Journal of Experimental Psychology:
Learning, Memory, and Cognition
1990, Vol. 16, No. 2, 179–195

0278-7393/90/$00.75

List-Strength Effect: II. Theoretical Mechanisms

Richard M. Shiffrin
Indiana University

Roger Ratcliff
Northwestern University

Steven E. Clark
University of California, Riverside

Ratcliff, Clark, and Shiffrin (1990) examined the *list-strength* effect: the effect of strengthening (or weakening) some list items upon memory for other list items. The list-strength effect was missing or negative in recognition, missing or positive in cued recall, and large and positive in free recall. We show that a large number of current models fail to predict these findings. A variant of the SAM model of Gillund and Shiffrin (1984), involving a differentiation hypothesis, can handle the data. A variant of MINERVA 2 (Hintzman, 1986, 1988) comes close but has some problems. Successful variants of a variety of composite and network models were not found (e.g., Ackley, Hinton, & Sejnowski, 1985; Anderson, 1972, 1973; Metcalfe Eich, 1982; Murdock, 1982; Pike, 1984). The results suggest constraints on the future development of such models.

Ratcliff, Clark, and Shiffrin (1990) examined evidence for a *list-strength effect*: They defined a (positive) list-strength effect as a decrease in retrieval of a given set of list items when *other* items on the list are strengthened or as an increase in retrieval of a given set of list items when *other* items are weakened. Manipulating strength by varying either presentation time or number of repetitions, they found a strong list-strength effect in free recall, at most a weak list-strength effect in cued recall, and a missing or negative list-strength effect in recognition. Such results may be contrasted with the *list-length effect* that is routinely found in all three paradigms, in which retrieval decreases as the list-length increases. It should also be noted the failure to obtain a list-strength effect in recognition could not easily be attributed to rehearsal artifacts involving redistribution of rehearsal from strong to weak items in lists mixing items of two different strengths. Evidence was amassed against the hypothesis that rehearsal was redistributed to items from temporally adjacent study items. Also many attempts were made to control rehearsal strategies, including use of an incidental learning condition, without changing the results. The possibility remains that items in lists of different strength compositions are given study effort determined by the strength composition of the list as well as the nominal strength of the item. However, because such a generalized effort hypothesis has no direct evidentiary support, is not well specified, and is not a feature of any current recognition models, we shall make a provisional assumption throughout this article that rehearsal or coding redistribution in mixed lists, or general shifts in effort across lists of different types, is not an important factor.

This research was supported by National Institute of Mental Health (NIMH) Grant 12717 and Air Force Office of Scientific Research Grant 870089 to Richard M. Shiffrin, and National Science Foundation Grant BNS 8510361 and NIMH Grant 44640 to Roger Ratcliff.

Correspondence concerning this article should be addressed to Richard M. Shiffrin, Psychology Department, Indiana University, Bloomington, Indiana 47405.

In this article we consider the theoretical implications of the results, exploring such questions as the following: (a) Can current models predict the findings? (b) What general constraints on memory models are imposed by the results? (c) Can current models be modified to handle the results?

The basic finding to be explicated is based on what was termed the "mixed-pure" paradigm in Ratcliff et al. (1990): Three list types are studied, each containing $2N$ distinct, different words (ignoring repetitions). The pure weak list consists of $2N$ items presented for brief periods of time, t (or for few repetitions, n). The pure strong list consists of $2N$ items presented for long periods of time, Zt (or for many repetitions, Zn), $Z > 1$. The mixed list consists of N items of time t (or repetitions n), and N items of time Zt (or repetitions Zn). Let the memory performance be denoted as follows: In the pure–weak condition, M(pw); in the pure–strong condition, M(ps); weak items in the mixed list, M(mw); strong items in the mixed list, M(ms). The models must predict the ratios M(ps)/M(pw) and M(ms)/M(mw), as well as the ratio of ratios: $[M(\text{ms})/M(\text{mw})]/[M(\text{ps})/M(\text{pw})] = R_r$. R_r should be greater than 1.0 if a list-strength effect exists. In our studies R_r was much greater than 1.0 for free recall, slightly greater than 1.0 for cued recall, and equal to or less than 1.0 for recognition.

In the sections to follow, we examine the list-strength predictions for several current models and consider how various modifications might bring the predictions in line with the data. First, it is useful to describe in general terms some of the factors that are most often crucial to the list-strength predictions.

Many of the current models describe recognition and sometimes recall in terms of the match between a cue used at the time of retrieval and the contents of memory. This cue-to-memory match has some mean strength E(M) and a variance, Var(M), due to noise in the system.

In recognition, the match is often termed familiarity (F), and it is assumed that an "old" response is given if $F > Cr$ and a "new" response is given otherwise, where Cr is a criterion chosen by the subject. Let $(F|i)$ denote the familiarity when a target (list item i) is tested and $(F|x)$ the familiarity when a distractor is tested. Then recognition performance is usually given by d':

$$d' = \frac{E(F|i) - E(F|x)}{\{\mathrm{Var}\ [F|x]\}^{1/2}}. \tag{1}$$

Let the recognition performance measures in the mixed–pure paradigm be denoted $d'(\mathrm{ps})$, $d'(\mathrm{pw})$, $d'(\mathrm{ms})$, and $d'(\mathrm{mw})$. Let the numerators of Equation 1 in these cases be denoted $u(\mathrm{ps})$, $u(\mathrm{pw})$, $u(\mathrm{ms})$, and $u(\mathrm{mw})$, and let the denominators of Equation 1 be denoted $\sigma(\mathrm{ps})$, $\sigma(\mathrm{pw})$, $\sigma(\mathrm{ms})$, and $\sigma(\mathrm{mw})$.

For many models the following relation holds: $u(\mathrm{ps}) = u(\mathrm{ms}) > u(\mathrm{pw}) = u(\mathrm{mw})$. That is, the difference in mean familiarity between targets and distractors depends only on the target strength, not on the strength of other items in the list. Also for virtually any model, we have $\sigma(\mathrm{ms}) = \sigma(\mathrm{mw})$, because in both cases a distractor is tested, and the list is the same (i.e., mixed). Because

$$d'(\mathrm{ps})/d'(\mathrm{pw}) = [u(\mathrm{ps})/\sigma(\mathrm{ps})]/[u(\mathrm{pw})/\sigma(\mathrm{pw})],$$

and

$$d'(\mathrm{ms})/d'(\mathrm{mw}) = [u(\mathrm{ms})/\sigma(\mathrm{ms})]/[u(\mathrm{mw})/\sigma(\mathrm{mw})],$$

we get the list-strength effect ratio:

$$R_r = [d'(\mathrm{ms})/d'(\mathrm{mw})]/[d'(\mathrm{ps})/d'(\mathrm{pw})] = \sigma(\mathrm{ps})/\sigma(\mathrm{pw}). \tag{2}$$

That is, if the pure–strong list variance is greater than the pure–weak list variance, a (positive) list-strength effect is predicted. (A *negative* list-strength effect will be the term used to denote the case when $R_r < 1$; the case $R_r = 1$ will be described as an absence of a list-strength effect.)

In addition, in many models, the pure list variance can be decomposed into a sum of equal independent components, each component corresponding to one of the presented items, denoted Var (i) for the ith item presented on a list. Let i_s represent an item on a pure–strong list, and i_w represent an item on a pure–weak list. Then,

$$R_r = \sigma(\mathrm{ps})/\sigma(\mathrm{pw}) = \{N\ \mathrm{Var}(i_s)\}^{1/2}/\{N\ \mathrm{Var}(i_w)\}^{1/2}$$
$$= \{\mathrm{Var}(i_s)/\mathrm{Var}(i_w)\}^{1/2}. \tag{3}$$

Thus a positive list-strength effect will be predicted by such models if the variance component associated with a strong item is larger than that associated with a weak item.

Virtually all the models of current interest predict $\sigma(\mathrm{ps})$ to be greater than $\sigma(\mathrm{pw})$ and hence fail to predict the data. The search for model variants capable of predicting the results will therefore focus on ways to equate the two terms in Equation 2 (or equivalently in many cases, Equation 3). Of course, such a model variant must still predict a small list-strength effect for cued recall and a large list-strength effect for free recall. It must also predict list-length effects for recognition and free and cued recall.

We begin the detailed consideration of models with the SAM model of Gillund and Shiffrin (1984) because variants have been found that can predict the data. The remaining models are taken up in an order that allows the exposition to proceed smoothly.

The SAM Model

The SAM model posits search and sampling to underlie free or cued recall and summed activation to underlie recognition. A list-strength effect is predicted for both paradigms, although for different reasons. The reasons are analogous to those that underlie the *list-length* predictions covered in Gillund and Shiffrin (1984). A simplified exposition may make this clear.

When memory is tested for old–new recognition with a single item, a memory probe is constructed with two cues, a *context cue*, C, and *an item cue*, I, each given a retrieval weight, W_c and W_r, respectively. Retrieval has limited capacity, so $W_c + W_r = 1.0$ (see Gronlund & Shiffrin, 1986). Each item (word) presented produces an image (I_j) in memory, and each of these images (from the list being tested) is activated by the probe. The activation of I_j by cues C and I_i is

$$A(\mathrm{I}_j/\mathrm{I}_i,\mathrm{C}) = S(\mathrm{I}_i, \mathrm{I}_j)^{W_i} S(\mathrm{C}, \mathrm{I}_j)^{W_c},$$

where S(a,b) represents the retrieval strength between Cue a and Image b. To make a recognition judgment, the subject simply adds the activation of all the list images; let the sum be termed familiarity (F). Then an "old" response is made if F is greater than a criterion, and "new" if not. We have

$$F(\mathrm{I}_i, \mathrm{C}) = \sum_{j=1}^{N} A(\mathrm{I}_j|\mathrm{I}_i, \mathrm{C}) \tag{4a}$$

$$F(\mathrm{I}_x, \mathrm{C}) = \sum_{j=1}^{N} A(\mathrm{I}_j|\mathrm{I}_x, \mathrm{C}), \tag{4b}$$

and if $F < \mathrm{C_R}$, respond new, but if $F > \mathrm{C_R}$, respond old, where I_x refers to a distractor item (not from the list) and $\mathrm{C_R}$ is the criterion selected by the subject. It is assumed in SAM that the activations of different images by a given cue set (i.e., the terms in the sums in Equations 4a and 4b) are independent.

The strength of an item cue to its own image, $S(\mathrm{I}_i, \mathrm{I}_i)$, is termed the *self-strength*, and its mean is assumed to equal a base value, d, plus a parameter c times an increasing function of the study time, $f(t)$. Thus $E[S(\mathrm{I}_i, \mathrm{I}_i)] = cf(t) + d$. The mean strength of an item cue to an image of an item with which the cue had been rehearsed is termed the *interitem strength* and is similarly a function of the rehearsal time: $E[S(\mathrm{I}_j, \mathrm{I}_i)] = bf(t) + d$. The mean strength between the context cue and an image is termed the *context strength* and is also assumed to rise with rehearsal time: $E[S(\mathrm{C},\mathrm{I}_i)] = af(t)$. The mean strength of an item cue to the image of an item with which it had not been rehearsed is termed the *residual strength*, d.

The distribution of the strength values is assumed to have a standard deviation that is linearly related to the mean strength value. In Gillund and Shiffrin (1984) a particular assumption was made: If the mean strength value was X, the distribution was

$$\begin{array}{rl} 0.5X & p = 1/3 \\ X & p = 1/3 \\ 1.5X & p = 1/3. \end{array} \tag{5}$$

This distribution is generalized in Appendix A, but suffices to illustrate the list-strength and list-length predictions.

In general, targets are better recognized than distractors because the mean target familiarity (Equation 4a) has one or more activations higher than residual values (e.g., $cf(t) + d > d$), whereas the mean distractor familiarity (Equation 4b) is based solely on residual values. Thus the numerator of Equation 1 will be positive, and $d' > 0$.

If items are added to a list, they add an equal amount of mean familiarity to both terms in the numerator of Equation 1, so the difference is unchanged. However, the denominator increases with extra list items: To be precise, the variance increases linearly with the number of items because the variance of a sum of independent variables (Equation 4) is the sum of the variances. Thus d' drops as the square root of the list length.[1]

If an image is increased in strength through extra presentation time and if some other item is tested that was not rehearsed with it, then again both target and distractor familiarity will increase by the same mean amount (due to increases in context strength in both cases). Again the variance will increase, by assumption, as indicated in Equation 5. Thus a list-strength effect will be predicted.

When an item is repeated, it can be treated as producing a new image, in which case it will be as if the list length was increased, or treated as increasing the strength of one image, in which case it will be as if presentation time was increased. In either case, a list-strength effect will be predicted.

Thus we see that the SAM model makes the qualitatively incorrect prediction of a list-strength effect in recognition. To examine quantitative issues and to explore model variants, it is helpful to derive analytical predictions.

To derive predictions, both for SAM and other models, assume the following "typical" paradigm: N different items are arranged into $N/2$ pairs and studied. The weak items are studied for t s, the strong for Zt s. The only interitem strengths due to rehearsal occur within pairs. Assume all lists have N distinct items, with the mixed list having M strong and $N - M$ weak items.

First note that the independence of activations for different images allows us to assess list strength in the form of Equation 3:

$$\mathrm{R_r} = \{\mathrm{Var}\,[A(\mathrm{I_s}|\mathrm{I}_x, \mathrm{C})]/\mathrm{Var}\,[A(\mathrm{I_w}|\mathrm{I}_x, \mathrm{C})]\}^{1/2}, \tag{6}$$

where I_s and I_w refer to strong and weak images, respectively.

In Appendix A it is shown that Equation 6 takes the form

$$R_r = \left\{\frac{\mathrm{E}[S(\mathrm{I}_x, \mathrm{I}_s)]}{\mathrm{E}[S(\mathrm{I}_x, \mathrm{I}_w)]}\right\}^{W_I} \left\{\frac{\mathrm{E}[S(\mathrm{C},\mathrm{I}_s)]}{\mathrm{E}[S(\mathrm{C},\mathrm{I}_w)]}\right\}^{W_C} \tag{7}$$

if $S(\mathrm{I_x},\mathrm{I})$ is independent of $S(\mathrm{C},\mathrm{I})$. Letting $\mathrm{E}[S(\mathrm{I_x}, \mathrm{I_s})] = \beta\, \mathrm{E}[S(\mathrm{I_x}, \mathrm{I_w})]$ and $\mathrm{E}[S(\mathrm{C}, \mathrm{I_s})] = \alpha\, \mathrm{E}[S(\mathrm{C}, \mathrm{I_w})]$, we can rewrite Equation 7 as

$$R_r = \alpha^{W_C}\beta^{W_I}. \tag{8}$$

Because in the previous version of SAM, β was assumed to be 1.0, we get

$$R_r = \alpha^{W_C}. \tag{9}$$

In words, if the strong items are a factor α stronger than the weak items (due to extra study time in the present analysis), then the list-strength effect ratio should be α^{W_c}, where W_c is the weight given to context.

If strength is increased by repetitions, rather than time, two models suggest themselves. In one, termed the *single stronger image* model, the repetitions are all accumulated into a single image of strength α stronger than the strength of the image for some fewer number of repetitions. This, of course, produces a situation identical to that of time variation, and $R_r = \alpha^{W_c}$. In the other, termed the *multiple image* model, repetitions are all represented by separate images. In this case, a list-strength effect occurs because the repetitions in effect increase the list *length*: $u(\mathrm{ps}) = u(\mathrm{ms})$ and $u(\mathrm{pw}) = u(\mathrm{mw})$, as before, but $\sigma(\mathrm{ps}) > \sigma(\mathrm{ms}) = \sigma(\mathrm{mw}) > \sigma(\mathrm{pw})$ because the number of images is ordered in this way, and each image adds a variance component that is greater than zero.

To be more specific about the multiple-image model, Equation 2 holds because of the equality of the us for pure and mixed lists. If different items are independent, then a version of Equation 4 holds as well:

$$R_r = \{\mathrm{Var}[\sum_{j=1}^{r_s} A(\mathrm{I}_{ij}|\mathrm{I}_x, \mathrm{C})]/\mathrm{Var}[\sum_{j=1}^{r_w} A(\mathrm{I}_{ij}|\mathrm{I}_x, \mathrm{C})]\}^{1/2}, \tag{10}$$

where r_s and r_w are the number of strong and weak repetitions for Item I_i, and subscript j denotes the different images for the repetition of I_i. If all the repetitions are equal and independent, then Equation 10 becomes

$$R_r = (r_s/r_w)^{1/2}. \tag{11}$$

If repetitions are not equal in expectation, the picture is more complex. Suppose $\delta_j = \mathrm{E}[S(\mathrm{C},\mathrm{I}_{ij})]$ is the mean context strength for the jth repetition for item i. Then assuming independence of the repetitions (dependence will make the variance higher, so this should be a conservative assumption), it is shown in Appendix A that

$$R_r = \left\{\sum_{j=1}^{r_s} \delta_j^{2W_C} \Big/ \sum_{j=1}^{r_w} \delta_j^{2W_I}\right\}^{1/2}. \tag{12}$$

If the δ_j are constant, or if $\mathrm{W_C}$ equals zero, then Equation 12 reduces to Equation 11. Regardless, as long as $\delta_j > 0$, R_r will be greater than 1.0, and a list-strength effect will be predicted.

Discussion and Model Variations

The independent multiple-image model for repetitions predicts a list-strength effect, as in Equation 12. Even if repetitions are not independent, Equation 10 is likely to be greater than 1.0. It is possible to envision situations in which greater strength stored on one repetition is compensated for by lesser strength stored on the next repetition, in such a way that variance does not rise, or even goes down, with repetitions (e.g., see the "closed loop hypothesis" of Murdock & Lamon,

[1] Of course, if the context cue does not perfectly focus search on the images of items in the most recent list, then the denominator will contain an extra variance component (possibly independent of list length), and the square root relation would no longer apply.

1988). However, this seems to require knowledge of the current number of repetitions, and this knowledge is not explicitly available to subjects in the experimental situations to which the model must be applied. Also, certain technical problems arise if the variance is not to increase. We will focus therefore on the single-stronger-image model for repetitions. The assumption that repetitions are accumulated into a single, stronger image equates theoretically the time and repetition versions of the mixed-pure paradigm, so only one model need be considered.[2]

There are a number of variants of SAM, some quite simple, that can predict R_r to be 1.0, but all but one are flawed in one or more ways. Our favored variant, the *differentiation* model, is discussed next. The others are discussed in Appendix B.

A Differentiation Model

In this variant of SAM, we relax the assumption that d, the mean item-to-item residual strength, is a constant. Even in Gillund and Shiffrin (1984), d was allowed to vary with such factors as natural language word frequency, and it was stated that d ought to be affected by factors like similarity (e.g., Gillund & Shiffrin, 1984, p. 32), so that allowing d to vary with storage strength seems a natural extension, well within the logic of the approach. Suppose that d is lower when the image being activated is stronger (had been rehearsed more, say). This assumption is based on a *differentiation* argument: The better encoded is an image, the more clear are the differences between it and the test item. The argument based on differentiation is an old one (e.g., Gibson, 1940; Salz, 1961, 1963) and is also common in the literature on similarity (e.g., Gibson & Gibson, 1955; Nosofsky, 1987). Indeed the basis for differentiation can be couched in terms of similarity: When a "clear" test stimulus is used to activate a strongly stored quite different memory image, the similarity between the two should be low, reflected in a low activation strength. As the storage strength drops, the uncertainty about features stored can cause an increase in similarity and activation strength. Of course, as storage strength continues to drop toward zero, eventually so must activation strength. Thus, technically, the function relating storage strength to activation strength for an "unrelated" test item must be nonmonotonic, starting at zero, rising fairly quickly to a maximum and then falling, perhaps toward an asymptotic value. For the studies reported in Part I, it will be assumed that the storage strengths will always be large enough that the activation strengths will be on the downward sloping part of the function.

The idea behind the differentiation model is illustrated in Figure 1. The left-hand panels show the mean retrieval strength $\bar{S}(C,I)$, between the context cue and an image, as a function of the presentation time (or number of presentations) of the item encoded in that image. The middle panels show the mean retrieval strength $\bar{S}(I_x,I)$ between the test item, I_x, and the image, also as a function of the presentation time of the item encoded in the image (when the test item and item in the activated image had not been encoded together). The right panels show mean total activation by both cues, assuming the cues are given the weights shown. Note also that the standard deviation of activation for this model is just a constant times the mean activation, as was true of the original model and as shown in Appendix A.

Row 1 shows how a trade-off of increasing context strength and decreasing item (residual) strength can produce a nearly flat mean and variance of activation as a function of item strength. For items having strengths on the flat portion of this function, no list-strength effect would be predicted. Row 2 has the same context and item strength functions as Row 1 but shows that an increase in the context weight (and decrease in item weight) can cause the mean and variance of activation to rise with strength, producing a positive list-strength effect (a movement of the weights in the other direction could, of course, produce a negative list-strength effect). Row 3 keeps the weights equal to .5 but shows that different item strength functions can produce either positive or negative list-strength effects.

The cases shown in the figure illustrate the flexibility of the differentiation model. Slight differences in the shapes of the strength functions, or in the weights given to the cues, could produce slight decreases or increases in the mean and variance of activation as item strength increases, thereby allowing the model to account for a range of list-strength effects. The shapes of the functions are presumably a function of the item types and similarities, whereas the weights are chosen by the subject.

Explicit predictions for the mixed-pure paradigm are not hard to generate. Assume that the mean item and context strengths are functions of rehearsal time for the item in the activated image when the test item had not been rehearsed with the activated item (along the lines of the functions illustrated in Figure 1). Assume that the noises associated with these mean strengths are otherwise independent. Then Equations 7 and 8 hold. Next assume that $\alpha > 1$ and $\beta < 1$ for the levels of strength used in a given experiment, thus producing the kind of trade-off exhibited in Figure 1, in the right-hand portion of the panels. Then for fixed α and β, raising the value of W_C and/or lowering the value of W_I will increase the list-strength effect.[3]

This differentiation model has a number of interesting features. Because both the mean and variance of activation remain roughly constant as strength of storage increases, when context and an unrelated item are used as cues, then a recall probe using these cues will not be affected by strength of

[2] It should be noted that there exists a rather long and inconclusive literature concerning the appropriate representation for repeated items (e.g., see Hintzman, 1988, for one recent treatment). We do not intend to review this evidence in this article. The choice of the single-image representation is made here because, of the two choices, it is the only one offering a ready explanation of the list-strength results.

[3] Our particular assumptions concerning rehearsal time for an item, negative dependence of mean strengths, and independence of noise components were chosen in part because they give rise to the pleasingly simple result of Equation 5. Many other assumptions could have been made that do not qualitatively change the arguments we have made but that are mathematically more complex.

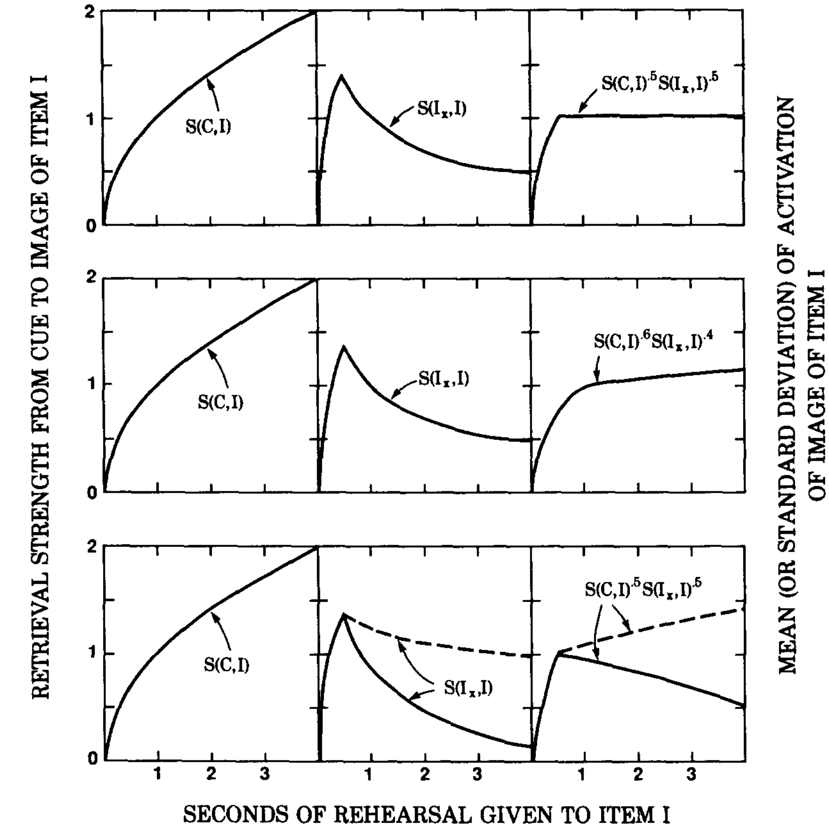

Figure 1. An illustration of the way in which activation of an image (right-hand panels) is determined by the retrieval strengths due to the context and (extraneous) item cues. (In a *differentiation* version of the SAM model, the residual item strength [the center panels in each row] decreases as the storage strength of the activated item increases, counteracting the effect of the context cue.) The two left-hand columns give mean strengths, and the right-hand column gives mean activation (which differs only by a constant factor from the standard deviation of activation—see Equation A2). (A positive list-strength effect is predicted if the items with higher amounts of rehearsal have a higher level of activation [right-hand panel] and hence higher variance. The type of list-strength effect predicted depends upon the cue weightings [first compared with second row] and upon the shape of the strength functions [first compared with third row].)

storage of other items. In the SAM model for free recall, it is assumed that some probes of memory are made with the context cue only, and other probes by an item and the context cue together. When the context cue only is utilized, a strong list-strength effect will be predicted (because the "stronger" images will be selectively sampled in mixed-strength lists). Thus, overall, a list-strength effect will be predicted in free recall to the extent that context cuing occurs alone during the course of retrieval. However, in cued recall (and in free recall whenever context and item cues are used jointly), the situation will be quite different: Because all (non-jointly-rehearsed) images will be activated about equally, regardless of strength, no selective sampling of stronger items will occur. Thus the main factor leading to a list-strength effect will be eliminated.

One way to predict a modest list-strength effect in cued recall, nonetheless, involves variations of the weights assigned to cues. Suppose that for cued recall a higher weight is assigned to context, and a lower weight assigned to the item cue than would be true for recognition. Reference to Equation 7 makes it clear that these expressions represent not only R_r but also the ratios of mean activation strength for a strong image to mean activation strength for a weak image. Therefore anything raising R_r, such as a weighting shift, will also increase the mean activation of strong items relative to weak items and hence lead to increased sampling of strong items from mixed lists. The data from Ratcliff et al. (1990) are consistent with the hypothesis that a slightly greater list-strength effect occurs in cued recall than in recognition (although additional data are needed).

Aside from weighting shifts, another factor in SAM may produce a list-strength effect in cued recall when none is seen in recognition. In SAM, after an image is sampled in the recall search process, the probability of correctly recalling that item is *not* based on the activation strength, but rather on the probabilistic combination of the recovery probabilities due to the separate cue strengths (see Gillund & Shiffrin, 1984). To be precise,

$$P_R(I_j|I_i, C) = 1 - \exp\{-W_I S(I_i, I_j) - W_C S(C, I_j)\}. \quad (13)$$

Because $S(C,I_j)$ will generally be higher than $S(I_i,I_j)$ (i.e., the context strength will be higher than the residual), Equation 13 will be higher for images with higher context strengths, even when the total activation of images with different context strengths is equal. Now suppose that in cued recall, recoveries of incorrect images will lead to a strengthening of item connections between the context and item cues used in the probe and to the image of the incorrect item recovered (Gillund & Shiffrin, 1984, termed this *incrementing*). Because stronger items will be selectively favored in recovery, this strengthening will occur more for stronger items than weaker ones. The effect of strengthening will be to increase the activation of that image to those cues from that point onward in the memory search, thereby reducing the probabilities of sampling the yet to be recovered correct item. Overall, then, stronger other items tend to reduce cued recall, and weaker other items tend to increase cued recall, over the entire course of the recall period. (This same factor could apply in free recall when item cues are used, but the continual switching of item cues in free recall will tend to minimize its contribution.)

In summary, the new variant of SAM predicts a good sized list-strength effect in free recall because of search phases in which context cues only are used. It predicts a somewhat greater list-strength effect in cued recall than in recognition if an incrementing process takes place during the memory search in cued recall for irrelevant items that are recovered. The advantage of the list-strength effect in cued recall over recognition will be further enhanced if the weight given to the item cue is higher, and the weight given to the context cue is lower, in recognition compared with cued recall.

It must next be asked whether the new variant of SAM will alter the previous predictions made by the theory, such as those put forward in Gillund and Shiffrin (1984) and Raaijmakers and Shiffrin (1980, 1981). The majority of such predictions occurred in situations where strength was not explicitly varied, and hence we will see only very slight quantitative changes in predictions (due, say, to slightly differing results when items are given differing amounts of rehearsal in a buffer). In cases where strength is varied, such as variations in presentation rate for a list, larger quantitative changes in predictions occur. Longer rehearsal times per item will produce two effects: (a) a direct strengthening of the test item's image (as in the old version of SAM) and (b) no strengthening of the activations of images not rehearsed with the test item (in contrast to the old version of SAM). Factor 2 will mean that the new variant of SAM will predict a larger effect of increased rehearsal time than the previous version (for both recall and recognition). As a consequence, it will probably be necessary to assume that the function relating growth in mean strength to rehearsal time will bend over more sharply than would have been necessary in the previous version of SAM. Clearly, quantitative testing of SAM with the differentiation assumption added must be carried out, both on the theoretical and empirical levels, but the topic will not be pursued further in this article.

Predictions of Other Models for the List-Strength Effect

None of the other models we have looked at predict the failure to obtain a list-strength effect in recognition. Variants of these models capable of handling the findings are difficult to find, though one based upon the MINERVA 2 model by Hintzman (1986, 1988) has fewer problems than the others. We begin, therefore, with MINERVA 2.

The MINERVA 2 Model of Hintzman (1986)

In MINERVA 2 items are vectors of feature values (+1, 0, or −1), and a pair of items is a longer vector consisting of two item vectors end-to-end. Each item or pair is stored separately. Each feature value is stored with probability L, and with probability $1 - L$, zero is stored for that feature. At test, the test vector is compared in parallel with each memory item. In particular, the dot product between the test vector and each memory vector is taken, normalized in a manner described below, and cubed. The result is the activation value for that image. The sum of all activation values is a measure of familiarity (F) used to make a recognition decision:

$$F = \sum_{i=1}^{N} \{(1/N_{R,i}) \sum_{j=1}^{M} P_j T_{i,j}\}^3. \quad (14)$$

In Equation 14, N is the number of vectors stored, M is the number of features in each vector, P_j is the value of feature j in the test probe, $T_{i,j}$ is the value of feature j in vector i, and $N_{R,i}$, the normalizing factor, is the number of features in the ith trace for which either P_j or $T_{i,j}$ is nonzero.

Applications to the mixed–pure paradigm are straightforward. If an item is given more study time, the value of L is higher; spaced repetitions result in the storage of new traces. It is easy to see that Equation 2 holds for this model because u(ms) = u(ps) and u(mw) = u(pw). The us are equal because for each trace activated by a target there is an equivalent mean activation of that trace by a distractor, save only for the trace matching the target. Thus the composition of the list, except for the target trace, has no effect on the numerator of d'. Because Equation 2 holds, we need only to determine the variances for pure strong and pure weak lists to derive list-strength predictions.

Ignoring for the moment the case of increased study time, the prediction for repetitions is clear: Additional presentations increase the number of vectors stored in memory. Each such vector contributes a positive variance term to the total activation, so additional presentations produce additional variance and hence lead to the prediction of a positive list-strength effect. Because this prediction is in error, the remaining discussion will focus on variants of MINERVA 2 in which all repetitions of a given item are collapsed into a single stored vector. We will assume that this is done in such a way that increases in presentations become analogous to increases in presentation time. That is, to start with, assume that extra presentations act to produce a stored vector equivalent to what would have been stored in one presentation with a larger value of L.

For increased presentation time, Equation 3 holds as well, in the following form:

$$R_r = \{\mathrm{Var}_s\left[\sum_{i=1}^{N} \{(1/N_{R,i}) \sum_{j=1}^{M} P_j T_{i,j}\}^3\right] / $$
$$\mathrm{Var}_w\left[\sum_{i=1}^{N} \{(1/N_{R,i}) \sum_{j=1}^{M} P_j T_{i,j}\}^3\right]\}^{1/2}, \quad (15)$$

where the subscripts *s* and *w* refer to storage of a *s*trong trace T_i, or a *w*eak trace T_i.

Unfortunately, it is not very easy to determine analytically the magnitude of the expression in Equation 15. In the limiting case when all elements in the presentation or test vectors are nonzero, the normalizing term, $N_{R,i}$ will be a constant equal to M, regardless of the value of L (i.e., the storage strength). In this case it is not hard to show that the variance of activation will increase as L increases, and a positive list-strength effect will be predicted. However, when vectors have a (random) number of zero entries, then $N_{R,i}$ will be a random variable whose mean will increase with increases in L and whose effect will be to reduce the variance as L increases, counteracting to some degree the other factors that are operating. We have decided not to investigate this model in greater depth for two reasons. First, the particular choice of normalizing factor made by Hintzman does not seem essential to the theory and was probably made somewhat arbitrarily. It would be appropriate to investigate a range of models for the normalizing factor (e.g., $N_{R,i}$ could be set to a constant value; $N_{R,i}$ could count the number of features for which the product $P_jT_{i,j}$ is nonzero; $N_{R,i}$ could be a power function of some feature count, etc.). Second, even though one of these models might have the desired property that variance remains constant when L changes, we prefer to discuss instead a much simpler variant having the same property. We take up this variant next.

The key to the next approach involves a slight alteration of the learning assumptions. Suppose the number of nonzero entries in the stored vector does not vary with the learning parameter, L. Suppose instead that a feature value matching the presented item's feature value is stored with probability L, and with probability $1 - L$ a value is stored that is chosen at random from the distribution giving rise to the feature values in the first place. In effect, storage will consist of replacing random feature values with appropriate ones, more appropriate ones being inserted, the stronger is the item (at least if repetitions are accumulated into one stored vector). If the item that is tested is a different one than the one encoded in the stored vector, then the mean and variance of activation will be the same, regardless of the "strength" of storage (because all vector elements will be independent with respect to the test vector and because the number of zero entries in the stored vector will not change with strength). This model, therefore, predicts no list-strength effect in recognition ($R_r = 1$).

A problem for this model is the negative list-strength effects that are sometimes observed. It would be desirable to have a model with the flexibility to account for at least a small range of list-strength effects. Another problem involves recall.

Recall in MINERVA 2 is slightly more complex than recognition. Although a full model of free recall has not been worked out, cued recall is well specified. For cued recall, the probe item is used (as before) to activate each stored image. Then the activation value for a given vector multiplies all the values in that vector to produce a *weighted vector*. This is done for all stored vectors, and then all the weighted vectors are summed to produce a vector termed the *content echo*. That is, the value (i.e., activation) of feature j in the content echo is

$$C_j = \sum_{i=1}^{N} A_i T_{i,j}, \quad (16)$$

where A_i is the activation of the ith vector stored in memory (given by the term to the right of the summation in Equation 14). Several methods have been proposed to "clean up" the content echo and produce a recall, but, in general, better recall is expected if the dot product of the content echo with the correct response vector is higher.

If repetitions are stored separately, the effect of strengthening is straightforward: More extraneous values will enter into the sum in Equation 16, increasing noise, reducing performance, and leading to a prediction of a positive list-strength effect. However, if it is assumed that repetitions are accumulated into one trace for recognition, the same must be assumed for recall. Thus the case of varying presentation time is the one needing investigation. In general, the recall predictions for the list-strength effect mimic those for recognition. In particular, if the modified learning rule is assumed, it is not hard to see that the "noise" contribution to the content echo from traces other than the desired target will not be affected by the storage strength of those other items. Thus no list-strength effect will be predicted for cued-recall, a problem to the degree that a small positive effect was observed in the data.

A larger potential problem for MINERVA 2 lies in the paradigm of free recall. If it is assumed that free recall is just a concatenation of a series of cued recall operations, no list-strength effect would again be predicted. It may be necessary for this reason to append a free recall model to MINERVA 2 mimicking the sampling assumptions in SAM: If stronger items are sampled preferentially from a mixed list, a large list-strength effect will be predicted (as was true for SAM).

In summary, we have not found a variant of MINERVA 2 that can handle all our findings, but certain versions in which all repetitions are accumulated into one memory trace come closest and deserve further exploration. One problem that must be resolved involves the need to allow prediction of a range of list-strength effects in recognition (including some negative). Another problem is the similar predictions that tend to arise for cued recall and recognition. (However, if a mechanism is invented to allow a range of list-strength effects in recognition, that mechanism might be able to predict small positive effects in cued recall.) The problem is greatest for free recall. One way to solve this problem involves producing a free recall model that allows sampling in proportion to strength, as in SAM. Perhaps some part of each stored vector could be reserved for context features. When these features are used to probe memory, the way in which the content echo is "cleaned up" to produce a recall on a given recall cycle might somehow be made to produce stronger items, preferentially.

Pike's Matrix Model (1984)

Each item is represented by a vector of size n of feature values. An item (word) has features that are chosen randomly and independently from a distribution with mean μ, and variance σ^2. A pair is treated as a matrix formed from the vector product (i.e., the matrix product of the vectors treated as row and column matrices) of the two component items. The matrices representing successively presented pairs on a list are summed together, cell by cell, to form a composite memory matrix, **M**. For recognition, a test matrix is formed from the presented pair (if only one item is tested, it is multiplied by a vector of ones to form a test matrix), and the dot product between the test matrix and **M** is taken. The result, F, is used to make an old–new decision.

To apply this model to the mixed–pure paradigm, one must have a mechanism for strengthening items. In the case of spaced repetitions, the most direct approach seems to involve storing the repeated pair (matrix) again. Perhaps later presentations may not be as efficacious as earlier ones, so suppose that r repetitions produce storage in memory of a matrix which is the original matrix multiplied by a scalar α ($1 < \alpha < r$; this may be thought of as an approximation to the *closed loop* hypothesis of Murdock & Lamon, 1988). The case of extra presentation time is most simply handled similarly, by assuming that study time of rt produces storage of the matrix multiplied by α.

This simple version of the model unfortunately predicts no improvement in performance for pure lists of increasing strengths (we shall see this shortly—α increases both the mean and standard deviation in corresponding fashion, so d' does not change). Two methods suggest themselves to solve this problem: Probabilistic encoding and/or noise in the memory matrix **M** due to inputs prior to the current list.

We consider first the hypothesis that the matrix **M** already contains values at the start of list presentation. Suppose the simplicity (it does no harm to do so) that j matrices have already been accumulated in memory, with $\alpha = 1$. For pure lists, k matrices are then stored during list presentation, all with strength $\alpha = \alpha_1$ or all with strength $\alpha = \alpha_2$ (along with j prior matrices with strength $\alpha = 1$).

For this model Equation 2 holds because $u(\mathrm{ms}) = u(\mathrm{ps})$ and $u(\mathrm{mw}) = u(\mathrm{pw})$ (and $\sigma^2[\mathrm{ms}] = \sigma^2[\mathrm{mw}]$, of course). The us are equal because the mean activations by a distractor cancel all the mean activations by a target for all matrices making up **M** but for one: that one corresponding to the target. Thus the composition of the rest of the list is irrelevant. Because Equation 2 holds, we need to determine only the variances for pure-strong and pure-weak lists. Some extensive algebraic computations (help may be found in Humphreys, Pike, Bain, & Tehan, 1989), when the pure list items are stored with strength α, lead to

$$\mathrm{Var}(F \mid \alpha) = (j + \alpha^2 k)(n^2\sigma^6 + 2n^2\sigma^4\mu^2 + 2n^3\sigma^2\mu^4 + n^3\sigma^4\mu^2) + (j^2 + \alpha^2 k^2 + 2\alpha jk)\mu^4 n^3 \sigma^2, \quad (17)$$

and

$$R_r = \{\mathrm{Var}(F \mid \alpha = \alpha_1)/\mathrm{Var}(F \mid \alpha = \alpha_2)\}^{1/2}, \quad \alpha_1 > \alpha_2. \quad (18)$$

The variance obviously rises with k, so a list-length effect is predicted. To see that a list-strength effect is predicted (i.e., that R_r is greater than 1), it must merely be noted that Equation 17 is an increasing function of α. Although it is not hard to obtain the expression for d' for the different conditions, we have not carried out a quantitative fit because the predictions are qualitatively in error for the cases in which R_r is significantly less than 1.0

An alternative approach to strengthening items would involve probabilistic encoding (as in MINERVA 2), either at the level of the whole item (the probability of matrix storage would rise with extra time or repetitions) or at the level of the feature (features would be stored with a probability that rises with extra time or that would apply again with each repetition). When correct storage does not occur for a feature, it could be replaced by a random choice or by some constant value. When items are given extra study time, at least one approach can handle the absence of a list-strength effect while still having stronger items give higher performance: Replacing random features with a number of veridical features, the number rising with study time, would leave the variance unaltered, and no list-strength effect would be predicted (just as in Hintzman's MINERVA 2 model). For the present model, and other composite storage models, it does not seem possible to solve the problem in an equivalent way in the spaced repetitions case. When an item is first presented, its trace joins the composite trace. At a later repetition, there is no sensible way to extract from the composite the feature elements stored at previous presentations (so that some of the random features may be switched to correct ones prior to restorage). We have not found a way to amend the model properly in the case of spaced repetitions (the extra storage events cause the variance to rise in all cases). Because it seems clear that the predictions for spaced repetitions will be in error in all variants and because analytical derivations are cumbersome for the probabilistic storage versions of the matrix model, we have not carried out quantitative tests.

It is not clear to us how to amend the matrix model to handle the list-strength findings in recognition. Assuming positive correlations among items does not seem to help, and letting μ and/or σ vary with strength lends to bizarre complications (and still fails in the case of spaced repetitions). Even probabilistic storage combined with normalization (which could be considered in MINERVA 2 because items are stored separately) does not seem to be sensible when the stored matrix is a composite. Another potential problem involves recall. The list-strength predictions for recall, in the simplest variants, follow those for recognition because the underlying mechanisms are similar: For cued-recall an item probes memory by multiplying the memory matrix; the resultant vector is used to produce recall. Although an explicit recall model is not yet available, Humphreys, Bain, and Pike (1989) suggest, in effect, that the output vector be compared with each possible candidate for recall by taking a dot product; they suggest that recall would follow the ordering of strengths so produced. This procedure produces strengths equivalent to those that would be available in an n-alternative forced-choice recognition test, and there is every reason to believe that the list-strength predictions would mimic qualitatively those in recognition (though the effect of increasing noise should be quantitatively larger, the larger is n, so recall would produce a larger list-strength effect than recognition).

In summary, the matrix model predicts some positive list-strength effects in recognition in all variants we have considered and predicts list-length and list-strength effects to occur together. A potential additional problem lies in wait: It is not clear how the model could produce the different list-strength results found in recognition, cued recall, and free recall.

Murdock's TODAM Model (1982)

Each item is represented by a vector of N feature values, each value chosen from a distribution with mean zero and variance σ^2 (usually set to $1/N$). A pair is stored as the sum of three weighted vectors: Item One's vector $\mathbf{f}_i$ multiplied by γ_1, plus Item Two's vector $\mathbf{g}_i$ multiplied by γ_2, plus the convolution of the two vectors $\mathbf{f}_i * \mathbf{g}_i$ multiplied by γ_3. This vector is then added to the previous memory vector multiplied by a forgetting factor α:

$$\mathbf{M}_i = \alpha\mathbf{M}_{i-1} + \gamma_1\mathbf{f}_i + \gamma_2\mathbf{g}_i + \gamma_3(\mathbf{f}_i * \mathbf{g}_i). \qquad (19)$$

For single item recognition (in the simplest case), the test vector is compared with the memory vector by taking a dot product.

The arguments now closely parallel those for Pike's matrix model. For spaced repetitions, it seems most sensible to add the repetition to memory, just as for any other pair (though possibly multiplied by a scalar less than one, to allow for less rehearsal for later presentations). It seems easiest then to let extra study time be handled by assuming that the stored vector is multiplied by a scalar value, η, whose size rises with the amount of study time. Just as for the matrix model, such multiplication does not change performance in a pure list if one assumes $\mathbf{M}$ contains only the list items (we shall see this in Equation 20). Therefore, we might produce performance increases by (a) assuming prior items not from the current list to be in $\mathbf{M}$ already or by (b) assuming probabilistic storage of items or features, the probability rising with study time or the probability applying more often with extra study time (e.g., see Murdock & Lamon, 1988). These two possibilities are addressed in turn.

Prior items in memory. Suppose there are prior items stored in $\mathbf{M}$, and then j weak paris ($\eta = \eta_1$) and k strong pairs ($\eta = \eta_2$) are presented in a list, with strength manipulated via presentation time or massed repetitions. For a recognition test of a single distractor, the variance of the dot product (see Murdock, 1982) is

$$Var(F) = \left[\frac{1}{N}(\gamma_1{}^2 + \gamma_2{}^2) + \frac{3N^2+1}{4N^3}\gamma_3{}^2\right] \left[C_p + \eta_1{}^2\sum_j C_i^2 + \eta_2{}^2\sum_k C_i^2\right]. \qquad (20)$$

In Equation 20, C_p is a constant representing contributions to the variance of items already stored in memory before list presentation, and C_i represents a serial position constant whose value for serial position i in a list of p items is just $C_i = \alpha^{p-i}$; the first sum in Equation 20 is to be taken over the serial positions occupied by the j weak items, and the second sum over the serial positions occupied by the k strong items. To complete the story, note that the numerator of d' is, for the first member of a pair, say, $\eta\gamma_1C_i$ for an item studied in position i, of strength η.

Because $u(\text{ms}) = u(\text{ps}) = \eta_2\gamma_1C_i$, and $u(\text{mw}) = u(\text{pw}) = \eta_1\gamma_1C_i$, Equation 2 holds for items in comparable serial positions relative to the end of the list. Thus we need to compare only the variances for pure lists, each of size m, of two different strengths:

$$R_r = \{\text{Var}(F\,|\,j=0, k=m)/\text{Var}\,F\,|\,j=m, k=0)\}^{1/2}. \qquad (21)$$

It is obvious from Equation 20 that $\eta_2 > \eta_1$ makes the left term of Equation 21 larger, and a positive list-strength effect is predicted.

The case of spaced repetitions is more complex because of serial position effects and possible differential forgetting of prior items in memory. For a pure list, say, let δ_i be the scalar constant governing storage on the ith presentation of the same item (δ_i might be smaller for larger i because of such factors as decreasing rehearsal, etc.). Let $C_{k(i),j}$ be the serial position constant that applies to the ith presentation of item j, occurring in position k. We can then derive

$$\text{Var}[F] = \left[\frac{1}{N}(\gamma_1{}^2 + \gamma_2{}^2) + \frac{3N^2+1}{4N^3}\gamma_3{}^2\right] \left[C_p\beta^2 + \sum_j\left(\sum_i \delta_i C_{k(i),j}\right)^2\right], \qquad (22)$$

in which the sum over j is a sum over items, and the sum over i is a sum over repetitions of a given item. β represents the fate of prior items in memory. If prior items are assumed to contribute a constant amount to memory regardless of list length, session length, and so forth, then $\beta = 1$. If, near the other extreme, all presentations cause equal forgetting of prior memory contents, then $\beta = \alpha^R$, where R is the total number of presentations in a list counting repetitions separately. If $\beta = \alpha^R$, then adding items to a list or strengthening items on a list has two opposing effects: a variance increase due to the list items added or strengthened and a variance decrease due to the lessened contribution of the prior list items undergoing forgetting. Although this model deserves further exploration, we will not pursue it for several reasons: First, a trade-off resulting in no variance increase would apply both to list-length and list-strength effects, whereas the data exhibit opposing effects. Second, the explanation requires enormous recency effects, which are not seen in the data. Third, the explanation requires enormous influences of prior lists, which are not normally seen in the data. In the following, then, let us suppose $\beta = 1$.

If we again focus on items whose first presentation and repetitions (if any) occupy comparable serial positions relative to list end in pure and mixed lists, then Equation 2 holds. In this case, we must examine Equation 22 to see if extra repetitions will cause an increase in variance. Depending on the values of δ_i and $C_{k(i),j}$, extra repetitions may or may not increase the variance. However, noting that the numerator of d' (say, for the first member of a pair) is $\gamma_1 \sum_i \delta_i C_{k(i),j}$, it is not hard to see that the same factors determining the magnitude of the variance determine the magnitude of d'. Thus the variance should increase with repetitions when d' increases with repetitions. Because the d' performance level is indeed observed to be higher for pure stronger lists, then the variance ought to be higher as well, and hence a positive list-strength effect ought to be seen.

As an aside, we have not found explanations based on serial position effects to be of much help in explaining our findings, in the context of any of the models under consideration (not just TODAM). The reason is twofold: First, we did not observe large enough study or test position effects to explain our findings, especially because the lack of a list-strength effect did not depend on the blocking or spacing, or placement, of items of differing strength. Second, the models that could be made to deal with the repetitions paradigm could not be carried over to the presentation time paradigm, could not usually explain different effects in recall, and usually could not predict $R_r \leq 1$ while still predicting a main effect of strength.

Probabilistic encoding. Consider probabilistic encoding as a basis for strengthening items. If features are encoded with probability L per unit of rehearsal time, say, and if nonencoded features are stored as zeros, then it is easy to see that a positive list strength effect is predicted—the more features are encoded, the larger the variance (because, as opposed to MINERVA 2, N is fixed regardless of the number of zeros in the vector). In general, probabilistic storage of vectors as wholes will also lead to increased variance with increased strength.

More interesting is the assumption that features not stored correctly are stored instead (with probability $1 - L$) as a random sample from the original sampling distribution ($\mu = 0$, σ^2). For the case of increased time, L will increase (or be applied more often), leading to storage of more veridical features and fewer random features. However, when a *different* item is tested, the number of veridical features stored for the item in question is irrelevant because in any event the items in question will be orthogonal. Thus no list-strength effect would be predicted. Of course, negative list-strength effects would remain a problem.

The same assumption would not help in the case of spaced repetitions, because new (partial) traces are laid down, thereby increasing the variance (though the equation would have a somewhat more complex form than Equation 22). It would not make sense, once the first presentation of an item has joined the composite memory, that its exact stored features could be extracted upon a later presentation, altered in the direction of greater veridicality, and then added back to memory.

For the case of spaced repetitions, each of the above approaches (prior memory contents and probabilistic encoding) are complicated considerably when the amount stored for a repetition depends inversely on the current level of activation produced by that item. Murdock and Lamon (1988) call this a "closed loop" hypothesis. As repetitions proceed, the items tend to equalize their activation so that the target variance drops. How this should affect the list-strength effect is less clear (especially since analytical derivations for this model do not seem possible). Although we see no reason why the addition of a closed-loop hypothesis should alter the list-strength predictions of the model, intuitions are not easy in this case, and further research would be desirable.

A possible final problem for TODAM involves the similar list-strength predictions produced by the theory for recall and recognition. As recall has been characterized thus far in TODAM, it largely acts mathematically like multiple-item forced-choice recognition of pairs. Thus the model in its present form correctly predicts a list-strength effect in recall. However, if new assumptions allowed the variance for pure–strong lists and pure–weak lists to be equal, it is very likely that the same would apply in recall, and the prediction of a positive list-strength effect would disappear. At least for free recall, a solution (should the problem arise) might involve a sampling process having properties like SAM, perhaps along the lines suggested by Metcalfe and Murdock (1981).

Metcalfe's CHARM Model (Eich, 1982, 1985)

The assumptions of CHARM largely mimic those of TODAM for pairs. Most of the applications of CHARM have involved recall tasks (Metcalfe Eich, 1982, 1985), but we take up briefly one approach to recognition that differs somewhat from TODAM (Metcalfe Eich, 1985). Assume pairs (a_i,b_i) are presented for study. Single items are stored as autoassociations, so that the memory vector, **M** is

$$\mathbf{M} = \sum_{i=1}^{K} \alpha_1(a_i * b_i) + \alpha_2(a_i * a_i) + \alpha_3(b_i * b_i), \quad (23)$$

ignoring forgetting factors and serial position factors. To recognize a single item, one probes memory by correlating the test item with **M**, producing a resultant vector. One attempts to recall the resultant or to compare it with the test probe. For simplicity, suppose the dot product between the test probe and the resultant is used for recognition judgments.

Mathematically, this is equivalent[4] to taking the dot product of $(a_t * a_t)$ with **M**; in other words, recognition of the autoassociation is used for a decision. We will not derive the list-strength predictions for this model, but the model seems to exhibit the important properties of TODAM that bear on the list-strength effect. The model can be made sensitive to strength either by starting with background noise in **M** or by probabilistic encoding. In the case of background noise, stronger other items or repeated other items will add noise that reduces performance, and hence a positive list-strength effect should be predicted (unless the strength difference itself becomes negligibly small). Certainly, negative list-strength effects would be a problem. Like TODAM, this model tends to predict much larger list-strength effects for cued-recall than for recognition because cued recall acts like n-alternative forced-choice recognition of convolutions. Thus the observed

[4] The vectors are normally truncated to a "central group of N" after each operation, and the formal equivalence holds in this case. It holds also if no truncation occurs except when required by different vector sizes.

relation among list-strength effects in recognition, cued recall, and free recall could be difficult to handle. Finally, list-length and list-strength effects would tend to be tied together, so that it might be hard to predict the absence of one and the presence of the other. The case of probabilistic storage of features as a means of strengthening items is somewhat different. Such a model could be made to eliminate the positive list-strength prediction in the case of increased presentation time, in the manner suggested for MINERVA 2 and TODAM, but not in the case of repetitions.

James Anderson's Models

Anderson's vector model (1973) is easily seen as a special case of Murdock's TODAM model (minus the convolutions) and fails to predict the data. His (1972) matrix model in simplest form is similar to Pike's and would have similar problems. Anderson, Silverstein, Ritz, and Jones (1977) presented a generalized version of the approach (sometimes termed BSB) with a matrix of weights (synaptic connections) to associate two vectors representing items, the matrix accumulating the weights for the storage of many associations. The model has not been applied to recognition in a way that makes it a direct matter to derive predictions for the present tasks, but it shares enough features with the models we have already discussed (and the PDP models to be discussed later) that it is highly likely that a positive list-strength effect would be predicted. Basically, for a new item to reduce performance, the items must not be orthogonal to one another, and the new item's contribution to the memory matrix must add "noise" when another item is tested. In such a system, another presentation of the item, or a longer presentation, should add additional noise. The details of such arguments must await specific implementations of the theory for our paradigms.

Glanzer and Bowles' Marking Theory (1976)

The marking theory (also as elucidated in Bowles & Glanzer, 1983), assumes items to be represented by a collection of S features. At study a subset s of the S features are sampled randomly and marked as "old." Different items have features that are shared, so "indirect" marking will occur for some features of items that were not presented on that trial, including distractors that are never presented at all. At test, a sample of features is again taken, as is the number of marked features used to make a decision.

Bowles and Glanzer show that adding items to a list reduces the mean difference between target and distractor tests and also increases the variance because of the extra indirect marking of features that occurs. Thus in contrast to the other models we have discussed, the numerator of d', as well as the denominator, is affected by list length and list-strength variations. However, this just makes matters worse because both factors push performance in the same direction: Marking extra features (due, say, to extra presentation time, or extra presentations) reduces performance for two reasons: a reduced difference between targets and distractors and an increased variance.

"Connectionist" and "Neural Net" Models

There are an enormous and rapidly increasing number of composite, parallel models in this class. Very few of these have been developed as memory models and applied to standard memory phenomena, so it will not be possible to test well worked-out samples. Ratcliff (in press) has extended a few simple models in this class to memory paradigms and data and has shown that they have a number of severe problems in areas other than the list-strength effect. We think that the list-strength findings impose important additional constraints on the development of such models. A typical model in the class shall be described and qualitative arguments shall be given why such models would predict positive list-strength effects. The arguments will be bolstered by a particular instantiation also described in Ratcliff (in press).

The models to be discussed are feedforward, multilayer, network models, with learning due to back propagation (e.g., Rumelhart, Hinton, & Williams, 1986). The particular model to be applied to recognition is the "encoder network" of Ackley, Hinton, and Sejnowski (1985). This system consists of three layers of nodes: an input layer of N feature nodes, a "hidden" layer of fewer than N nodes, and an output layer of N nodes. An item is input as a vector of feature values, Q_j, on the input nodes. Each input node, j, is connected to each hidden node, i, by a weight, W_{ij}. Input to a given node, i, in the hidden layer is given by

$$\mathrm{net}_i = \sum_j W_{ij} O_j. \tag{24}$$

The output from hidden node, i, is then given by transforming the net_i value from a range of $-\infty$ to $+\infty$ to a range of 0 to 1:

$$O_i = 1/(1 + \exp(-\mathrm{net}_i)). \tag{25}$$

The weights between the hidden and output nodes are then combined with these O_i values according to Equation 24, and the results then transformed according to Equation 25, producing the final output activations of the system.

The system is then trained, by adjusting the weights suitably, to tend to reproduce its inputs as its output activations. Then when a test item is presented, the outputs may be compared with the inputs, say, by taking a dot product, and the value is used to generate a recognition decision. The weight adjustments occur by the method of back propagation. When an item traverses the system, the discrepancy between the actual and desired output at node i is given by (Rumelhart et al., 1986):

$$\delta_i = (t_i - O_i)O_i(1 - O_i), \tag{26}$$

where t_i is the desired output. The error, δ_i, is used to adjust the weights between the hidden and output layers by

$$\Delta W_{ij} = \eta \delta_i O_j, \tag{27}$$

where η is a learning rate parameter. Once the weights between the output and hidden layers are modified, another error signal is computed at the hidden nodes by

$$\delta_i = O_i(1 - O_i) \sum_k W_{ik} \delta_k . \tag{28}$$

Then the change in weights from the input to hidden nodes is given by using Equation 27 with the δ_i from Equation 28.

Each time an item is studied, and/or rehearsed, a complete cycle like that described above takes place. When an item is tested for recognition, it is input to the system, and the output activation vector is obtained. The dot product of the output and input vectors is used to generate an old–new decision. We shall present some simulation results from this system shortly, but first it will be helpful to discuss some general characteristics that illustrate likely difficulties that will beset not only this system but others like it.

One may characterize the current state of such an encoder system by the values of all weights taken together, in other words, as a point in a multidimensional "weight space." Changes in weights can be thought of as a movement of this point through weight space. When a new item is presented for the first time, the current state moves along a trajectory in weight space, so that (at least on the average), the new state is closer to one that will reproduce the presented input. If the input were orthogonal to previous inputs, the movement would not have to harm the system's ability to reproduce those previous inputs. However, we know that new inputs do harm performance (e.g., the list-length effect), so it seems reasonable to conclude that the movement "toward" a just presented item will, on the average, be associated with a movement "away" from the previously stored items. It is hard to escape the conclusion that a repetition or additional study item should move the current state even farther toward the strengthened item and even farther from the previously stored items (especially because the data show that we are operating well below ceiling—the items do become strengthened by the extra time or presentations).

The effects we have been discussing are, of course, retroactive in nature. That is, a list-strength effect is predicted when the items varying in strength follow the critical items in the list. (Indeed, these models exhibit strong retroactive effects in many areas, leading to severe difficulties, as shown by Ratcliff, in press). The situation when the items varying in strength precede the critical items is more complex. In some cases, negligible effects of strength variation on the subsequent items are predicted. It is even possible to have stronger items facilitate performance for the following critical items; this can happen when "momentum" is incorporated in the system. Momentum refers to letting a given weight change be a (linear) combination of the current change (from Equation 27) and the preceding change of weights (from the previous training or rehearsal cycle). In such a case, learning of initial items well can lead to small weight changes, and hence small carryover momentum to the critical items; the critical items are then free to adapt efficiently. However, weak learning of initial items can lead to large carryover momentum to the critical items, interfering with their weight changes.

The system as described in Equations 24–28 was used to simulate list-strength effects in the following way. There were 32 input nodes, 16 hidden nodes, and 32 output nodes. Input vectors representing different items had 32 randomly selected values that were 0 or 1. The weight space was initialized with random values in the range $-.3$ to $+.3$. The value of momentum was set equal to .5, and η was set equal to .25. It was assumed that each studied list had 16 single items, and each item was rehearsed η_i times when presented, and then not again. The values of η_i were higher for "strong" items and were systematically varied for different list types.

The study list of 16 items was divided into three groups: the first 6, the next 6, and the last 4. Different numbers of rehearsals (i.e., learning cycles) were given to the items in these three groups, as shown in Table 1; the code represents the number of rehearsals given to the items in each group, and the match values in the table (i.e., the data products) are for items in the group with the slash (/) next to it.

Strong retroactive interference effects are seen in the table, as expected. In general, this model would not be adequate on a number of grounds; for example, d' does not vary for pure lists of different strengths (2-2-2 vs. 4-4-4 vs. 8-8-8). Proactive facilitation is seen due to momentum, also as expected (e.g.,

Table 1
List Strength Predictions for the Encoder Model

Old item match	New item match	*SD* in new match	d'	Group type
.0583	.00410	.0294	1.844	2/-2-2
.0482	.00512	.0330	1.305	2/-1-2
.0448	.00436	.0332	1.218	2/-2-1
.0381	.00547	.0338	.965	2/-2-2
.0251	.00636	.0410	.457	2/-4-2
.0270	.00773	.0493	.391	2/-2-4
.0212	.00899	.0482	.253	2/-4-4
.0695	.00335	.0309	2.141	1-2/-1
.0682	.00436	.0332	1.923	2-2/-1
.0588	.00456	.0344	1.577	1-2/-2
.0584	.00547	.0338	1.566	2-2/-2
.0550	.00721	.0350	1.365	4-2/-2
.0422	.00773	.0493	.699	2-2/-4
.0430	.00931	.0450	.749	4-2/-4
.0766	.00721	.0350	1.983	4/-2-2
.0526	.00796	.0382	1.168	4/-4-2
.0577	.00931	.0450	1.075	4/-2-4
.0447	.0104	.0439	.781	4/-4-4
.0356	.0129	.0459	.496	4/-8-4
.0346	.0144	.0658	.307	4/-4-8
.0345	.0175	.0595	.286	4/-8-8
.105	.00636	.0410	2.406	2-4/-2
.100	.00796	.0382	2.401	4-4/-2
.0790	.00899	.0482	1.452	2-4/-4
.0801	.0104	.0439	1.588	4-4/-4
.0756	.0126	.0367	1.717	8-4/-4
.0561	.0144	.0658	.634	4-4/-8
.0623	.0175	.0556	.806	8-4/-8
.0859	.0126	.0367	1.997	8/-4-4
.0703	.0155	.0411	1.333	8/-8-4
.0733	.0175	.0556	1.003	8/-4-8
.0683	.0204	.0537	.892	8/-8-8
.0705	.0273	.0561	.770	8/-16-8
.0584	.0260	.0863	.375	8/-8-16
.0638	.0323	.0827	.381	8/-16-16
.128	.0129	.0459	2.508	4-8/-4
.130	.0155	.0411	2.785	8-8/-4
.100	.0175	.0595	1.387	4-8/-8
.109	.0204	.0537	1.650	8-8/-8
.116	.0267	.0496	1.800	16-8/-8
.0805	.0260	.0863	.632	8-8/-16
.0915	.0332	.0806	.723	16-8/16

4-4/-4 vs. 8-4/-4). A fuller discussion of these and other effects may be found in Ratcliff (in press). Because the model predicts such strong serial position effects, it is not appropriate to utilize our normal measure of list-strength effects, R_r. Rather we shall focus upon items of a given strength in a given set of serial positions and shall consider the effects of variations in the strengths of items in *other* serial positions. The list types in each block in the table may be compared in this fashion, and they are ordered from top to bottom in each block from minimum to maximum strength of *other* items. Blocks 1, 3, and 5 illustrate retroactive effects, and uniformly demonstrate strong list-strength effects: Stronger other items reduce performance. Blocks, 2, 4, and 6 illustrate proactive effects and are less clear. Positive list-strength effects are generally seen, but a number of reversals occur especially in Block 6, where the items become so strong that differential momentum effects are seen. Because momentum can be thought of as a kind of transfer of processing effort from later items in a list to earlier weaker items, such reversals are not surprising.

All in all, these simulation results provide strong verification of the general line of reasoning discussed earlier, and they demonstrate the difficulty encountered in attempting to make such models account for our list-strength results. Although the simulation results speak directly only to this particular model instantiation, we believe that list-strength findings may be difficult to handle in the context of many models in this class.

Assessment of Models

It is clear that none of the current models in their prototypical form can handle the findings laid out in Ratcliff et al. (1990) concerning the list-strength effect. The problem may be summed up as follows. For the SAM, MINERVA 2, TODAM, MATRIX, and CHARM models a positive list-strength effect is predicted for recognition because the variance of activation caused by a test item is higher when other items (than that being tested) are stronger. The other models (e.g., the network theories, the marking theory, the BSB theory) fail partly for the same reason, but also because the mean activation difference between targets and distractors is affected by the strength of other items.

As a step toward solving the problem, we looked for ways to make the activation variances approximately equal when other list items vary in strength. No way to do so suggested itself when strength was manipulated by spaced repetitions for any of the composite models (though for some of these models, solutions could be found when strength was varied by presentation time manipulations). In the case of SAM and MINERVA 2 (which assumed separate storage of traces), it was assumed that repetitions were accumulated into a single memory trace, thereby making the repetitions condition equivalent to the presentation time condition. For these two models, several methods to produce constant variance were considered. The solutions discussed for MINERVA 2 left several problems unresolved: First, recall and recognition tended to give rise to equivalent list-strength predictions, and second, no flexibility in the list-strength predictions for recognition was available (so that zero or negative list-strength effects could both be predicted).

The preferred solution within the SAM framework seemed to resolve most of the problems. It was assumed that strengthening of an item produces two effects: (a) The activation of that item's trace by the context cue *increases* as the strength increases; (b) The activation of that item's trace by an extraneous different, test item *decreases* as the strength increases. The latter is a *differentiation* assumption. These two factors counteract each other, so that the mean and variance of activation of a trace by an extraneous test item are approximately constant as strength of that trace is varied. However, the degree of trade-off between the two factors will be determined by the exact shape of the relevant strength functions, the two levels of strength involved in the study, and the relative weights given to the context and item cues. Thus the model has the flexibility to predict a range of list-strength effects (including some negative) across different recognition studies and also the ability to predict different list-strength effects for recognition and cued recall within a study (if the cue weightings at retrieval differ slightly for cued recall and recognition). Finally, the model has no trouble predicting large list-strength effects in free recall because according to SAM in free recall the context cue is sometimes used alone to probe memory. In such a case, stronger images will be preferentially sampled at the expense of weaker ones, producing a large positive list-strength effect.

It should be emphasized that our theoretical conclusions are firm only for models and variants for which derivations have been produced. The fact that we could not find appropriate variants within the context of a large number of models does not mean that they cannot be found. In this sense, the results in this article (and Part I) can be thought of as a challenge to proponents of particular theories, and as a guide to future theory development. The models to date all seem to have been designed to predict list-*length* effects; in the future it would be desirable to have them predict the pattern of list-*strength* effects as well.

The implication of the list-strength results for theory are, of course, predicated upon the assumption that effort redistribution in mixed lists is not enhancing weak items at the expense of strong items. Although conditions utilized in Part I (Ratcliff et al., 1990) tended to rule out redistribution of rehearsal or coding to nearby positions and although instructions in some studies, and conditions of incidental learning in Experiment 7, make redistribution explanations less likely, additional and more direct evidence concerning effort redistribution would be desirable. It should be noted that redistribution of effort in mixed lists, if it is to explain our findings, must have the following effects: (a) Strength differences in mixed lists must get smaller but remain positive; the shift in strength must roughly cancel out the factors otherwise tending to produce a positive list-strength effect in recognition, and to a large degree in cued recall, while leaving large list-strength effects predicted in free recall. (b) The mechanism underlying redistribution must be insensitive to local effects in neighboring study positions and insensitive to instructions and conditions of incidental learning.

If we rule out redistribution provisionally, then models must predict missing or negative list-strength effects in recognition, a small positive list-strength effect in cued recall, and a large positive list-strength effect in free recall, whether strength variations are produced by variations in presentation time or by variations in numbers of massed or spaced repetitions.

At the present time, the only solution we see to the puzzle posed by such list-strength findings (if not due to redistribution) involves the following hypotheses that apply in the context of our studies:

1. Different items are stored separately in memory (on at least one level in the system—composite storage might be quite possible on other levels).

2. Repetitions are accumulated into a single, separate, memory trace (which, curiously, is a form of composite storage), at least for the conditions of our studies.

3. Free recall operates differently than recognition, possibly on the basis of sampling images according to their strength.

In addition, our preferred solution involves another assumption, one easy to implement within the SAM framework, that the part of the activation of an image that is induced by a different item-cue, one not rehearsed with the item encoded in that image, gets lower as the image is encoded more strongly. This is a form of *differentiation* hypothesis.

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. (1972). A simple neural network generating an interactive memory. *Mathematical Biosciences, 14,* 197–220.

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

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

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

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 Floweree Symposium on Cognition* (pp. 313–344). Hillsdale, NJ: Erlbaum.

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

Clark, S. E. (1988). *A theory for classification and memory retrieval.* Unpublished doctoral dissertation, Indiana University, Bloomington, Indiana.

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

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

Gibson, E. J. (1940). A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review, 47,* 196–229.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32–41.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 19,* 1–65.

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

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.

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

Hintzman, D. (1988). Judgments of frequency and recognition memory in a multiple trace memory model. *Psychological Review, 95,* 528–551.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96,* 208–233.

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.

Metcalfe, J., & Murdock, B. B., Jr. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behaviour, 20,* 161–189.

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

Murdock, B. B., Jr., & Lamon, M. (1988). The replacement effect: Repeating some items while replacing others. *Memory & Cognition, 16,* 91–101.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87–108.

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

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* (Vol. 14, pp. 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. (in press). Connectionist models of memory: Constraints imposed by learning and forgetting functions. *Psychological Review.*

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.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition: Vol. 1 Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Saltz, E. (1961). Response pretraining: Differentiation or availability? *Journal of Experimental Psychology, 62,* 583–587.

Saltz, E. (1963). Compound stimuli in verbal learning: Cognitive and sensory differentiation versus stimulus selection. *Journal of Experimental Psychology, 66,* 1–5.

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

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13–34). New York: Wiley.

Appendix A

Derivation of List-Strength Predictions for the SAM Model

The first thing to note is that $u(\text{ps}) = u(\text{ms}) > u(\text{pw}) = u(\text{mw})$. Why? By assumption, activation of an image is the same residual value, d, whether the test item is new or an item from some list pair *other* than the one containing the image in question (because rehearsal is assumed to occur only *between* members of a study pair). Thus every term in the sum making up Equation 4a is equal to a corresponding term in Equation 4b, except for the two terms representing items in the pair containing the test item. Thus,

$$\text{u(ms)} = u(\text{ps}) = \bar{A}(\text{I}_{is}\,|\,\text{I}_{is}, \text{C}) + \bar{A}(\text{I}^*_{js}\,|\,\text{I}_{is}, \text{C}) - \bar{A}(\text{I}_{is}\,|\,\text{I}_x, \text{C}) - \bar{A}(\text{I}^*_{is}\,|\,\text{I}_x, \text{C}),$$

where the subscript, is, refers to a strong image or item, and a superscript * refers to the image or item which is the other member of the same pair. Similarly,

$$u(\text{mw}) = u(\text{pw}) = \bar{A}(\text{I}_{iw}\,|\,\text{I}_{iw}, \text{C}) + \bar{A}(\text{I}^*_{iw}\,|\,\text{I}_{iw}, \text{C}) - \bar{A}(\text{I}_{iw}\,|\,\text{I}_x, \text{C}) - \bar{A}(\text{I}^*_{iw}\,|\,\text{I}_x, \text{C}).$$

As a consequence, Equations 2 and 3 hold for SAM, in the form given by Equation 6 in the test.

This derivation confirms what intuition suggests: If the variance associated with a strong image is larger than that associated with a weak image, a list-strength effect is predicted. To go further, some additional details are needed. First, in order to have wider applicability for the results, the distributions of strengths about the means are generalized from those in Equation 5. Assume that for any strength of mean X,

$$\text{P}[S = \gamma_i X] = \text{p}_i;\ \sum_i \text{P}_i = 1;\ \sum_i \text{p}_i\gamma_i = 1;\ \gamma_i > 0. \quad \text{(A1)}$$

Then,

$$\text{E}[S] = \sum_i \gamma_i X \text{p}_i = X$$

as stipulated.

$$\text{Var}^{1/2}[S] = \left(\sum_i \gamma_i^2 X^2 \text{p}_i - X^2\right)^{1/2} = X\left(\sum_i \gamma_i^2 \text{p}_i - 1\right)^{1/2},$$

so that the standard deviation rises linearly with the mean, as desired.

In the derivation we will need

$$\text{E}[S^W] = \sum_i \gamma_i^W X^W \text{p}_i = X^W \sum_i \gamma_i^W \text{p}_i.$$

Let

$$\theta_W = \sum_i \gamma_i^W \text{p}_i.$$

Then

$$\text{E}[S^W] = (\text{E}[S])^W \theta_W.$$

The remaining derivations use the facts that

$$\text{E}(\Sigma X_i) = \Sigma\text{E}(X_i),$$

$$\text{Var}(\Sigma X_i) = \sum_i \text{Var}(X_i)$$

if X_i are independent, and

$$\text{E}(\Pi X_i) = \Pi\text{E}(X_i)$$

if X_i are independent.

We are now ready to proceed. Assume that $S(\text{C, I})$ is independent of $S(\text{I}_x,\text{I})$. Then,

$$\begin{aligned}
\text{Var}[A(\text{I}_s\,|\,\text{I}_x, \text{C})] &= \text{Var}[S(\text{I}_x, \text{I}_s)^{W_1}S(\text{C}, \text{I}_s)^{W_C}] \\
&= \text{E}[\{S(\text{I}_x, \text{I}_s)^{W_1}S(\text{C}, \text{I}_s)^{W_C}\}^2] \\
&\quad - \text{E}^2[S(\text{I}_x, \text{I}_s)^{W_1}S(\text{C}, \text{I}_s)^{W_C}] \\
&= \text{E}[S(\text{I}_x, \text{I}_s)^{2W_1}]\text{E}[S(\text{C}, \text{I}_s)^{2W_C}] \\
&\quad - \text{E}^2[S(\text{I}_x, \text{I}_s)^{W_1}]\text{E}^2[S(\text{C}, \text{I}_s)^{W_C}] \\
&= \text{E}^{2W_1}[S(\text{I}_x, \text{I}_s)]\theta_{2W_1}\text{E}^{2W_C}[S(\text{C}, \text{I}_s)]\theta_{2W_C} \\
&\quad - \text{E}^{2W_1}[S(\text{I}_x, \text{I}_s)]\theta^2_{W_1}\text{E}^{2W_C}[S(\text{C}, \text{I}_s)]\theta^2_{W_C} \\
&= \text{E}^{2W_1}[S(\text{I}_x, \text{I}_s)]\text{E}^{2W_C}[S(\text{C}, \text{I}_s)][\theta_{2W_1}\theta_{2W_C} - \theta^2_{W_1}\theta^2_{W_C}].
\end{aligned} \quad \text{(A2)}$$

Similarly,

$$\text{Var}[A(\text{I}_w\,|\,\text{I}_x, \text{C})] = \text{E}^{2W_1}[S(\text{I}_x, \text{I}_w)]\text{E}^{2W_C}[S(\text{C}, \text{I}_w)][\theta_{2W_1}\theta_{2W_C} - \theta^2_{W_1}\theta^2_{W_C}].$$

Using this result and substituting into Equation 6, we get Equation 7, given in the text. Because in the old version of SAM (i.e., Gillund & Shiffrin, 1984), $\text{E}[S(\text{I}_x,\text{I}_s)] = \text{E}[S(\text{I}_x,\text{I}_w)]$, and assuming that $\text{E}[S(\text{C},\text{I}_s)] = a\alpha f(t)$, $\text{E}[S(\text{C},\text{I}_w)] = af(t)$ we have

$$R_r = \alpha^{W_C}; \qquad \alpha > 1; \qquad W_C > 0.$$

For the case of spaced repetitions, when the repetitions are not necessarily equal in expectation, $\delta_j = \text{E}[S(\text{C},\text{I}_{ij})]$ is the mean context strength for the jth repetition for item i. Assume that repetitions are stored separately and independently of each other (and that they are independent of different items). Then we need to calculate the variance of the activations for the two terms in Equation 10. For an item with r repetitions,

$$\begin{aligned}
\text{Var}\left[\sum_{j=1}^{r} A(\text{I}_{ij}\,|\,\text{I}_x, \text{C})\right] &= \sum_{j=1}^{r} \text{Var}[A(\text{I}_{ij}\,|\,\text{I}_x, \text{C})] \\
&= \sum_{j=1}^{r} \text{E}^{2W_1}[S(\text{I}_x, \text{I}_{ij})]\text{E}^{2W_C}[S(\text{C}, \text{I}_{ij})][\theta_{2W_1}\theta_{2W_C} - \theta^2_{W_1}\theta^2{}_{W_C}],
\end{aligned}$$

(from Equation A2)

$$= d^{2W_1}[\theta_{2W_1}\theta_{2W_C} - \theta^2_{W_1}\theta^2_{W_C}] \sum_{j=1}^{r} \delta_j^{2W_C}.$$

Substituting in Equation 10 gives the result in Equation 12 in the text.

Appendix B

Variants of the SAM Model Designed To Handle the List-Strength Findings

No Context Cue

Equation 9 shows that SAM does not have to predict a list-strength effect: Setting the weight given to the context cue equal to zero will cause R_r to be 1.0. While technically correct, this observation is not a very good solution to the puzzle posed by the present data. The fundamental problem is the fact that SAM was developed and applied to many forms of data with a substantial value for the weight given to the context cue. In applications to date, W_c has almost always been as large as any other cue weight, has usually been equal to .5, and has never been set to a value lower than .33. (We have explored the model's parameter space when W_c was set to .5 and have shown that the model fails badly to predict our data.) The most important reason for a high value of W_c lies in the need to focus memory access upon images in the target list, rather than upon the myriad other images in memory. Putting this another way, in recognition every memory image I potentially contributes a mean value of d^{W_I} (the weighted, residual, item to item strength) to the total activation, were it not for the other term multiplicatively determining activation: $S(\mathrm{C,I})^{W_C}$. If W_C were zero, then no focusing on the recently presented list would be possible. If W_C were very small, then $S(\mathrm{C,I})$ would have to be extremely small indeed, so that the sum of terms $d^{W_I}S(\mathrm{C,I})^{W_C}$ across all images in memory other than the test list would be negligible. That is, the context cue would have to be strongly connected to images of items on the recent list, but almost totally unconnected to other images in memory. If this were so, it would be hard to explain the existence of intrusions from recent lists, ordered in terms of list recovery, in recall tasks (e.g., Shiffrin, 1970). It would also be hard to explain the great difficulty that occurs when distractors are items from previous lists (e.g., Anderson & Bower, 1972). Also, there would be no way to predict the occasional findings of recognition decrements when context is changed between study and test (e.g., Bjork & Richardson-Klavehn, 1989; Smith, 1988). In addition to these problems, many of the applications of SAM have assumed a large value of W_c, and a change in this assumption would require alternative versions of SAM to explain the findings. Finally, we should point out that no assumptions concerning the value of W_c could explain the significantly negative list-strength effects that were found in some of our studies.

We therefore turn to possible variations on the basic SAM model that might be able to handle the present findings. Attention is restricted to variants in which a substantial weight is given to the context cue (say, on the order of .5). The variations involve methods by which the variance terms in Equation 6 may be made a (approximately) constant function of the mean activation strength. That is, a probe of memory by a distractor cue and a context cue must produce image activation whose variance does not increase with the strength of storage of the image.

Constant Context Strength

Assume that context strength does not increase when an item is studied longer or more times. That is, the presentation of an item causes an initial increase in context strength, but no further increase with extra time or repetitions. Reference to Equations 7 and 9 show that R_r will be 1.0 (ultimately deriving from the fact that the variance will be a constant, as shown in Appendix A). However, this assumption violates the logic of the SAM model. For example, if an item is presented and stored with some context strength and then presented again after a spaced interval, it is hard to imagine what mechanism would prevent any additional context strength from being stored at the second presentation. Aside from such conceptual difficulties, the present assumption would alter substantially the predictions of SAM described in Raaijmakers and Shiffrin (1980, 1981) and Gillund and Shiffrin (1984), because all learning effects would have to be due to the growth of interitem associations. Even ignoring such difficulties, the assumption of constant context strengths cannot produce significantly negative list-strength effects. Although such a result might be predicted by a model variant in which increased time or presentations produce a decrease in context strength, such an assumption is unreasonable, and we have not further pursued variations of this hypothesis.

Independence of Mean Strength from Variance

Assume that context strength grows but that strength is independent of variance. The simplest version of this variant replaces the variance assumptions incorporated in Equation 5 (or the generalization in Appendix A) with the assumption that the variance of the activation of any image is constant regardless of the mean level of activation. Such an assumption would be consistent with the hypothesis that variance arises in the retrieval process of image activation, rather than arising from item–subject differences, from differing coding and rehearsal amounts and strategies, or from differences in particular cue–image relation. Because these latter sources of variance undoubtedly exist, perhaps it would be best to suggest that the variance they contribute is negligibly small in comparison with the variance due to the process of image activation.

An advantage of this theory is the ease with which certain predictions can be derived for the SAM model. If the constant variance is denoted V for any image, then the denominator in Equation 1 is just $(NV)^{1/2}$ when there are N images in the part of memory accessed by the context cue (i.e., the recent list). Thus d′ is simply determined by a difference in mean activation levels, and these can usually be written down by inspection.

Another useful characteristic of this theory is the fact that the pattern of prediction true of the original SAM theory is also true of the modified theory. The one exception we have found lies in the area of context shift between study and test. The original form of the theory predicted no effect of context shift on recognition performance, only upon recall (see Gillund & Shiffrin, 1984). The modified theory predicts context should harm performance in recognition as well. This difference in predictions does not argue strongly for one form of the theory over the other because the current data concerning context shifts in recognition studies are highly variable in outcome (e.g., see Bjork & Richardson-Klavehn, 1989; Smith, 1988).

This new variant of SAM is not without problems, however. The old version of SAM, because it assigned variance to particular combinations of cues and images, could account naturally for covariances and correlations. For example, if forced choice recognition were utilized, the context cue would be common to the probes of memory with each of the test items from which the choice must be made. In the original model, the covariance of the activations for each probe could be calculated by taking into account the fact that each image is activated by one cue that is the same for each probe. There is no natural way to handle covariance in the new version of SAM: Some sort of covariance assumptions would have to be added to the model in ad hoc fashion (see Clark, 1988, for one approach).

Another problem with this variant of SAM is common to all the versions we have discussed thus far: There is no way to explain the range of list-strength effects and the fact that some are significantly negative. It could be argued that some borrowing of rehearsal from strong items to give to weak items could explain negative list-strength effects, even though the theory predicts $R_r = 1.0$. This argument is weakened by Experiment 7, in Part I, however, in which the incidental task that should have reduced rehearsal redistribution to a minimum produced a strongly negative list-strength effect.

Received August 19, 1988
Revision received July 18, 1989
Accepted July 24, 1989 ■