functions simply give total cumulated recall over subjects divided by the number of subjects, so that after a long "time" most of the "subjects" have ceased trying to recall, and only a few of the subjects are contributing new recalls to the cumulative functions. This explains why the lower functions in the figure reach a nongrowing asymptote—all subjects have stopped retrieving. These predicted functions show a very important property; the rate of approach to a higher final asymptote is slower. In fact, a considerable literature attests to just this fact (see Johnson, Johnson & Mark, 1951; Bousfield, Sedgewick, & Cohen, 1954; Indow & Togano, 1970). Note that predicted recall grows at a reasonable rate for quite a long time when the list length and presentation time per item are large. What should be the most reasonable stopping rule in such a case is difficult to judge.

The predictions in Fig. 11 show what happens when subjects are assumed to use a normal stopping rule, so that search ceases relatively quickly. It might be asked, what are the predicted cumulative output functions if subjects are induced to search for very lengthy periods without stopping. Typical predictions are shown in Fig. 12. The Robert's parameters are used, except there is no stopping rule at all. List length is set to 40, presentation time to 4 sec/word. The dashed curve is an extreme case in which *no* rechecking is assumed, so that almost all samples late in retrieval use context only as a cue (only after a new recall occurs is there a brief period, with criterion = L_{MAX} = 3, of cuing with word + context). The solid curve gives an extreme case in which rechecking occurs whenever a multiple of 50 samples occurs (unless rechecking is still underway at that point). Early in search, for this list length and rate, rechecking actually harms recall (see the portion of the curves between 50 and 100). This occurs because rechecking gives rise to new words very slowly. On the other hand, rechecking ensures that new retrieval routes become available. That is, an image may have been sampled but not recovered with context and words $i_1, i_2, \ldots, i_M$ as cues; this image can still be recovered if sampled with a new cue, word i_{M+1}. Thus the rechecking curve continues to rise, albeit slowly, until it surpasses the other curve. This is a general property of the rechecking assumption: it causes cumulative output functions to continue to grow for longer periods of time. Finally, it may be interesting to compare the levels of recall after 500 samples under either assumption, about 29 words, with the predicted level if a total failure rule of 30 is used (see Fig. 8), 21.4 words. For the rechecking curve, such a level corresponds to about 125 samples. Clearly, more words are predicted to be available in memory when search

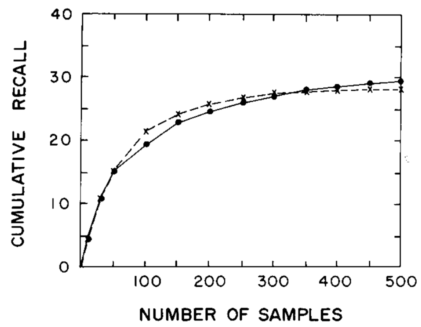

Fig. 12. Cumulative output predictions for SAM without a stopping rule. Dashed curve: no rechecking assumed. Solid curve: rechecking assumed every 50 samples. Parameter values as in Fig. 8; list length = 40, presentation time = 4 sec/word.

stops with K_{MAX} = 30, but an enormous effort may be required to retrieve them.

We might now ask whether cumulative *data* functions show any of these properties. Figure 13 shows cumulative functions from Roediger and Thorpe (1978) who induced subjects to continue to try to recall for 21 min, corresponding to the assumptions used in Fig. 12. Without attempting to estimate parameters, is seems clear that real subjects show increases in recall over quite long periods of time, and that the growth functions are quite similar in form to SAM's predictions.

Figure 13 also shows what happens when the subject is given three consecutive recall periods of 7 min each, the subject beginning over in each new recall period. The cumulative curves shown for this case ignore any multiply recalled words and simply count new words recalled. On the other hand, if one counts total words recalled during each 7 min period, then this total increases in each period, especially for pictures, as shown in the top left panel of Fig. 14. This phenomenon has been called "hypermnesia" by Erdelyi and his colleagues (see Erdelyi & Kleinbard, 1978) and interpreted as some sort of "negative forgetting." Roediger and Thorpe's data shown in Figs. 13 and 14 seem to make it clear that the effect is merely a consequence of more total time available for recall, along with a result that fewer previously recalled items are forgotten than new words are recalled.

It is not immediately obvious that SAM should predict this "hypermnesia" result, despite the predicted growth in cumulative output, since it is difficult to judge intuitively how many previously recalled words will be predicted to be forgotten in a following recall period. It is easy to apply SAM to this task, however. Since Roediger and Thorpe

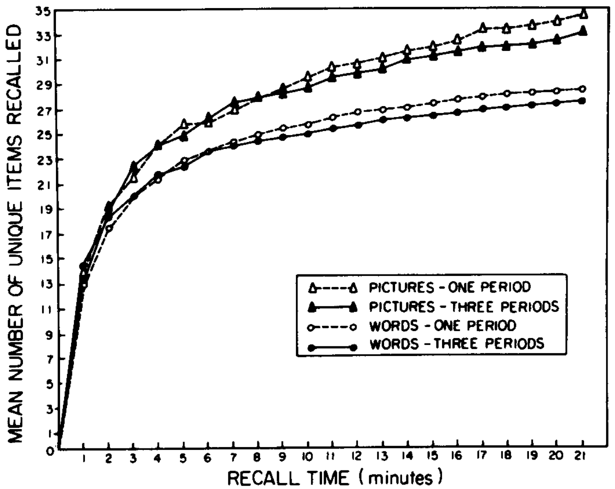

Fig. 13. Mean cumulative recall of unique items for subjects presented pictures or words and given either three 7-min tests or one 21-min test (taken from Roediger & Thorpe, 1978).

(1978) use long recall periods (7 min) and report that the hypermnesia effect is obtained only in this case, we deleted the stop rule from SAM and simply assumed that 150 samples were made in each recall period. Rechecking of all previously recalled items within the current recall period was assumed after each 50 samples. List length was set equal to 40, presentation time per item to 1 sec, and parameters chosen to represent high strengths (a = .2, b = .1, e = .7) or low strengths (a = .1, b = .05, e = .7). The other parameters were those for Roberts data (r = 4, L_{MAX} = 3). The predictions are shown in the middle upper panel of Fig. 14. Under either high or low strength assumptions, the increase across successive recall periods is observed. The upper right-hand panel shows similar predictions when the incrementing parameter is greatly increased, to 3.0.

In order to determine which features of the model are responsible for the "hypermnesia" prediction, several alternate assumptions were used. The "normal" version already discussed assumes that an image that has previously been sampled but not recovered may still be recovered later if the cue-set contains at least one cue that is new for that image. This assumption is denoted "alternate retrieval routes." The lower left-hand panel shows that the hypermnesia prediction is reduced but not eliminated if the "alternate routes" possibility is eliminated. In this case, only one recovery chance is possible for a given image, but incrementing remains.

The center lower panel shows that hypermnesia is also predicted if incrementing is set to .0, but alternate routes remain. However, if *both* "alternate retrieval routes" and "incrementing" are removed from SAM, then it may be shown that no change in recall is predicted for successive recall periods. Finally, the right-hand lower panel shows that the amount of increase is lowered but not eliminated if increments are allowed to take place anew in each successive recall period (thus an item already incremented to a cue, can receive another increment to that cue in a later recall period).

One or two final points should be mentioned about the "hypermnesia" prediction. First, if the number of samples per recall period is reduced, the predicted increase in recall lessens considerably. Such a prediction accords with data reported by Tulving (1967) and Donaldson (1971) (although these results are difficult to interpret because the first recall

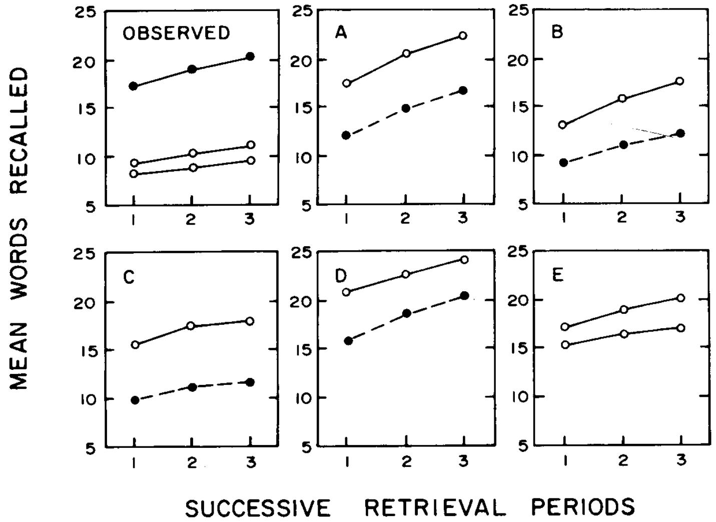

Fig. 14. Mean number of words recalled in each of three successive retrieval periods for the same list. Top left panel: observed data from Roediger and Thorpe (1978). Panels A to E give predictions. (A) Alternate retrieval routes assumed: solid curve: $a=c=.2$, $b=.1$, $d=.02$, $e=f=g=.7$; dashed curve: $a=c=.1$, $b=.05$, $d=.01$, $e=f=g=.7$. (B) Alternate retrieval routes assumed: solid curve: $a=c=.2$, $b=.1$, $d=.02$, $e=f=g=3.0$; dashed curve: $a=c=.1$, $b=.05$, $d=.01$, $e=f=g=3.0$. (C) No alternate routes: solid curve: $a=c=.2$, $b=.1$, $d=.02$, $e=f=g=.7$; dashed curve: $a=c=.1$, $b=.05$, $d=.01$, $e=f=g=.7$. (D) Alternate routes assumed: solid curve: $a=c=.2$, $b=.1$, $d=.02$, $e=f=g=0$; dashed curve: $a=c=.1$, $b=.05$, $d=.01$, $e=f=g=0$. (E) Increments apply again each new retrieval period; top curve: alternate routes assumed: $a=c=.2$, $b=.1$, $d=.02$, $e=f=g=.7$; bottom curve: no alternate routes: $a=c=.2$, $b=.1$, $d=.02$, $e=f=g=.7$.

includes an STS component). On the other hand, if the number of samples remains high, but rechecking is eliminated, a large increase across recall periods is still predicted. Finally, if the assumptions are changed so that each recall period ends when a stopping criterion is reached, a fairly large "hypermnesia" effect is still predicted (for either type of stopping rule).

The various effects and predictions of this section have all been concerned with cumulative recall over time. We now wish to look at a finer-grained temporal aspect of recall, interresponse times (IRTs). The basic data of interest were collected by Murdock and Okada (1970). Each of 72 subjects was given 20 lists of 20 words each to free recall. Words were presented visually, and the free recall was spoken and tape recorded. Two presentation rates were used, 1 word/sec or 2 words/sec, but these were collapsed together in the reported analyses. Figure 15 shows the mean interresponse time between each of the consecutive ordinal output positions, partitioned separately for each different number of total words output. Because there were insufficient data when fewer than four or greater than nine words were recalled, these curves are not shown.

We did not attempt to fit the exact data of Murdock and Okada (1970), since their data include a STS-component. Simulation of their experiment would therefore necessitate the prediction of retrieval rates from STS. Because our main interest is on retrieval from LTS, this did not seem to

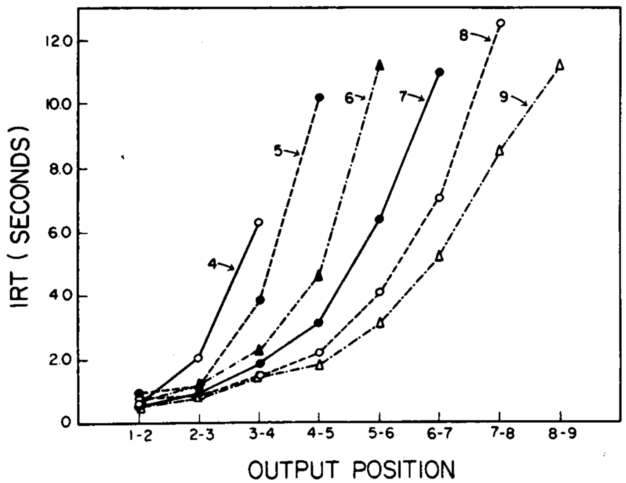

Fig. 15. Mean interresponse time (IRT) as a function of ordinal position in output. Data partitioned according to total number of words recalled (4–9). (From Murdock & Okada, 1970.) Data include STS retrieval.

be desirable. We set list length equal to 15 and presentation rate equal to 2 sec per item. The same model was used as for Roberts' data, except that the STS-buffer was cleared before recall began, and the rechecking assumptions are slightly altered. Since the subjects in this study used spoken, not written, recall, they would not have all previously recalled items available. It was therefore assumed that after the K_{MAX} criterion had been reached, the "subject" continues the search via a rechecking process which uses as cues items recovered subsequent to the K_{MAX} point, even if these items had already been recalled. It was assumed that the number of failures in this second phase of search, K_{MAX2}, would be set equal to the total recall in the first phase multiplied by L_{MAX}. This assumption makes the rechecking effort similar to that assumed for earlier versions of the model. In phase two, search begins with the context cue, and as soon as *any* recoverable item is sampled, switches to it as a cue along with context. Then this cue set is used until a new recall occurs, in which case a switch is made to the new item as a cue, or until L_{MAX} failures accumulate. In this case, the context cue alone is used in the successive samples, until a recoverable item is reached. This process continues until K_{MAX2} failures are reached.

Figure 16 shows the predictions, based on 5000 simulation runs with Robert's parameters. The predicted curves are very similar to the empirical curves obtained by Murdock and Okada (1970). Several features that were noted by Murdock and Okada are also evident in the simulated data. First of all, the interresponse times increase in a positively accelerated fashion as recall proceeds. Second, for a fixed output position, the interresponse times were shorter the more words there were yet to recall. Finally, at any given output position the interresponse time is a good predictor of the number of words yet to recall. Of course these predictions are based on the particular rechecking assumptions that have been made, and a different set of assumptions would undoubtedly shift the predictions (for example, without *any* rechecking or a comparable process the predicted curves are almost linear, with decreasing slopes toward the right of the figure). The lesson from this simulation is simply that IRT results like those of Murdock and Okada are quite consistent with a SAM-like model.

C. CATEGORIZED FREE RECALL: CUING, OUTPUT INTERFERENCE, TEST ORDER

In categorized free recall the list of words that is presented to a subject is divided into a number of conceptual categories (e.g., four-footed animals, professions, tools, etc.). The words belonging to a particular category may be presented contiguously (blocked presentation) or in random

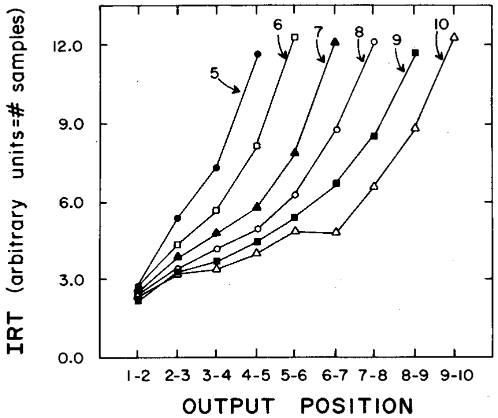

Fig. 16. Predicted mean interresponse times partitioned according to the total number of words recalled (5–10); STS retrieval deleted from predictions. Parameters as in Fig. 8; list length = 15; presentation time = 2 sec/word; special model assumptions given in text.

list positions (random presentation). The present theory will be applied to a blocked presentation procedure only. (In any event, the results mentioned by Cohen, 1966, suggest that if subjects are aware of the categorical nature of the list there are no qualitative differences in the gross results between blocked and random presentation.)

In principle, extending the SAM model to the categorized situation is quite simple. Just as is the case for context information, it is assumed that category information is stored as part of each image, and that category information may be used as a cue. Let us denote an item in a cued category by using a prefix c, and an item in a different category than that of the cue by using a prefix nc. Then the SAM model sets the strength of category cue to c-image to be a linear function of rehearsal time, and the strength of category cue to nc-image to be equal to a residual value. Furthermore, there are separate increments for the category to c-image strength, and the category to nc-image strength, when either of these is recovered and output in the presence of the category cue. Such a model has been fit quite successfully to a variety of data from categorized paradigms.

We have found, however, that not all this machinery is needed to capture the essence of the mechanisms of the categorized situation. A much less powerful model is quite capable of predicting almost all the effects. This simpler model ignores all interitem strengths and retrieval routes (i.e., items cues are not used). Instead, only context and category

cues need be considered. It is this model that shall be presented and utilized in the following sections, since it illustrates the main points without confusing the issue.

Since we are not concerned with serial presentation position effects in the categorized task, it is also much simpler to replace the buffer storage system with a fixed strength assumption. In particular, the context to image strength is set equal to a times the presentation time for that item, and the category to c-image strength is set equal to B times the presentation time. The category to nc-image residual strength is set equal to D. The increment for context-to-image strength was set equal to e and the increment for category to c-image strength was also set equal to e.

An excellent demonstration of the power of even this simplified model to explain the results of categorized studies may be obtained by applying it to a well controlled study by Tulving and Pearlstone (1966). They varied three independent variables: (1) list length—12, 24, or 48 list items, (b) number of words or items per category—1, 2, or 4 items per category, and (3) type of recall test—either a cued or a noncued recall test. In noncued recall the subjects were given a standard free recall instruction, i.e., they were told to write down all the words they could remember as having been on the list. In cued recall the subjects were given a list of all the category names and then tried to recall as many words as possible. In this experiment the members of each category were presented in a blocked fashion, preceded by the category name. Subjects were instructed carefully that they were to remember only the category members, not the category names. The presentation time was 1 sec for each item and 3 sec for each category name. The amount of recall time given was proportional to the list length (1 min for every 12 items). Following the first recall test all subjects were given a second recall test. This second test was always a cued recall test. The results for the first test are given in Figs. 17 and 18 as the solid points.

The application of the model to this data is fairly straightforward. In the case of cued recall it is assumed that each category cue is used until a criterion of L_{MAX} total failures is reached (no rechecking). The items are sampled using both the context and the category cue. The probability of sampling is therefore proportional to the product of the item-to-context associative strength and the associative strength between that item and the category that is tested. The probability of recovery is given by the usual exponential transformation of the sum of these two associative strengths. Retrieval of an item outside the category being tested is assumed to be a failure, on the reasonable basis that a subject always recognized whether an item belongs to the category being tested. The contextual and category associative strengths are incremented upon successful recall of an item.

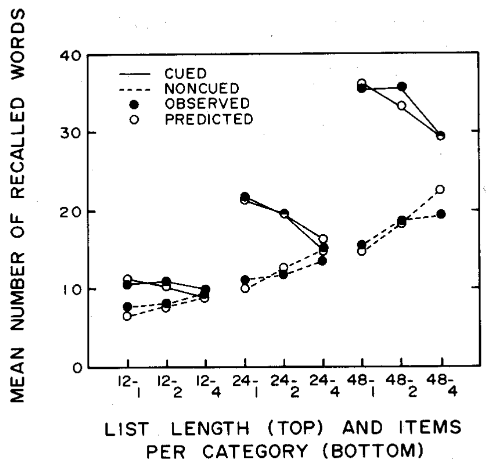

Fig. 17. Mean number of words recalled in the first recall test as a function of list length and words per category, for testing with or without category names provided. (Data from Tulving & Pearlstone, 1966.) Predictions from simplified SAM model, described in text.

Note that due to the assumption of a nonzero residual strength of the association between a category and an item belonging to a different category the model predicts an effect of the number of, and the contextual associative strength of, the items belonging to other categories.

In the case of noncued recall it is assumed that the subject first samples one of the items using only the context cue. Upon successful retrieval of an item the subject will generate (with probability 1.0) the category name of which that word is a member. Both the context and the category strengths are incremented. Next, the subject tries to recall items from within that category until he reaches a criterion of L_{MAX} failures. Contextual cues as well as the category cue are used in this restricted search. As before, retrieval of an item outside the category tested is counted as a failure. Thus, this category search is exactly the same as in the case of cued recall. When the criterion of L_{MAX} failures has been reached the subject discards the category cue and continues sampling using only the context cue. This goes on until a criterion of K_{MAX} total failures with the *context* cue has been reached. Note that failures that are made during category searches are not counted as part of these K_{MAX} failures. Thus, recall stops when the subject believes that he can find no more new categories.

The above model for categorized free recall is similar to the model proposed by Shiffrin (1970) and Rundus (1973). However, neither of

these models allows for sampling items outside the cued category. Therefore, they have difficulty explaining total list-length effects upon within-category recall, and order effects of testing successive categories in cued recall (e.g., Smith, 1971; Roediger, 1973).

The model was fit to the data in Figs. 17 and 18 and a good fit was obtained for the following parameter values: $a = 1.2$; $B = 1.7$; $D = 0.2$; $e = 2.0$; $L_{\mathrm{MAX}} = 15$; $K_{\mathrm{MAX}} = 20$. The predictions are also given in Figs. 17 and 18.

Inspection of Figs. 17 and 18 reveals that the advantage of cuing with the category name decreases with increasing category size (for constant list length) and increases with increasing list length (for constant category size). The model predicts these effects because with increasing category size relatively more categories are accessed in noncued recall, thereby eliminating the advantage of the cued group. With increasing list length, however, relatively fewer categories are accessed in noncued recall, which increases the advantage of cuing.

These data were analyzed by Tulving and Pearlstone (1966) in terms of the two response measures used earlier by Cohen (1963); category recall (R_c), the number of categories of which at least one member was recalled, and words-within-category recall ($R_{w/c}$), the ratio of the total number of words recalled to the number of categories recalled. Thus, a

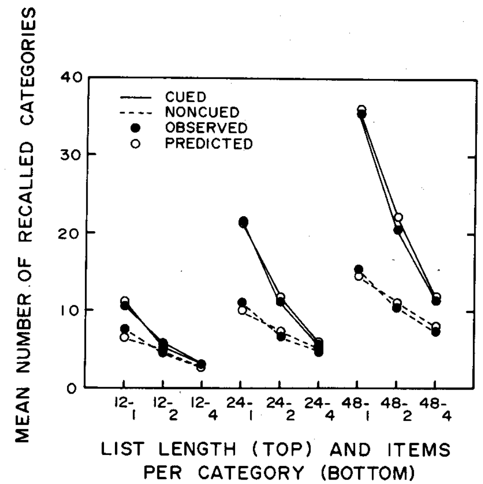

Fig. 18. Mean number of categories recalled in the first recall test as a function of list length and words per category, for testing with or without category names provided. (Data from Tulving & Pearlstone, 1966.) Predictions from simplied SAM model, described in text.

category is not considered to be "recalled" when no member of that category is recalled. Analyzed in this way the data show that the probability of recalling a category was higher for the cued group, that this probability decreased with increasing list length in both the cued recall and the noncued recall condition, and increased with increasing category size in both conditions. Naturally (judging by our fit), when the predicted data are analyzed in the same way, the same effects are obtained. The reasons why the model predicts these results are evident, if we keep in mind that sampling *within* a category is predicted to depend upon the number and strengths of items in *other* categories, due to the residual associations, D. Thus, for example, the probability of recalling an item from a cued category will go down if there are more items in other categories on the list.

In seeming conflict with this reasoning is the finding that words recalled per category having at least one recalled member did not vary much with total list length (in both the data and predictions). However, this seeming paradox disappears when the data are reanalyzed in a nonconditional fashion. The possibility of *no* recalled members from a cued category must be taken into account. Figure 19 gives the observed data and the predictions for items per cued category (with unconditional scoring). Clearly the list length effect is present in both data and predictions.

One other interesting finding observed by Tulving and Pearlstone (1966) for the first test is that the probability of recalling a member of a category, given that at least one member was recalled, was a decreasing

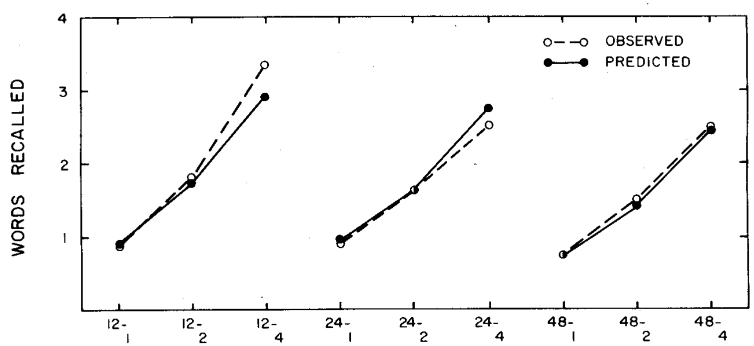

Fig. 19. Predicted and observed mean number of words recalled per cued category in test one as a function of list length and words per category. (Data from Tulving & Pearlstone, 1966.) Predictions from simplified SAM model, described in text.

function of category size. This is of course similar to the usual list-length effect in (uncategorized) free recall, and this result is predicted by our model for exactly the same reasons.

As was mentioned above, Tulving and Pearlstone (1966) gave all subjects a second recall test which was always a cued recall test. They found, of course, that cued recall showed a large increase over a previous noncued recall for the usual reason—the cues gave access to additional categories. Of greater interest are comparisons among the three cued tests: test 1 cued, and test 2 cued after either uncued or cued test 1. Tulving and Pearlstone found test 2 cued after test 1 cued gave recall practically identical to test 1 cued. However a previous *uncued* test seemed to reduce test 2 cued performance. This is illustrated in Fig. 20 at the *top*. The differences were not explained by Tulving and Pearlstone.

In the bottom of Fig. 20 we show the predictions when the present SAM model is applied to the second test. In this application the final test matrix after the first test (uncued) was for each simulation run the starting

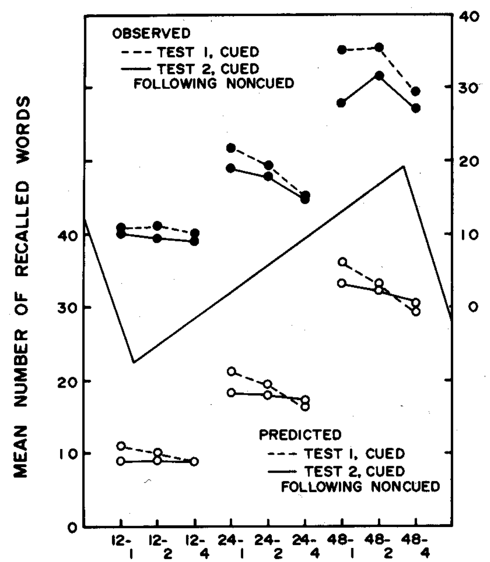

Fig. 20. Mean number of words recalled in cued recall as a function of list length and words per category for subjects who were given these cues either on the first recall test, or not until the second recall test. (Data from Tulving & Pearlstone, 1966.) Predictions from simplified SAM model, described in text.

test matrix for the second test (cued). Otherwise the second test assumptions were identical to those used for a cued first test (including the occurrence of new incrementing for items recalled in test 2, regardless of whether test 1 incrementing had already occurred). Clearly the predictions are matching the main features of the data, but why does a noncued first test hurt a cued second test? The answer depends upon the context-item increments that take place in test 1. In test 2 these increments tend to cause sampling of the previously recalled items in test 1, to the exclusion of items (and categories) that were not recalled in test 1, especially when category size is small.

There is one other set of important findings in categorized free recall that may be considered in the context of our model. These findings concern cued recall; they show that the probability of recall of a category member decreases slightly but systematically as successive categories are cued. This result seems to have been found first (independently of each other) by Dong (1972) and Smith, D'Agostino, and Reid (1970). More systematic studies are reported in Smith (1971) and Roediger (1973). In the experiments of Smith (1971) blocked presentation of categorized words was followed by cued recall. A significant decline in word recall for successive categories tested was observed. This output interference effect was not dependent on the inclusion of the last input category nor was it decreased by introducing an interpolated task between study and test. Thus the results cannot be attributed to a short-term forgetting process. More output interference was observed when a long recall time per category (60 sec) than when a short time (30 sec) was given. In one of Smith's experiments (Smith, 1971, Exp. IV) presentation time per item and category size were varied in a between-list design. More output interference was observed with longer categories and with a higher presentation rate. Roediger (1973) varied category size within a single list and found no effect of category size. Thus, one may conclude that the effect depends on the absolute number of items previously recalled rather than on the number of items per category stored in memory.

Roediger (1973) observed that the probability of recall for successively tested categories decreased in an approximately linear fashion with a slope of about $-.007$. Figure 21 shows the predictions of our model for cued recall for a list of 20 categories of 4 items each. These results are based on 1000 simulation runs with the parameter values that were estimated from the results of Tulving and Pearlstone (1966). It is evident that the model predicts this output interference effect. In our simulation we obtained a slope of $-.0063$, so the magnitude of the effect is also predicted quite well.

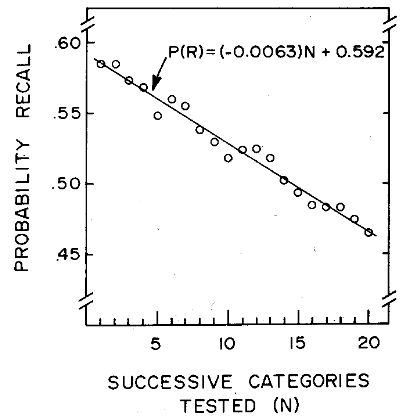

Fig. 21. Predicted probability of recall as a function of the output (i.e., test) position of the category in cued recall. Best fitting regression line is shown. The model assumed 20 categories of 4 words each, and used parameter values equal to those utilized for the fit to the Tulving and Pearlstone (1966) data (see Figs. 17–20, and the text).

The output interference prediction is due to the *incrementing* of context-to-item associations, when there are residual associations between category cue and nc-image. Later in the sequence of tested categories, there is an increasing tendency to sample recalled items from earlier categories, because their associations to context had been incremented. The model also predicts that the effect will be stronger the lower the initial strength of the context associations because the increment after retrieval will than be relatively higher. Thus, the model predicts that higher presentation rates (i.e., shorter presentation times per item) should lead to more output interference as was found by Smith (1971). For these same reasons, our model is consistent with the result obtained by Roediger (1973) that the output interference effect depends on the absolute number of items recalled previously.

A similar explanation handles an effect noted by Roediger (1978): providing some of the category names as retrieval cues increased the number of words recalled from the cued categories [i.e., the positive cuing effect observed by Tulving & Pearlstone (1966)], but decreased the number of words recalled from the noncued categories (i.e., the negative cuing effect observed by Slamecka and others, see below). In this case, there will be a tendency to sample the words recalled earlier from the cued categories, due to incrementing. (Note that such an explanation does *not* handle Slamecka's part list cuing effect, however, as discussed in the next section.) Verification of our reasoning comes from another condition in Roediger (1978): one group was given the category names with the instruction not to recall from those categories and another group was given those names with the instruction to recall especially from those categories. Relative to a control group a large decrement in the number of critical words recalled was observed for the second group but not for the first group. These results show that it is the act of recall that produces the interference.

If the category cue to nc-image residual, D, is raised, the model predicts cued performance to drop, but the magnitude of the output interference effect to remain virtually unchanged. For example, when we doubled the value of D in the simulation, the recall level in Fig. 21 dropped by 10%, but the slope of the function did not change. One way to increase D experimentally is to use categories that are more similar to one another. Roediger and Schmidt (1980, Exp. III) carried out such a study, and found just this predicted pattern of results. (Roediger & Schmidt, 1980, Exp. IV, showed a similar effect in cued recall of paired associates, a finding matching the predictions of the theory, as shall be described in the section on paired associates below.)

To summarize all these findings concerning categorized free recall, our very simple SAM model, without interitem associations, proved capable of predicting all the major results from this paradigm, including cued, and uncued, and partially cued recall, and the output interference effect. Furthermore, these results and simulations demonstrate clearly the need for residual associations between category cues and noncategory items from the list, to explain list length effects on within-category recall, and to explain the output interference effect. In addition, the need for incrementing is clear, to explain the output interference effect. Thus, although these factors are not needed in the model to deal with simple free recall of uncategorized lists, both residual associations and incrementing are necessary components of SAM. (The basis for interitem residuals will be discussed later.)

Finally, one might ask whether additions to our simplified model, such as word cues and interitem search routes, or recalls from categories different from the cue, can present any difficulties for SAM. Such changes add quite a few processes and parameters to the model; we have applied such an extended model to this data, with equal success, but do not present the results since no new insights are gained. However, we shall give such an extended version of the model when turning to our next application, where the two categories in a list consist of (1) random words, and (2) complex pictures.

D. FREE RECALL OF PICTURES AND WORDS

The study on pictures and words was carried out and analyzed by Gary Gillund at Indiana University. We give only a brief resume of the major points here (see Gillund & Shiffrin, in preparation). Some lists contained only words, others only pictures, and some contained some of both. The numbers of words and pictures in these various lists were covaried. In mixed lists, presentation was either blocked, or was alternated as evenly as possible. Arithmetic was used after presentation and before recall to empty STS. Recall of pictures was obtained by the method of Shiffrin (1973): the subjects wrote very brief descriptions of each recalled picture. Then, after recall was completed, all the list pictures were shown to the subjects, who matched their descriptions to the pictures. List lengths used were 10 and 20, and presentation time was 2 sec per item.

Some of the main results are shown in Figs. 22 and 23. Note that the usual list length effects were obtained in both pure and mixed lists, but

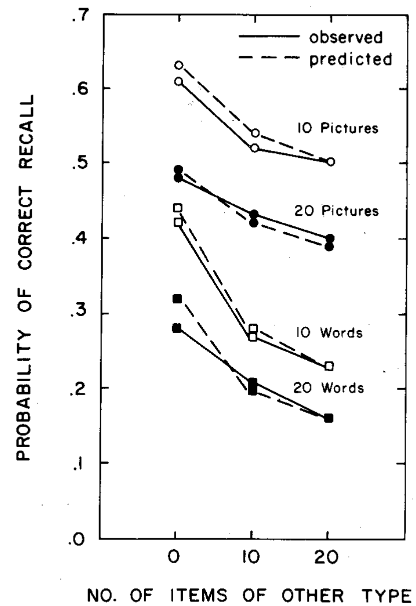

Fig. 22. Probability of recall for pictures and words as a function of both category size and other-category size. Predictions from the SAM model described in text.

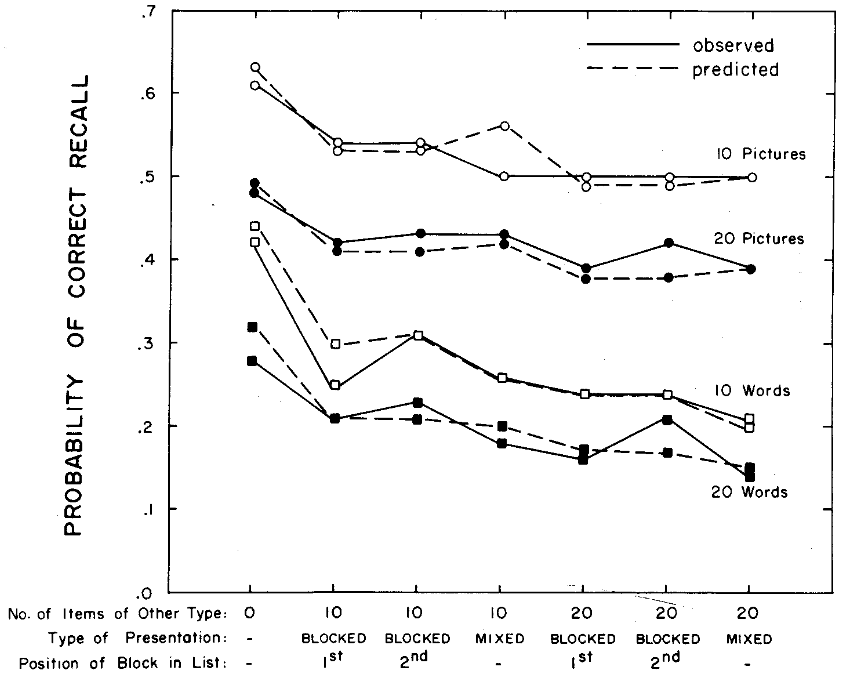

Fig. 23. Probability of recall for pictures and words, as a function of blocking order of testing of each block, and mixing, for each category size, and other-category size. Predictions from SAM model described in text.

also a large effect was obtained for recall of pictures when only the number of words was varied, and vice versa. This is of some interest since the words were chosen to be of low imagery value, and the pictures were complex and not easy to describe succinctly or accurately in words.

The model applied to this case is the basic SAM model for two categories (pictures and words), but with all residual associations, inter-item searches, category searches, and so forth, included. The storage process utilizes a buffer of size 4 for words, and size 1 for pictures; whenever a picture is presented, the previous buffer contents are cleared; when a picture is in the buffer a presented word replaces it. Words build up item–context (a_{w}) and item–item (b_{w}) strengths as a function of rehearsal time; pictures build up item–context strength (a_{p}) as a function of rehearsal time. An item's self-association strength is set equal to the context strength. Residual associations fill the rest of the test matrix: picture–picture (d_{pp}), picture–word (d_{pw}), word–picture (d_{wp}), and

word–word (d_{ww}, for jointly unrehearsed words). In addition, it is assumed that a category cue may be used during retrieval. Thus a within category residual, d_{cc}, and a residual from a category cue to an item in the other category, d_{nc}, are also assumed.

The retrieval plan is fairly straightforward. When a new item is recalled, the next cue-set consists of the recalled item, its category, and context. The search begins with the context cue and a randomly chosen category cue. Any time a cue-set including a word category cue fails L_{MAXW} consecutive times, or a cue set including a picture category cue fails L_{MAXP} consecutive times, the next cue-set consists of a change in category cue, and context (no item cue).

When an item is recalled, increments of the cue to image associations take place: context–word e_{w}; context–picture e_{p}; word–word and word–category–word, both f_{ww}; picture–picture and picture–category–picture, both f_{pp}; word–picture and word–category–picture, both f_{wp}; picture–word and picture–category–word, both f_{pw} (self-increments were equated to context increments). When the total failures (including those accumulated during item cue searches) reach K_{MAX}, search ceases. No rechecking is assumed.

One could guess that this model has the power to deal with the data in Figs. 22 and 23. In fact, many combinations of parameters give more or less equivalent fits to the data. Just as with the category data, more insights into the model would probably be gained by applying simpler versions of the model. Nevertheless, the assumed processes and parameters represent logically necessary components of the SAM model, and it is of some use to show the predictions of the complete model. The predictions in Figs. 22 and 23 represent the results of a rough parameter search, with $a_{\mathrm{w}} = .29$; $a_{\mathrm{p}} = .56$; $b_{\mathrm{w}} = .01$; $d_{\mathrm{pp}} = .0305$; $d_{\mathrm{wp}} = .03$; $d_{\mathrm{ww}} = .02$; $d_{\mathrm{pw}} = .025$; $d_{\mathrm{cc}} = .0385$; $d_{\mathrm{nc}} = .005$; $e_{\mathrm{w}} = .03$; $e_{\mathrm{p}} = .007$; $f_{\mathrm{ww}} = .03$; $f_{\mathrm{pp}} = .007$; $f_{\mathrm{pw}} = f_{\mathrm{wp}} = .001$; $L_{\mathrm{MAXW}} = 3$; $L_{\mathrm{MAXP}} = 4$; $K_{\mathrm{MAX}} = 32$. Quite clearly the model captures the main aspects of the data. It is reasonable to conclude that complex pictures, and words, are comparable entities that may be treated similarly in retrieval (at least within the context of a SAM-like model).

E. PART-LIST CUING

One of the more remarkable findings in free recall, primarily because it does not seem consistent with traditional associative theories of memory, is known as the part-list cuing effect (e.g., Slamecka, 1968, 1969). Suppose after list presentation that a random subset of the list items is presented to the subjects in the cued group, who are told to use them as cues to aid in recall of the remaining list-items, called "critical items." The control group is given no cues, and recalls freely, as usual. The control group actually recalls slightly *more* critical items than the cued group. This finding is robust, being found consistently in random lists, and within categories of categorized lists.

The mystery of the control group advantage was first discussed by Slamecka (1968, 1969). He argued that at least some of the critical items that were not recalled by the control group should have been recalled by the cued group due to the presence of cues that would not have been retrieved by the control group. This argument depends upon the formation during storage and use in retrieval of interitem associations. Slamecka and many later theorists therefore concluded from the part-list cuing effect that such interitem associations could not have been both stored and used in retrieval.

This reasoning is not, however, correct. We shall show next that a prediction of the part-list cuing effect is inherent in SAM-like models, despite the heavy use of interitem associative structure that is made in such models. In fact, it is this very structure and its use in retrieval that produces the effect. This entire problem is discussed in all possible variations, and the literature thoroughly reviewed, in Raaijmakers and Shiffrin (1981). In this article, therefore, we shall summarize these matters in very brief fashion.

One of the most surprising findings related to the part-list cuing effect concerns the effect of increasing the similarity of the list items to each other. Slamecka (1968, Exp. VI) used three lists: (1) 30 rare words; (2) 30 common words; (3) a list consisting of "butterfly" and 29 of its most popular associates. The control groups recalled 5.58, 7.04, and 8.50 critical words, respectively. The cue groups recalled 4.70, 6.79, and 8.97 critical words, respectively. Thus increasing similarity almost doubled recall, while only slightly altering the basic effect.

The SAM model will now be fit to an idealized part-list cuing paradigm. We assume that 30 words are presented (of varying interitem similarity in different lists), followed by arithmetic. The control group free recalls as usual. The cued group is given 15 randomly chosen words from the list, and told to use them to aid recall of the remaining words.

The model for the control condition is identical to that used for Robert's data discussed earlier; even the parameters are identical, except that the interitem strength parameter, *b*, is systematically varied for different lists. The model for the cued condition is almost the same, except that the subject is assumed to use the provided cues before reverting to the normal search. In particular, each provided cue is used in the cue-set, along with context, until L_{MAX} failures are reached. A recovery of another

cue word is *not* counted as a failure, except on the second and subsequent recoveries of the same cue word (to make the cue condition comparable to the control condition). Recalled items during this phase of cued search are simply "written down" but not used as cues themselves. If K_{MAX} is not reached when the provided cues are used up, then normal search commences, as in the control condition. When K_{MAX} failures are reached, then all previously recovered items (whether cues or critical items) are rechecked. All parameters are the same as in the control condition.

The predictions for the cued and control conditions, for various values of the interitem strength parameter, *b*, are shown in Fig. 24. Note that recall is predicted to double as *b* increases, but the control group advantage decreases only slightly over the same range.

In Raaijmakers and Shiffrin (1981), these predictions are exhaustively explored, through numerous versions of the SAM model. The basic predictions hold without incrementing, without rechecking, regardless of the particular rechecking assumptions made, whatever the value of K_{MAX} or L_{MAX} (within reason) and regardless of the particular stopping rule used, among other variations. Surprisingly, the control group advantage occurs in the face of a factor favoring the cue group: when an item cue plus context cue are used, recovery probability is higher than when a context

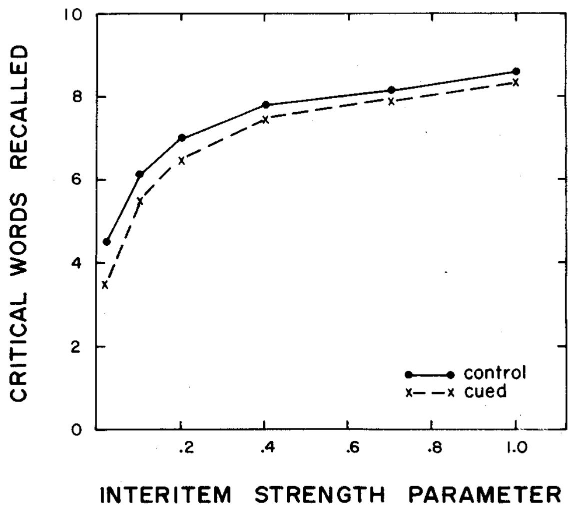

Fig. 24. Predictions for the part-list cuing effect for an idealized paradigm, as interitem strength parameter, *b*, is varies. List length is 30, number of cues is 15, presentation time per word is 2 sec; *d* is set to equal .2*b*. Other parameters as in Fig. 8.

cue is used alone; since the cue group uses more item searches, it obtains an advantage. For example, if recovery probability is set equal to .75 regardless of the cues used, then the cued group inferiority increases by about one-half a word.

What then explains SAM's prediction of the part-list cuing effect? The main factor is illustrated by Fig. 25. Suppose that LTS after presentation of 12 words consists of 4 triads, unrelated to each other but so strongly interrelated that recall of any one item in a triad leads to recall of the entire triad. The critical items and cues are indicated in the figure. Suppose that during search of this structure, the control and cued groups sample an equal number of triads (a simplification for the sake of the argument). The cued group's sampled triads will all contain a minimum of one cue word, and hence a relatively small number of critical words. The control group's sampled triads, on the other hand, will often contain no cue words and hence be relatively rich in critical words. Note well that

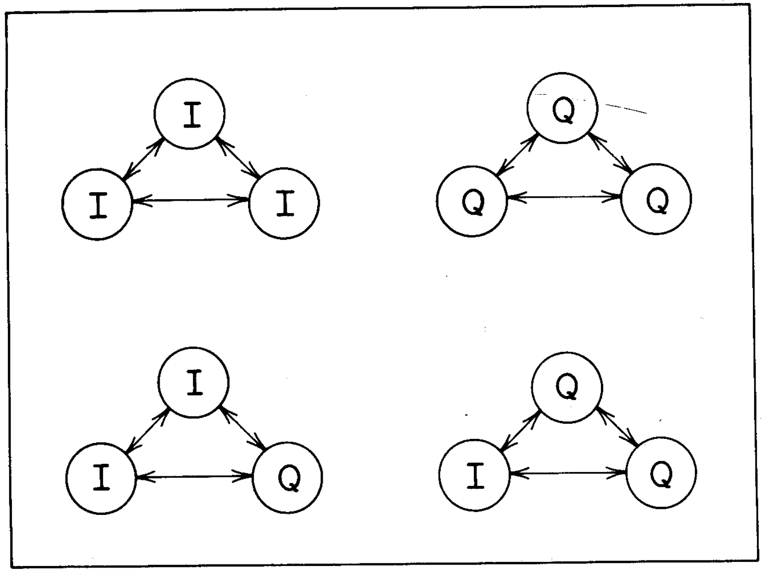

Fig. 25. A simplified associative network for a 12-word list stored as four triads. The six experimenter provided cue words have images denoted Q. The six remaining critical items are denoted I. The arrows denote associations between images. Each image has an association to context, not depicted. A context sample can access a triad rich in critical items (e.g., the triple-I triad). The cue word plus context samples can only sample triads relatively impoverished in critical items, since each such triad must contain at least one cue.

this effect depends upon both a nonuniform structure in LTS, and also extensive and effective use of interitem search.

SAM's prediction of the part-list cuing effect is thus dependent upon the fact that both groups make extensive use of interitem search. The control group uses a mixture of two types of cue sets; one type consists of context cues only, while the other type consists of both a context cue and a word cue that was generated by the subject earlier in the search. On the other hand, the cue group usually uses just one type of cue set, containing both a context cue and a word cue provided by the experimenter. The control group is superior under these circumstances for the reasons given above. Strangely, then, the control group advantage is the result of just the interitem structure and interitem retrieval routes that previous theorists have argued must be ruled out.

We mention, finally, that several other findings in part-list cuing are predicted by SAM. It has been found (especially in categorized lists) that increasing the number of provided cues slightly increases the control group advantage (e.g., Slamecka, 1968; Roediger, Stellon, & Tulving, 1977; Roediger, 1974). To apply SAM, we reduced L_{MAX} to 2 and raised K_{MAX} to 50, to ensure that all provided cues would always be used, and studied the predictions as the number of cues from a 30 item list varied from 0 to 25. The predictions (all other parameters as in Fig. 24) are given in Fig. 26. Clearly the model predicts a slight, almost linear, decrease as the number of cues increases.

A version of SAM has also been applied to the categorized list paradigm, in which cues may be provided from each category. Furthermore, these cues may be "extra list," from the relevant category, but not on the list. Bruce Williams at Indiana has applied the model to this situation. The model is similar to, and an extension of, the model fit by Gary Gillund in Section III,D to the two category situation. It is described in Raaijmakers and Shiffrin (1981). We show here the results only when the model was fit to data of Watkins (1975, Exp. 1). Figure 27 gives the data and predictions. Clearly the effects of number of cues and extra list cues are both well predicted by SAM.

One effect found by Mueller and Watkins (1977) could not be predicted by SAM, despite variations in parameters and assumptions. This effect was the fact that cues from categories other than the tested category did not produce a disadvantage in comparison with the control condition. For assumptions and parameter values which *would* produce a cue condition disadvantage for within-category, within-list, cues and for within-category, extra-list cues, a disadvantage was also predicted for the extra category within-list cues. Thus the fact that a particular version of SAM contains a profusion of parameters and processes does not necessarily

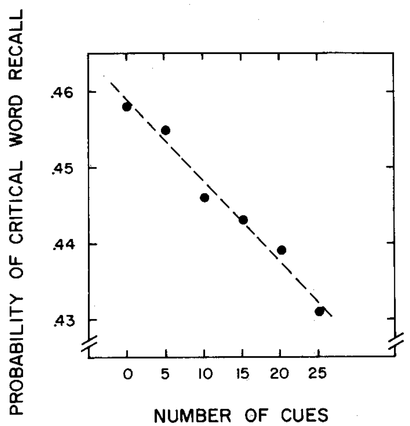

Fig. 26. The same model as in Fig. 24, with $b=.10$, $L_{MAX}=2$, $K_{MAX}=50$. The probability of critical word recall is given as a function of the number of experimenter-provided cues (list length = 30).

reduce the testability of the model. Why does SAM mispredict the extra-category cuing result? We suggest that subjects realize that the cues are from a different category, and thinking such cues to be worthless, ignore them. We suggest that a disadvantage would appear if only the subjects could somehow be induced to *use* the provided cues.

F. PAIRED-ASSOCIATE PARADIGMS

The reader will undoubtedly have noticed that the model developed for free recall contains all the ingredients necessary to predict cued testing of paired associates. We have in fact embarked upon an extensive research program in which paired associates are presented and tested by various methods (free recall, cued recall, recognition). In this article we will present only the first of these studies, and that in brief fashion (see Raaijmakers, 1979, for additional details).

The study was a natural generalization of the simple free recall paradigms that have been discussed. Pairs of items (paired-associates, denoted PA) were included in lists along with single items (denoted FR, for consistency with the previous studies). A single trial procedure was used, so each word was seen only one time. The 10 conditions were:

Number of PA items:	5	5	15	15	5	15	30	0	0	0
Number of FR items:	10	30	10	30	0	0	0	10	30	40

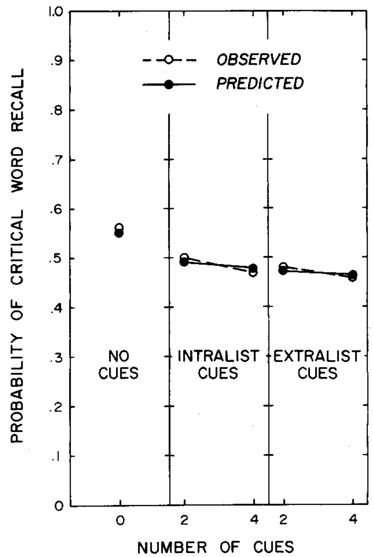

Fig. 27. Predicted and observed probabilities of critical word recall in a category, for the control condition, and for the intralist and extralist cue conditions, each with two or four cues. (Data from Watkins, 1975.) The model is described in Raaijmakers and Shiffrin, 1981; parameters: list length = 36; words per category = 6; presentation time per word = 3 sec; $r = 4$; K_{MAX}/category = 12; $L_{MAX} = 3$; $a = c = .38$; $b = .38$; item-category strength per sec = .38; category-cue and list-word-cue residuals to words on list = .1; all increments = .36; residual strength of extralist cues to list items in same category = .03; product of residual strengths when self-sampling an extra-list cue = 2.2.

(Note that the number of PA items is given in terms of the number of pairs; the number of words is therefore given by twice this number.) PA pairs and FR items were randomly mixed. For each condition, half the subjects were first given cued testing of the PA items (paired-associate testing), followed by free recall of the FR items; the other subjects were tested first on the FR items (free recall testing), followed by cued testing of the PA items. Subjects were not told before study of a list which items would be tested first. The words were presented visually, a single word for 2 sec, a pair for 4 sec. Paired words were tested either in a forward manner or in a backward manner: if the pair was A–B it was tested either as A-? or as ?-B.

Subjects were asked to allot an equal amount of effort in studying each word. The instructions emphasized that they should try to link together the two members of a word pair into a single unit, by forming a mental image or by using some kind of verbal code. After presentation of the list a 20-sec arithmetic task was given to eliminate short-term effects. A written recall procedure was used. Single words were tested using a 2-min free recall procedure, paired words were tested with a paired-associate testing procedure. In this case the subjects had 4 sec to write down their answer.

Figure 28 in the top panel shows the effect of list length on recall of the PA items and the FR items. These data are averaged over order of testing and over testing with the A member and with the B member of the A–B pair. It is evident that the results are quite consistent. In free recall testing the probability of recall decreases not only as a function of the number of FR items but also as a function of the number of PA items on the list. A similar list-length effect is observed for the PA items, and again the probability of recall decreases when other items are mixed in the list. In contrast with the word–picture study, where recall was not directed to one category or the other, the present "cross-category" list length effects take place even though recall is directed specifically to either FR or PA items. The results are similar to those found in cued recall of categories, where recall depends upon the number of items in the other categories.

Figure 29 gives the effects of test order. The main result to note here is

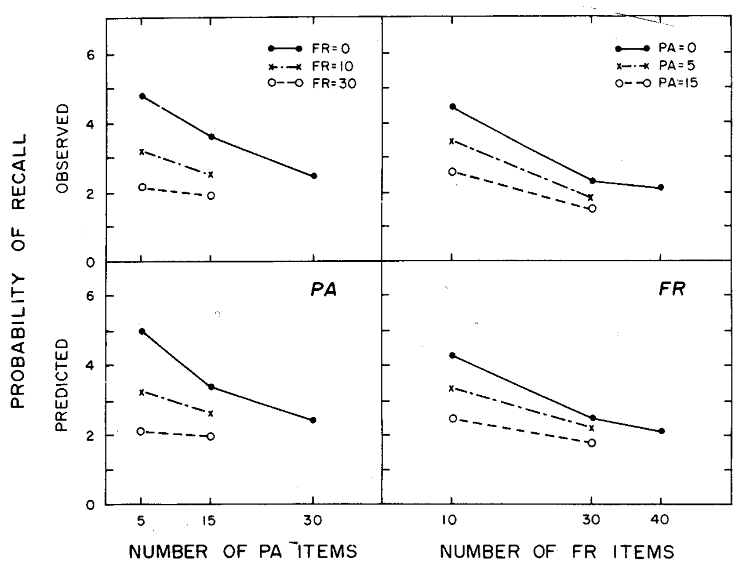

Fig. 28. The effect of the number of single words (FR) and the number of words pairs (PA) on the probability of recalling these two types of items. Predictions based on the SAM model described in the text.

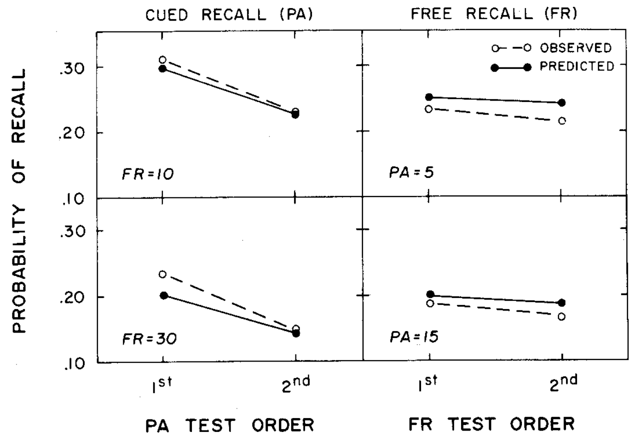

Fig. 29. The effect of the order of the FR and PA tests upon the probability of recall of each type, partitioned according to the number of items of the *other* type. Predictions based on the SAM model described in the text.

that free recall of the FR items is only slightly reduced by a prior cued test of PA items (about .02 reduction in probability of recall, overall). On the other hand, cued recall of PA items is considerably reduced by a prior free recall test of FR items (about .08 reduction in probability of recall, overall).

The final results to which we wish to call attention are given in Fig. 30; averaged across all conditions, this figure gives the overall probability of cued recall of PA items, broken down by test quartiles. That is, this figure shows that cued recall probability drops slightly as the test position of the pair is delayed.

Applying SAM to this study is quite easy, since all the groundwork has already been laid in the models for free recall. The storage assumptions are straightforward. Each pair of PA items clears the buffer. Each FR item clears the buffer of PA items, but adds to any FR items already in the buffer (up to the buffer size, r: then one of the previous buffer members is deleted). A PA pair builds up item–context strength (parameter $= a_{\mathrm{PA}}$) and interpair-strength (parameter $= b_{\mathrm{PA}}$) as a function of rehearsal time (always 4 sec in this case). An FR item builds up item–context strength (parameter $= a_{\mathrm{FR}}$) and item–item strength (parameter $= b_{\mathrm{FR}}$) as a function, respectively, of rehearsal time, and of joint rehearsal time. Note in each case that if there are m individual words in the buffer together for t sec, the rehearsal time for any word, or any pair of words, is equal to t/m.

The remainder of the test matrix is filled with various residual associa-

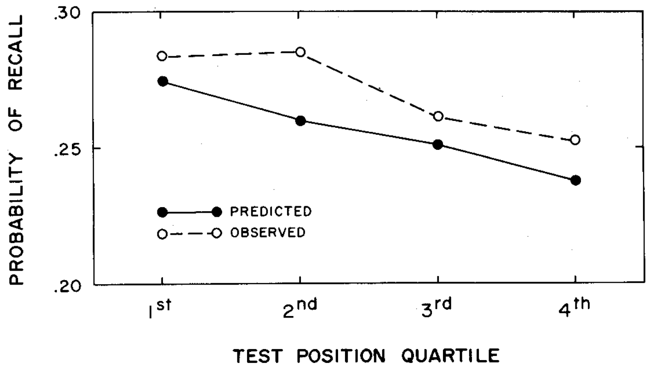

Fig. 30. Predicted and observed recall probability of PA items, for all list types lumped together, as a function of test quartile. Predictions based on model described in text.

tions. Any item's self-association strength was set equal to its context strength. To keep things simple, *all* strengths not mentioned above were set equal to a common residual value, d.

Finally, various increments are assumed after a successful recall. Recall of an FR item is followed by increments of e_{FR} for the context and self-strengths, and f_{FR} for the interitem strength if an item cue is used in the cue-set. Recall of a PA item is followed by an increment of e_{PA} for the context and self-strengths (and presumably by an increment of f_{PA} for the interitem strength, but this parameter is never needed in this application—see below).

Since intrusions of PA items in free recall, or FR items in cued recall, were very rare, we assume that any sampled and recovered item can be classified correctly as to type. Hence any sample of a "wrong" type during search is counted as a failure (and no incrementing occurs). With this proviso, the retrieval model for free recall is identical to that described for, say, Robert's data. For cued recall, it is assumed that each sample is made with both context and the provided item cue, and search ceases when L_{PAMAX} failures is reached (since only 4 sec were provided by cued recall, L_{PAMAX} was arbitrarily set to 1.0). Note that any increment between a PA cue and the correct response has no effect within this model, because such a cue will never be used again, either in PA or FR testing.

Only a very limited parameter search was carried out, but a reasonably adequate fit was obtained for the following values: $K_{MAX} = 30$; $L_{MAX} = 3$; $L_{PAMAX} = 1$; $r = 4$; $a_{FR} = .30$; $a_{PA} = .18$; $b_{FR} = .30$; $b_{PA} = .60$; $d = .035$; $e_{FR} = 3.0$; $e_{PA} = .80$; $f_{FR} = 3.0$. The predictions are shown in

Figs. 28, 29, and 30. It seems clear that the SAM model, essentially the same SAM model used for free recall, can handle the major findings of cued PA recall, even for the case when FR items and PA pairs are mixed within the same list. The list-length effects, the effects of order of PA and FR testing, and the effects of test order in PA testing are all well handled by the model.

In many ways, the theoretical conclusions to be drawn from this study and the application of the model parallel the conclusions reached from the categorized free recall situation. The list-length effects that cross test type, and that appear in cued recall, illustrate the importance of residual associations between items not rehearsed together, and even between items of different types. The effects of order, in both Figs. 29 and 30, illustrate the importance of the incrementing process. Of course, the basic phenomena of cued testing require interitem associations and search routes. Finally, we regard it as a strong point in favor of the model that a system developed for free recall can handle so accurately these various results from cued testing of paired-associates.

IV. General Discussion and Final Comments

We begin by calling attention to a problem that Smith (1978) has termed "the sufficiency/transparency tradeoff." The problem is that as a long-term memory model (especially a simulation model) becomes more and more complex, and increasingly encrusted with special assumptions, it gains the ability to predict a good deal of data (sufficiency), but becomes increasingly opaque to external observers (including the model's creators). That is, it becomes virtually impossible to extract the essential principles from the mass of details and interactions that comprise the model, and it is often impossible to anticipate what the model will predict for a given manipulation.

We have been quite concerned, even for our relatively simple model, with the "transparency" problem, and have adopted a series of measures to deal with it. First, we do not attach much significance to the fact that the model can fit any single study or type of study. Rather, we require the model to apply to many different tasks and types of tasks, with essentially the same set of assumptions, and the same set of basic mechanisms. Second, if the values of parameters are important to predict certain effects, those values should be consistent with the model's rationale and the task requirements. Third, the model should have testable aspects—there should be some results that the model cannot fit (the part-list cuing effect is an example of an inherent prediction of our model—in fact, we saw that the failure to attain a cue condition disadvantage for extracategory cues could not be handled by the model). Fourth, and perhaps most important, the model must be made understandable to the observers. We have attempted to do this by carrying out extensive explorations of the "assumption space" of the model, systematically adding and removing various processes, and examining the shifts in predictions that result. Our model, unfortunately, is so stochastically interactive, that it is difficult to make accurate intuitive predictions even for very simple combinations of assumptions. We have seen this to be the case especially when applying the model to "hypermnesia" and "part-list cuing." In light of our theoretical explorations, we hope the predictions of these effects, and indeed the basic workings of the model, have been illuminated.

Let us review now the basic tasks to which the model has been applied. Serial position effects, but more important, list-length and presentation time effects in single-trial free recall were easily handled. The temporal aspects of free recall were dealt with next, including the effects of instructions to extend the period of active retrieval, cumulative response curves, repeated recall, the effect known as "hypermnesia," and interresponse times. SAM was applied next to the basic phenomena in categorized free recall, not only handling the large effects of cuing, of category size, or number of categories, of mixtures of pictures and words, of the number of categories upon within category recall, and of the test order of categories (the output interference finding), but also explaining the subtle effects of cued recall following noncued recall. The model was next shown to predict the part-list cuing effect in its sundry variations, an important result since previous associative models have had difficulty dealing with the finding. Furthermore, the explanation was not post hoc; the model for free recall was applied "intact" to the part-list cuing paradigm, and the prediction proved to be an inherent property of the model, occurring in almost all model variations. Finally, the model was shown to predict cued recall of paired associates, in lists containing both paired-associates and single items. Since the model for free recall utilized extensive amounts of item-cuing, the extension to the paired-associate situation required no new assumptions. The predicted effects include those of list length, number of PA items, number of FR items, sequential effects during cued testing, and the relationship of free recall to cued recall for different test orders.

These are not the only tasks to which the model has been applied, but space restrictions prevent our presentation of these other paradigms. In brief, they include a variety of other paired associate tasks, and several recognition paradigms. Recognition may well involve an initial judgment of "familiarity," perhaps based on the value of the denominator of the sampling equation. If familiarity does not lead to a response, however, then the rest of the search is treated similarly to that for recall.

Let us conclude by reprising the important features of our retrieval theory. An associative retrieval structure and cue-dependent retrieval are essential, but are common to many theories. The sampling assumptions are the key to the present approach, in several different ways. First, the fact that sampling is probabilistic allows for a considerable degree of resampling in certain circumstances. Such resampling of previously sampled images is the basis for stopping the search, and hence an important contributor to the limitations upon retrieval. Second, the sampling equation [Eq. (1)] provides an explicit basis for combining cues. That is, the multiplication of strengths in an additive ratio rule provides a means of focusing the search when necessary or desired, and allows SAM to predict cued or free recall with equal facility. Turning now to recovery, it is obvious on logical and empirical grounds that some type of recovery rule is necessary (for example, the effects on free recall of doubling presentation time would be most difficult to handle without a strength-dependent recovery probability).

These factors notwithstanding, we make no claim for the uniqueness of the particular mathematical forms of Eqs. (1) and (2). These functions were chosen for simplicity, convenience, and historical factors, but slight variations in their forms would undoubtedly lead to an equally good description of most of the data. It is our position that the basic framework of the model has enough power to handle the data that small variations in quantification will do little to degrade the quality of the predictions. A test of this position must await further empirical and theoretical work.

Let us turn now to some of the subsidiary assumptions of SAM. The inclusion of residual associations makes our retrieval network "completely" interconnected, a rather novel feature. Such interconnectivity is needed to explain list-length effects in various types of cued recall, in both the category and paired-associate paradigms. Incrementing represents learning effects that occur during retrieval; it is essential to explain various types of test-order findings (as in successive testing of categories, for example). Still other factors in our model do not seem crucial for predicting present data, or have not yet been explored theoretically. Such factors include the conditionalization rules that apply after resampling of the same image and rechecking.

Finally, there are subject controlled strategies in our theory, such as search termination rules, and choice of cues at various stages of the search. We have tried to include reasonable strategies in our simulation, but are convinced that a theory would be very weak if its predictions depended in important ways on the choice of particular strategies (since different subjects probably choose different strategies, and the same subject probably changes strategies from time to time). It is for this reason that we have expended considerable effort in this article showing the effects upon the predictions of changing strategies. It is one of the successes of this model that the basic predictions are quite insensitive to "sensible" alterations in retrieval strategies, but that manipulations expressly designed to change strategies (such as encouragement to continue search) have effects in the data that are predicted through simple manipulations of the appropriate parameter in the model.

ACKNOWLEDGMENTS

This research was supported in part by a fellowship from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) to the first author, and Public Health Service Grant 12717 and National Science Foundation Grant BNS-77-00156 to the second author. We are indebted to Gary Gillund and Bruce Williams who carried out part of the research described in this article.

REFERENCES

Anderson, J. R. FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 5). New York: Academic Press, 1972. Pp. 315–378.

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D. C.: Winston, 1973.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press, 1968. Pp. 89–195.

Bousfield, W. A., Sedgewick, C. H. W., & Cohen, B. H. Certain temporal characteristics of the recall of verbal associates. *American Journal of Psychology,* 1954, **67,** 111–118.

Bower, G. H. Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, D. C.: Winston, 1972. Pp. 85–123.

Cohen, B. H. Recall of categorized word lists. *Journal of Experimental Psychology,* 1963, **66,** 227-234.

Cohen, B. H. Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5,** 182–187.

Donaldson, W. Output effects in multitrial free recall, *Journal of Verbal Learning and Verbal Behavior,* 1971, **10,** 577–585.

Dong, T. Cued partial recall of categorized words. *Journal of Experimental Psychology,* 1972, **93,** 123–129.

Erdelyi, M. H., & Kleinbard, J. Has Ebbinghaus decayed with time? The growth of recall (hypermnesia) over days. *Journal of Experimental Psychology: Human Learning and Memory,* 1978, **4,** 275–289.

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

Gillund, G., & Shiffrin, R. M. Free recall of pictures and words. In preparation.

Indow, T., & Togano, K: On retrieving sequence from long-term memory. *Psychological Review,* 1970, **77,** 317–331.

Johnson, D. M., Johnson, R. C., & Mark, A. L. A mathematical analysis of verbal fluency. *Journal of General Psychology,* 1951, **44,** 121–128.

Luce, R. D. *Individual choice behavior: A theoretical analysis.* New York: Wiley, 1959.

Mueller, C. W., & Watkins, M. J. Inhibition from part-set cueing: A cue-overload interpretation. *Journal of Verbal Learning and Verbal Behavior,* 1977, **16,** 699–709.

Murdock, B. B., Jr. The serial position effect in free recall. *Journal of Experimental Psychology,* 1962, **64,** 482–488.

Murdock, B. B., Jr., & Okada, R. Interresponse times in single-trial free recall. *Journal of Experimental Psychology,* 1970, **86,** 263–267.

Raaijmakers, J. G. W. *Retrieval from long-term store: A general theory and mathematical models.* Doctoral dissertation, University of Nijmegen, Nijmegen, The Netherlands, 1979.

Raaijmakers, J. G. W., & Shiffrin, R. M. Search of associative memory. *Psychological Review,* 1981, in press.

Roberts, W. A. Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology,* 1972, **92,** 365–372.

Roediger, H. L.,III. Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior,* 1973, **12,** 644–657.

Roediger, H. L.,III. Inhibiting effects of recall. *Memory and Cognition,* 1972, **2,** 261–269.

Roediger, H. L., III. Recall as a self-limiting process. *Memory and Cognition,* 1978, **6,** 54–63.

Roediger, H. L., III, & Schmidt, S. R. Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning and Memory,* 1980, **6,** 91–105.

Roediger, H. L.,III, Stellon, C. C., & Tulving, E. Inhibition from part-list cues and rate of recall. *Journal of Experimental Psychology: Human Learning and Memory,* 1977, **3,** 174–188.

Roediger, H. L.,III, & Thorpe, L. A. The role of recall time in producing hypermnesia. *Memory and Cognition,* 1978, **6,** 296–305.

Rundus, D. Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior,* 1973, **12,** 43–50.

Shiffrin, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970. Pp. 375–447.

Shiffrin, R. M. Visual free recall. *Science,* 1973, **180,** 980–982.

Shiffrin, R. M. Short-term store: The basis for a memory system. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1). Hillsdale, NJ: Erlbaum, 1975. Pp. 193–218.

Shiffrin, R. M. Capacity limitations in information processing, attention and memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 4): *Memory processes.* Hillsdale, NJ: Erlbaum, 1976. Pp. 177–236.

Slamecka, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology,* 1968, **76,** 504–513.

Slamecka, N. J. Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology,* 1969, **81,** 557–560.

Smith, A. D. Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1971, **10,** 400–408.

Smith, A. D., D'Agostino, P. R., & Reid, L. S. Output interference in long-term memory. *Canadian Journal of Psychology,* 1970, **24,** 85–89.

Smith, E. E. Theories of semantic memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 6): *Linguistic functions in cognitive theory.* Hillsdale, NJ: Erlbaum, 1978. Pp. 1–56.

Tulving, E. The effects of presentation and recall in free recall learning. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 175–184.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972. Pp. 382–404.

Tulving, E. Cue-dependent forgetting. *American Scientist,* 1974, **62,** 74–82.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5,** 381–391.

Watkins, M. J. Inhibition in recall with extralist 'cues'. *Journal of Verbal Learning and Verbal Behavior,* 1975, **14,** 294–303.

Waugh, N. C. Presentation time and free recall. *Journal of Experimental Psychology,* 1967, **73,** 39–44.

Williams, M. D. *Some observations on the process of retrieval from very long-term memory.* Doctoral dissertation, University of California, San Diego, 1977.