3

The SAM Retrieval Model: A Retrospective and Prospective

Richard M. Shiffrin
Indiana University

Jeroen Raaijmakers
TNO Soesterberg

A Brief Reminiscence from the First Author

My professional relationship with Bill Estes began as a graduate student at Stanford in the years 1964–1968. I spent my first year with Gordon Bower and the next 3 years with Dick Atkinson as research advisors, but nevertheless learned much from Bill in classes, research seminars, and discussions. Certainly a central impetus toward the development of the SAM retrieval model took place during a qualifications examination in which Bill asked me how I knew whether the memory effects I was describing arose during memory storage or memory retrieval. Much of the subsequent development of the SAM model and its attendant research has revolved around attempts to deal with this question. Subsequently I spent two different years on leave with Bill at Rockefeller University, and we developed a close friendship as well as a fruitful professional relationship. I have learned much from Bill, about research, science, humor, and life.

This chapter describes the evolution of the SAM model for memory, summarizing the model as described in Raaijmakers and Shiffrin (1980, 1981) and Gillund and Shiffrin (1984), and its precursors, but focusing primarily upon developments since 1984. It is particularly appropriate for this volume as the second author has recently utilized the idea of stimulus fluctuation theory (Estes, 1955a, 1955b) in an extension of SAM to deal with classical interference phenomena (Mensink & Raaijmakers, 1988, 1989).

PRECURSORS

Atkinson and Shiffrin (1968) discussed a search model for retrieval, but were most concerned with processes and control processes in short-term store. These short-term processes are still an important component of the model, but have not recently been a focus of empirical research or further theoretical development. Shiffrin and Atkinson (1969) discussed the structure of long-term store, and the way in which such a structure is used to facilitate retrieval. Both articles assumed unitized separate traces are stored in long-term store, and search to consist of random sampling of these traces. Applications of the search model were limited to free recall. Shiffrin (1970) quantified these notions more precisely, and proposed stages for the search of memory. The model was applied both to free recall studies and continuous paired associate memory studies. The idea of specifying retrieval cues, and their strengths, weights, and combination rules, appeared in the 1980 and 1981 articles by Raaijmakers and Shiffrin. These articles dealt mainly with recall, but pointed out that recognition could be carried out in two ways: (a) as a memory search (a model utilizing this component was adopted by Mensink and Raaijmakers, 1988, to deal with interference among lists when recognition tests are used); and (b) on the basis of the total activation of memory. This latter recognition process, in which the decision is based on a sum of activations across traces, was laid out in detail in the 1984 article by Gillund and Shiffrin. Admitting that both search and global summation could be used to recognize, Gillund and Shiffrin proposed that the global summation process is used predominantly.

THE *SAM* MODEL CIRCA 1984

Early formulations of SAM were designed to incorporate as few representational assumptions as possible, but two lay at the core of the theory: (a) experience is divided into relatively unitized events (partitioning occurring through the operation of control processes in short-term store); and (b) events are stored in memory as separate *images*. Both these assumptions are crucial to recent developments and are discussed later in the chapter.

Images (I_i) contain many kinds of information, consisting of some proportion of the information rehearsed and coded together in short-term store. For example, if a pair of words is presented, the information stored might contain sensory, lexical, semantic, and conceptual information about the two words, information linking, relating, or associating the words, and a wide variety of context information (e.g., the environment, the setting, the thoughts and feelings of the subject). The information is unitized in the sense that there is a tendency for it to be retrieved as a group during one stage of a memory search, in a recall task.

The information in long-term store (LTS) is always accessed with the use of retrieval cues (Q_j). These cues have properties similar to those of images: Each is relatively unitized and separate. Several cues can be used together to probe memory, each being assigned a *weight*, w(j), representing that cue's relative salience. Retrieval is assumed to be (at least weakly) a limited capacity system; that is, cues cannot be accumulated in the probe set in indefinite numbers without some cost. The Raaijmakers and Shiffrin 1980 and 1981 articles were concerned with situations involving just two cues, so capacity limitations were discussed but not incorporated explicitly in the fits of the model to data. However, Gillund and Shiffrin (1984) varied number and types of cues in recognition and collected evidence that the sum of all weights should add to a constant (a constant that was set to 1.0 in their article for convenience; the limitation on retrieval capacity was needed also in subsequent research such as that by Gronlund & Shiffrin, 1986, Clark & Shiffrin, 1987, and Shiffrin, Murnane, Gronlund, & Roth, 1989). Examples of cues are a *word*, and *context*. We assume that a context cue is always one of the cues used in any episodic memory task; its role is the focusing of retrieval upon the recently presented information that is being tested.

Each cue has a tendency to activate each memory image, represented by a weighted strength,

$$S(Q_j, I_i)^{w(j)} = A(j,i).$$

The strengths are determined by rehearsal and coding processes carried out in short-term store during study, and by preexisting relationships between cues and items. Total activation of image I_i is determined by the product of the weighted strengths across all cues:

$$A(i) = \prod_j A(j, i).$$

In a free or cued recall task, a search of memory is carried out. The search consists of a series of search cycles. On each cycle cues and weights are selected, a sample of one of the memory images is made, information is recovered from that image and assessed, a recall may be emitted, new information may be added to memory, and a decision is made whether or not to continue the search.

The probability of sampling image I_i given a set of cues and weights $\{Q_k; w(k); k = 1, \ldots, m\}$ is just the ratio of the activation strength of I_i divided by the sum of the activations of all the images:

$$P_s(I_i) = \frac{A(i)}{\sum_j A(j)}. \tag{1}$$

Once an image has been sampled, the information in it must be recovered and used to generate an appropriate decision and response. The simplest case occurs when the sampled image contains just one word, and has not previously been sampled using any of the present cues. In this case the probability of being able to recover the correct name is assumed to be:

$$P_R(I_i) = 1 - \exp\left\{-\sum_j w(j)S(Q_j, I_i)\right\} \qquad (2)$$

where the sum is taken over the m cues. Special assumptions are used when an image has previously been sampled using one or more of the present cues and the previous recovery did not succeed. In this case the components of the sum in Equation 2 corresponding to the previously unsuccessful cues are removed (in effect, only one recovery chance per cue is allowed for a given image).

When a successful recovery does occur, it is assumed that the strengths between each of the cues used and the image in question are increased (a process called *incrementing* in previous work; e.g., in Raaijmakers & Shiffrin, 1981, the value added depended on the cue but not the initial level of strength). This process tends to produce an increased sampling probability for previously recalled or recovered items. This in turn means that recall of as yet unrecovered items from memory decreases as recall proceeds, at least if new cues are not utilized (an hypothesis well supported by data; see Bjork, 1988).

To model memory processes in laboratory tasks that utilize presentations of lists of items, particular assumptions are made concerning the rehearsal and coding processes occurring during list presentation; these assumptions determine the strengths between cues and images (except when no coding or rehearsal took place between cue and image; in this case a constant residual strength reflecting preexperimental factors was assumed). For example, when lists of single words are presented, a rehearsal buffer of r items is commonly assumed; the mean strengths between context cue and item image, and between item cue and its own image, are assumed to be monotonically related to the total time that item stays in the rehearsal buffer. Similarly, the strength between an item cue and the image of some other item is assumed to be monotonically related to the total time both items are in the rehearsal buffer together. As another example, if pairs of items are presented, it is generally assumed that only the current pair of items is rehearsed; thus the mean strengths are monotonically related to the presentation time. The monotonic relationship would quite possibly be the same in each of these cases. In most applications to date we have assumed that the relationship is linear. For simplicity, a linear relation is also assumed in the following descriptions.

To be specific, suppose that words are studied (denoted by i, j, . . .), and that the images stored in memory (denoted by I_i, I_j, . . .) encode words, but not more than one word per image. Suppose word i has been rehearsed for t_i seconds, and has been rehearsed for t_{ij} seconds with word j. Then there are retrieval strengths, $S(Q,I_i)$, to image I_i from various cues. The most important cues and their strengths are: (a) word i—ct_i (the self-strength); (b) word j—bt_{ji} (the interitem strength); (c) word x (not rehearsed with word i)—d (the residual strength); and (d) the context cue—at_i (the context strength).

Finally, the retrieval strengths are assumed to have a distribution (due to item variability and many other factors). The variance of the strength distribution does not much affect the recall predictions because the large variance associated with sampling washes out the effect of strength variability. However, strength variability is essential to obtain sensible predictions for recognition, as distributional variance is often the only source of variability in this paradigm. For various technical reasons we have assumed that the standard deviation of the distribution is a constant times the mean (in many cases a symmetrical three-point distribution was utilized; the details are not crucial because the summation of independent activations across images produces a near-normal distribution).

To model retrieval during recall tasks, a strategy is assumed by which the subject generates cues at each stage of the search, and by which the search is terminated. In free recall the subject is assumed to start with the context cue only. Whenever a new item is recalled it is used in combination with context as joint cues. If I_{max} consecutive unsuccessful search cycles occur, then the context cue only is used; if K_{max} total failures occur the search ceases (actually in many instantiations an additional search phase then occurs, termed "rechecking": Each recalled word is used along with context as cues, each until L_{max} samples occur). In cued recall, the subject is assumed to use both context and the provided cue on each cycle of the search, stopping when some fixed number of unsuccessful cycles has occurred.

In the Gillund and Shiffrin (1984) version of SAM, recognition is not assumed to proceed as a memory search (though there is nothing to prevent a subject from searching during recognition if he or she desires; in fact Mensink & Raaijmakers, 1988, have modeled recognition as a search process). Rather it is assumed that global activation determines the decision. In particular, context and the test item(s) are used as cues. The sum of all the image activations (i.e., the denominator of Equation 1) is termed *familiarity* (F). A criterion, C, is chosen. If $F > C$, the subject gives an "old" response; else "new." New storage can also occur during recognition testing; Gillund and Shiffrin (1984) assumed that each item tested resulted in a new image being added to memory with some fixed strengths.

The model in this form has been used to predict numerous findings in free recall, cued recall, and recognition. Raaijmakers and Shiffrin (1980, 1981) demonstrated SAM predicted the following effects: (a) in free recall, the effects of serial position, list-length, presentation time, recall time, repeated recall attempts, picture or word stimuli, ordinal output position on interresponse times, and providing some list words as retrieval cues; (b) in categorized free recall (category cues, strengths, and weights are added to the basic model) the effects of category cuing, cuing with studied and nonstudied category members, number of items per category, recall of one category member upon recall of the others, noncued testing upon subsequent cued testing, and order and ordinal position of testing of categories; and (c) in cued paired-associate paradigms the effects of list length, presentation time, relative proportion of paired versus single items on a list, and ordinal test position.

Gillund and Shiffrin (1984) showed that SAM predicted the effects of most of these variables in recognition paradigms. They also showed for both recall and recognition paradigms that SAM predicted the effects of time available for test, orienting task at study, type of distractor used, type of coding and rehearsal, test expectancy, match of study and test encodings, context shifts between study and test, test delay, serial test position, natural language word frequency (high frequency words are assumed to have higher retrieval strengths than low frequency words), and type of recognition test (cued, single, or paired).

DEVELOPMENTS SINCE 1984

Contextual Fluctuation: Interference and Estes' Stimulus Fluctuation Theory

The original SAM model (Raaijmakers & Shiffrin, 1980, 1981) used context cues primarily to allow the memory search to be focused on the target list of items. The possibility of changes in context between study and text and the implications of such changes were also discussed in these articles and in Gillund and Shiffrin (1984).

Context and changes in context play an important role in the prediction of forgetting phenomena, due to the operation of two basic factors: First, the context cue used after a shorter delay may be more strongly associated to an image than the context cue used after a longer delay. Second, the strength and number of other images associated to the context cue may be greater after a longer delay. Both factors are based on the assumption that the strength of the context cue to an image is related to the similarity of the context at retrieval to the context at storage. When the current context is used as a cue to probe memory, context changes between study and test will lower the strength of activation of the target image and increase the activation strength of images that are more similar to the test context (typically the intervening items). These factors lead to a reduction in the proportion of activation due to the target image, lowering sampling probability in recall and lowering the signal to noise ratio in recognition.

In addition to the type of discrete, discontinuous changes typical for studies that explicitly manipulate the test context (e.g., Godden & Baddeley, 1975; Smith, 1979), there may also be a more gradual type of context change, more typical for the changes occurring within an experimental session. If the experimental paradigm is quite homogeneous in character (as may occur in a continuous paired associate paradigm; see Shiffrin, 1970) then the similarity of study and test contexts might decrease smoothly as the study test interval increases. In most other situations the context changes might well be fairly discontinuous in nature, occurring, for example, at boundaries between lists, at breaks between study and test periods, and at switches between distinct classes of study items. State-dependent learning studies (e.g., Bjork & Richardson-Klavehn, 1989) often include conditions in which attempts are made to reinstate the study context after a delay interval, and performance is shown to improve. In most other studies, however, context similarity will be a nonincreasing function of delay.

Mensink and Raaijmakers (1988, 1989) presented an extension of the SAM model that was designed to handle time-dependent changes in context. In particular, it was assumed that changes in context could be modeled by adopting the notion of fluctuation, as used in Estes' Stimulus Sampling Theory (Estes, 1955a, 1955b). As in the original fluctuation model, the basic idea is that there is a random fluctuation of elements between two sets, a set of available elements and a set of (temporarily) unavailable elements. Performance is a function of the relationship between sets of available elements at different points in time (viz., study and test trials).

In this version of the SAM model, the experimental context is represented as a set of contextual elements. At any given time only a part of this set is perceived by the subject and this subset is denoted the current context. Elements in this set are said to be in the active state. All other elements are inactive. With the passage of time, the current context changes due to a fluctuation process: Some inactive elements become active and some active ones will become inactive. At storage, only active elements are encoded in the memory image. If there are multiple study trials, each study trial gives a new opportunity for encoding a particular element in the image. As in Estes (1955a), the encoding of an element is all-or-none, that is: Each contextual element is either encoded or not in a given image; once an element is encoded in an image, further opportunities for encoding that element have no effect. The context strength at test is proportional to the overlap between the set of context elements encoded in the image and the set of context elements that are active at the time of testing.

Mensink and Raaijmakers (1989) gave general equations similar to those of Estes (1955a) that may be used to compute the probability that any given element is present both at the time of storage and at the time of retrieval, and hence compute the overlap between encoding and test context. Although the general idea is similar to that of Estes, there are a number of important differences. First, in the SAM model, context elements may be "conditioned" to (i.e., stored in) many memory images, not just to one. Second, these elements do not directly determine recall or recognition probabilities, and only contribute toward the activation value for an image. That is, the overlap measures determines the strength of the contextual association to a particular image; this strength combines with the strengths of any other cues in the probe set (using the weights) to determine activation. The activations are then used to determine retrieval (via sampling and recovery functions).

Mensink and Raaijmakers (1988) showed that this elaborated SAM model can handle a number of important interference and forgetting phenomena. Not all of these predictions are specifically due to the contextual fluctuation assumption, however. A number of interference phenomena are due to the increase in the number of images associated to the retrieval cues and not primarily to the strength of these associations.

For example, suppose two lists of pairs are learned in succession, both with the same stimulus terms, but with different response terms (an A-B A-C paradigm). The model predicts both retroactive and proactive inhibition, even when the subject is asked to try to produce the responses associated with each of the two lists (termed *MMFR testing*). The reason for this prediction is simple: retroactive and proactive inhibition in a two-list design occur because the probe cues at test (say A and context) are associated to more images in the interference conditions (say B and C) than in the control conditions (after learning A-B and D-C, say, B will be the strong response to A and context). Increasing the number of trials on the interfering list increases the strength of both the item and the context cues to the "other" response, and hence the interference effect. (This description is conceptually straightforward, although the quantitative fitting of the model to the results of various conditions from a number of studies is far from trivial).

Traditional interference theory (e.g., Postman & Underwood, 1973) always has had great difficulty in explaining the occurrence of retroactive and especially proactive interference on a MMFR test because it assumed that such a test was not affected by response competition. To explain the retroactive interference, it had to be assumed that second list learning led to unlearning of the first list associations. This however could not explain the observation of proactive interference on a MMFR test. The SAM model does not have this problem because it predicts that "competition" (in this model due to the fact that sampling is influenced by other images) will not be eliminated by MMFR testing.

Of the predictions that do depend on the contextual fluctuation model, two are especially interesting. The first has to do with the phenomenon of spontaneous recovery and the dependence of proactive inhibition on the length of the retention interval. In the SAM model, the probability of recalling (sampling) a particular List-1 image is a function of its relative strength. Let s1 and s2 be the contextual strengths for a List-1 and a List-2 image, respectively. The probability that a List-1 image is sampled depends (among other things) on the ratio s1/s2. Mensink and Raaijmakers (1989) show that the fluctuation model predicts that this ratio increases as a function of the retention interval. This implies a relative and possibly an absolute increase in the probability of recalling List-1 responses as the retention interval increases. That is, the model predicts spontaneous recovery of List-1 responses. By the same mechanism, the probability of recalling a List-2 image decreases relative to a control condition (proactive interference). Thus, the model predicts an increase in proactive interference as a function of the retention interval.

A second phenomenon that is due to the contextual fluctuation assumption is "normal" forgetting, that is forgetting in single-list paradigms. As the retention interval increases, the expected overlap between the storage and test contexts, and hence the probability of recall, decreases. Mensink and Raaijmakers (1988) showed that the model predicts that forgetting curves for lists differing in initial associative strength (due to e.g., differing numbers of study trials) fall off more or less in parallel, a somewhat controversial phenomenon observed by Slamecka and McElree (1983).

Recently, this contextual fluctuation model has been used to explain results concerning the spacing of repetitions. Suppose an item is presented twice for study (P1 and P2) and tested at a later time T. If the retention interval (i.e., the interval P2-T) is relatively long, the probability of recall increases as a function of the spacing between the two presentations (the interval P1-P2). With short retention intervals, however, the probability of recall decreases as a function of the spacing between the presentations. With intermediate retention intervals, the results are more complicated, often showing a nonmonotonic effect of spacing.

Unpublished work by Raaijmakers and van Winsum-Westra shows that this complicated state of affairs can be predicted by the present version of the SAM model (supplemented by a few assumptions appropriate for the paradigm). This is in large part due to contextual fluctuation. The reasons for this are basically similar to those described by Estes (1955b). As the interpresentation interval increases, the context at P2 will include more new, not yet encoded, elements that may be added to the memory image. Encoding more elements in the image increases the expected overlap between the test context and the context elements on the image.

Although this sounds very simple, the actual model requires supplementary assumptions that complicate matters. Crucial in this version of SAM is what happens on the second presentation, P2. It is assumed that on P2 an implicit retrieval attempt is made for the image stored on P1 (a study-phase retrieval assumption). New context elements that are present on P2 are only added to the image formed on P1 if that image is successfully retrieved on the second presentation. If it is not retrieved, a new storage attempt is made, based only on the information present on P2; if the attempt succeeds, a new image is stored. In addition, in order to accommodate dependencies due to differential storage strengths, it is assumed that each storage attempt is either successful or not. If it is not successful, the probability of sampling that image on a future retrieval attempt is zero. It is also assumed that no new storage takes place for any item that is still in STS on P2.

In this model, spacing of repetitions has a number of effects. As mentioned, due to context fluctuation more new context elements are stored, provided the item is "recognized" on P2. Second, as the spacing interval increases, there is a corresponding decrease in the probability that the item is still in STS on P2. Both of these effects lead to an increase in the probability of recall at test. However, spacing also has an opposite effect. That is, the longer the spacing interval, the lower the probability that the image is successfully retrieved on P2. This is a simple forgetting effect: As the interval increases, the expected overlap between the context at P1 and that at P2 decreases and this implies a decrease in the strength of the context cue at P2. Together, these factors produce a nonmonotonic effect of spacing. The spacing function shows an initial increase followed by a decrease, the maximum point depending on the length of the retention interval (P2 to T).

This version of the model has been used to fit the results of a number of well-known experiments (e.g., Glenberg, 1976; Rumelhart, 1967; Young, 1971). Although the success of this enterprise is perhaps not surprising (as most of these data have already been fitted by other models), these analyses again show the value of incorporating hypotheses about contextual fluctuation in SAM. In addition, the fact that it can handle data from experiments with many study trials (e.g., Rumelhart, 1967), demonstrates that SAM can handle the basic learning data that were the main focus of the Markov models of the 1960s.

The present model can also explain the intriguing results of Ross and Landauer (1978). According to their analysis, most theories that explain the effects of spacing by some sort of encoding or contextual variability assumption predict that there should not only be a beneficial effect of spacing for two presentations of the same item but also for the two presentations of two different items. That is, there should be an effect on the probability of recalling either of the two items. They showed that such a result is not obtained: A typical spacing effect is only obtained for one item presented twice, not for two items each presented once. The model handles this result because the second presentation of a single item is often stored within the trace formed for the first presentation, but two different items are never stored in the same trace. Because the probabilities of sampling and recovery depend on the overlap in elements with each image separately, the spacing of the presentations by and large only matters for the single item case (the model predicts only a very minor deviation from independence for the two item case).

Taken together, these studies show that contextual fluctuation hypotheses greatly extend the range of phenomena that can be handled by the SAM theory. Of course, this important step is only a start. Issues such as contextual fluctuations within list, and the ability of subjects to select probe context appropriate for the test situation (in autobiographical memory tests, to take one obvious example), are yet to be examined theoretically within the SAM framework.

Composition, Differentiation, and the List-Strength Effect

The previous section discusses interference in the traditional sense, between lists; this type of forgetting is based in part on context shifts and context fluctuations. Interference also occurs within lists, as demonstrated by the list length effect: Longer lists produce poorer memory in recognition, free recall, and cued recall. Ratcliff, Clark, and Shiffrin (1990) studied a related effect. The strengths of some items on a list are varied: If the unchanged list items are harmed by strength increases for other items (or helped by strength decreases), then a (positive) list-strength effect is said to have occurred. A list-strength effect occurs for free recall (e.g., Tulving & Hastie, 1972; Ratcliff et al., 1990). However it is weak at best for cued recall, and missing or negative for recognition (Ratcliff et al., 1990). In seven studies, strength was varied either by the number of spaced repetitions or by the amount of study time. The results are summarized in Fig. 3.1. Points below the diagonal represent positive list-strength findings. None were signifi-

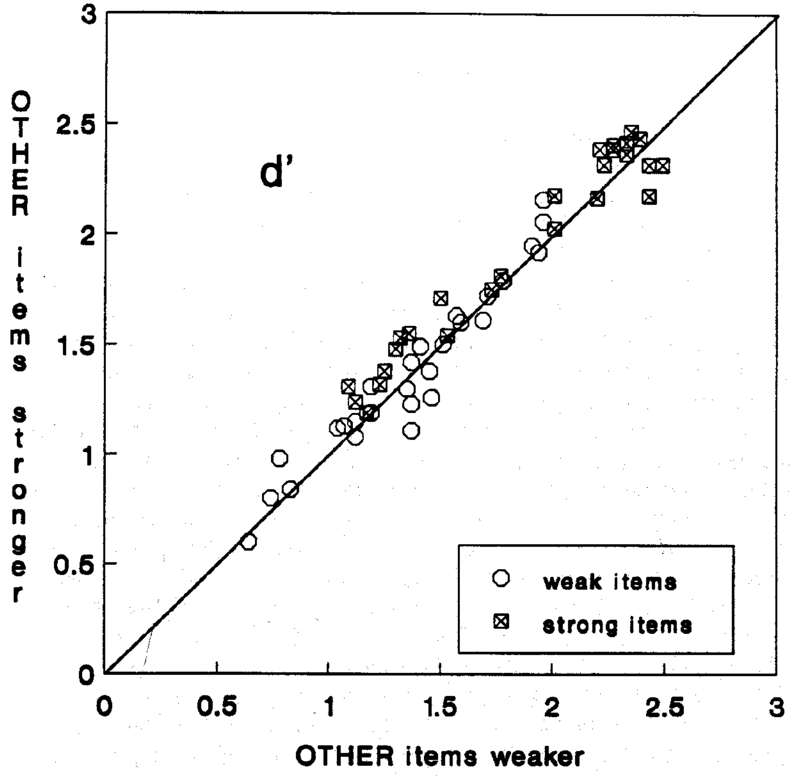

FIG. 3.1. Each point graphed represents a pair of d' values, one from each of two conditions that are matched for presentation time and number of presentations, total number of different items on the list, serial presentation position, test position, and lag from study to test, but differ in the strength of *other* items on the list. Data from many conditions in several studies is accumulated in this figure, and strong and weak test items are indicated separately. A positive list-strength effect is represented by points below the diagonal.

cantly below, but several were significantly above (although of small magnitude).

Before drawing conclusions it is necessary to establish that the results are not due to some artifact of experimental design or between-list strategy differences. In the course of the first seven studies, and subsequent experiments by Murnane and Shiffrin (in press-a, in press-b), we ruled out contamination by serial position effects at study or test, delay between study and test, within-list context shift effects, and redistribution of rehearsal or coding effort from strong to weak items in mixed lists.

We take the list-strength findings to provide strong prima facie evidence against the hypothesis that structural interference occurs in the process of storing multiple inputs in memory. Most composite storage models store items in such a way that the successive inputs interact and produce mutual degradation. According to these models it is assumed that many items are stored (in part) in the same units of memory, in such a way that increases in the number of items stored reduces the quality of representation of the individual items. One simple example comes from Anderson (1973); he represented an item by a vector of feature values; multiple items are stored by adding the vectors together to form a single vector. The resultant vector is correlated with the individual item vectors, but the correlations keep dropping as new items are added. Almost all current composite models have this character (e.g., Ackley, Hinton, & Sejnowski, 1985; Anderson, 1973; Metcalfe Eich, 1982; Murdock, 1982; Pike, 1984). Shiffrin, Ratcliff, and Clark (1990) showed that these models could not handle the list-strength findings; in addition, they could not find variants retaining structural interference that could be made to fit the results. Although Shiffrin et al. (1990) argued for separate rather than composite storage, this phrasing might be slightly misleading. The results really argue for models in which events are stored in noninterfering fashion. However, even models with this characteristic, like SAM, may not be able to predict the findings. For example, the traditional version of SAM could not do so. This led us to the following revision.

To simplify the exposition, suppose we are interested in recognition of a target item, A, or a distractor item, X, when an item B not rehearsed with A is varied in strength (B+ and B- will be used to refer to strong and weak items, respectively). The effect of B is produced by the activation of the image of B when A or X is tested. Suppose X is tested; then the cues used to probe memory are X and C (where C is the context cue). In the old version of SAM, the mean and variance of activation of B+ by C is greater than the activation of B-, but the activation of B+ by X is same as the activation of B-. As the net activation is produced by the product of these, the mean and variance of activation of B+ is more than of B-. The variance increase produces the predicted recognition deficit produced by stronger items.

To revise SAM, we settled upon a *differentiation* hypothesis: The activation (e.g., of B) produced by an unrelated item cue (e.g., A or X) is posited to be less when the image being activated is stronger (e.g., X activates B+ less than B-). This assumption has precedent in the literature and is also fairly plausible: As the cue and the image are in fact quite different, one can argue that the stronger is the stored image, the more salient will be the differences between it and the cue, and hence the less will be the activation. If so, the item cue and context cues will roughly counteract each other, strong and weak B's will be activated about equally, and list-strength effects near zero can be predicted.

It should be noted that the explanation of the list-strength effect within SAM requires a composite-like assumption, that repetitions of a given item are accumulated into a single, stronger image; otherwise, separate images would each be activated to a "standard" level, and their sum would reflect the number of repetitions. Murnane and Shiffrin (in press-a) induced separate storage of repeated words by placing them in the context of different sentences; as predicted, this manipulation produced a positive list-strength effect.

The differentiation assumption itself is a fairly modest modification of SAM (the residual strength parameter, d, is allowed to depend on the strength of the image), altering previously derived predictions only in small ways. The main result of the change is a magnification of the predicted effect of increasing list study time, so that we would need to assume that the increase in stored strength as study time increases would have to be highly damped.

In sum, we take the list-strength findings as evidence in favor of noninterfering rather than structurally interfering storage of events, and as evidence that interference in memory arises during retrieval rather than during storage. (Perhaps this is the answer to the question posed by Estes to the first author over 20 years ago.)

The differing results for free recall and recognition are accounted for by the model because free recall is assumed to involve some search cycles with only the context cue in the probe set. Similarly, the small effects in cued recall are predicted because both context and item cues are used in this paradigm. Thus both the findings and our theoretical account support different retrieval mechanisms for (free) recall and recognition.

The Units of Storage and Retrieval

The list strength results provide evidence for separate storage of events; what then are the units of event storage? We have been exploring this issue for several years. The difficulty lies in distinguishing a single higher order unit like a sentence from an interassociated collection of lower order units like the words making up a sentence. A related issue is the nature of unit used as the probe of memory. For example, when a sentence is presented for test, context will presumably be one of the cues used to probe memory, but what will be the other cues in the probe? All the words may be used jointly and separately in the probe, or the words may join into a single unit that is then used in the probe, or there may be successive probes each using one word at a time. Shiffrin et al. (1989) examined these model variants and tested them empirically. They presented five word sentences and examined cued recall for a word in a designated sentence position. The number of cues varied from one to four, and accuracy and response time were measured.

The results provided strong evidence for sentence units in both storage and test. As number of test cues increased, accuracy of response increased but latency of response decreased. The better match of the sentence unit cue to the stored sentence unit would cause such a result. However, separate storage of words would lead to slower latencies as the cues themselves would tend to be sampled during retrieval. In a subsequent study, the cue words were presented in scrambled positions. In this case accuracy increased with number of cues, but latency was flat. It was assumed that the scrambling of word order at test reduced the cue to image match, offsetting the advantage of extra cues, and thereby producing a flat latency function. However, once the relevant sentence image had been sampled, recovery from it depends on the number of sentence words in the cue much more than the order of those words, leading to the accuracy increase. This argument represents another extension of SAM in which sampling and recovery processes are at least somewhat disassociated.

Several other studies provided supporting evidence. In one, sentences, scrambled sentences, or nonsense sentences were studied. In response to the cue(s), the subject responded with as many of the remaining words as possible. Of the many salient results, we mention only one: When all the remaining words in a sentence were recalled correctly, the latencies for successive recalls were constant. We took this for evidence of a serial read out process from a sentence image, when the sentence image was well stored. If the sentence were stored as separate words, it should have taken increasing times to locate the remaining words as recall proceeded. These findings by no means exhaust the evidence in Shiffrin et al. (1989) for sentence level units. It would be pointed out, however, that their results also included evidence for word level storage in certain conditions when sentence encoding was difficult.

The sentence studies provide evidence for higher order sentence units, but the results may be generalized much further. That is, SAM must be construed as a flexible model in which the level of unit at storage and retrieval is determined by coding operations at these times. Shiffrin et al. (1989) outlined such an extension of SAM for recall tasks:

1. Grouping and coding operations in short-term store determine the units of storage. Often higher order units require longer and deeper encoding operations; if these do not succeed, lower order units may be stored, quite possibly with interassociations.
2. A variety of retrieval strategies may be used by subjects in different situations. When the test items are easy to combine into a higher order unit, the probe set will tend to consist of context plus this unit. When the test items are hard to combine (as when sentence words are presented for test in scrambled order), the items may be used separately and jointly at first (along with context), but may be combined into a higher order unit later in the memory search. When higher order units are even more difficult to form (as when nonsense sentences are tested) only lower order units may be used in the probe sets.

3. Once a higher order image is sampled, recovery is not generally based on the same strength that governs sampling (previous versions of SAM used the same strength for both, but did not consider the possibility of units at several levels). Overall match of cue set to image may determine sampling strength, but component strengths may determine recovery (if, e.g., the test sentence consists of words in scrambled order, sampling strength may be weak, but recovery strength may be quite high).
4. It seems likely that information is stored at several different levels simultaneously.

Evidence for higher order units may also be found in recognition tasks. Humphreys, Pike, Bain, and Tehan (1989) have argued for such units on the basis of tasks in which doubles and triples of items are studied and tested. Suppose the subject studies AB, AC, BC, and DEF. Suppose the task requires an "old" response when each of the test items is old. Recognition of DEF can be superior to ABC even when recognition of AB, AC, BC is superior to DE, DF, and EF. This suggests that recognition of DEF is not based on pairwise associations, and that DEF may be a unit (the argument is not airtight because the subject might have inverted the normal decision rule, but Gronlund, 1986, reported in Shiffrin et al., 1989, added some additional conditions and obtained quite strong evidence for higher order units).

Clark and Shiffrin (1987) studied many types of recognition tasks after study of word triples. They examined an alternative to higher order images as a basis of explaining findings not fitting the standard SAM model: *context sensitive encoding*. Suppose an item's encoding at study is different depending upon the other items with which it is studied. Suppose that the encoding of that item at test also depends upon the other test items. Then when test groups match study groups, the encodings will match better than otherwise, producing effects similar to those that would have occurred had the group of items been stored as a higher order unit. The context sensitive encoding model did fit the data well, but no attempt was made to fit the higher level unit model, and it is likely that this model would have done well also. We think both hypotheses are probably correct; certainly the weight of evidence from all the studies argues for multiple levels of units.

FUTURE DIRECTIONS

Despite the notable successes of the SAM theory to date, there are many opportunities for future developments, and large sets of findings that must be handled before the theory can be fairly assessed. These findings and issues include: (a) reaction times in all the basic paradigms (free and cued recall and recognition), including times for all the types of responses observed, correct and incorrect; (b) probabilities of responses in signal-to-respond paradigms (e.g., Dosher, 1984); (c) the way in which the subject or the memory system choose recognition criteria (including the "mirror" effect; see Glanzer & Adams, 1990); (d) the way in which memory is accessed for the purposes of categorization; (e) how information should be represented, and similarity among items taken into account; (f) how information is stored in and retrieved from semantic memory (can an accumulation of episodic images appear to act as a semantic image?); (g) how SAM may be best applied to the data concerning implicit and explicit memory tasks.

At the present stage of development of the field, the SAM theory is certainly one of the best worked out and most self-consistent approaches to the basic phenomena of memory. However, there are numerous strong theoretical competitors, including those couched in distributed, composite form. The field continues to evolve rapidly, and exciting prospects lie on the near horizon.

ACKNOWLEDGMENTS

This research was supported by NIMH Grant 12717 and AFOSR grant 870089 to the first author.

REFERENCES

Ackley, D. H., Hinton, G. E., & Sejnowski, J. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9,* 147–169.

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review, 80,* 417–438.

Atkinson, R. C., & Shiffrin, R. M. (1968). 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* (pp. 89–195). New York: Academic Press.

Bjork, R. A. (1988). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Flowerree symposium on cognition,* (pp. 313–344). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, S. E., & Shiffrin, R. M. (1987). Recognition of multiple-item probes. *Memory and Cognition, 15,* 367–378.

Dosher, B. A. (1984). Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 541–574.

Eich, J. Metcalfe. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

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

Estes, W. K. (1955b). Statistical theory of distributional phenomena in learning. *Psychological Review, 62,* 369–377.

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

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15,* 1–16.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66,* 325–331.

Gronlund, S. D. (1986). *Multi-level storage and retrieval: An empirical investigation and theoretical analysis.* Unpublished doctoral dissertation. Indiana University.

Gronlund, S. D., & Shiffrin, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 550–561.

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, MINERVA II, Matrix, and TODAM models. *Journal of Mathematical Psychology, 33,* 36–67.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1989). A model of contextual fluctuation. *Journal of Mathematical Psychology, 33,* 172–186.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murnane, K., & Shiffrin, R. M. (in press-a). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Murnane, K., & Shiffrin, R. M. (in press-b). Word repetitions in sentence recognition. *Memory & Cognition.*

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

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19–40.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory (Vol. 14,* 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R., Clark, S., & Shiffrin, R. M. (1990). The list-strength effect. I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Ross, B. H., & Landauer, T. K. (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior, 17,* 669–680.

Rumelhart, D. E. (1967). *The effects of interpresentation intervals on performance in a continuous paired-associate task.* (Tech. Rep. No. 16). Stanford, CA: Stanford University, *Institute for Mathematical Studies in Social Sciences.*

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

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 79,* 179–193.

Shiffrin, R. M., Murnane, K., Gronlund, S., & Roth, M. (1989). On units of storage and retrieval. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree symposium on cognition* (pp. 25–68). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shiffrin, R. M., Ratcliff, R., & Clark, S. (1990). The list-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 384–397.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460–471.

Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology, 92,* 297–304.

Young, J. L. (1971). Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology, 8,* 58–81.