

Memorability as a Measure of Processing: A Unit Analysis of Prose and List Learning

David C. Rubin
Duke University

SUMMARY

The percentage of subjects recalling each unit in a list or prose passage is considered as a dependent measure. When the same units are recalled in different tasks, processing is assumed to be the same; when different units are recalled, processing is assumed to be different. Two collections of memory tasks are presented, one for lists and one for prose. The relations found in these two collections are supported by an extensive reanalysis of the existing prose memory literature.

The same set of words were learned by 13 different groups of subjects under 13 different conditions. Included were intentional free-recall tasks, incidental free recall following lexical decision, and incidental free recall following ratings of orthographic distinctiveness and emotionality. Although the nine free-recall tasks varied widely with regard to the amount of recall, the relative probability of recall for the words was very similar among the tasks. Imagery encoding and recognition produced relative probabilities of recall that were different from each other and from the free-recall tasks. Similar results were obtained with a prose passage. A story was learned by 13 different groups of subjects under 13 different conditions. Eight free-recall tasks, which varied with respect to incidental or intentional learning, retention interval, and the age of the subjects, produced similar relative probabilities of recall, whereas recognition and prompted recall produced relative probabilities of recall that were different from each other and from the free-recall tasks.

A review of the prose literature was undertaken to test the generality of these results. Analysis of variance is the most common statistical procedure in this literature. If the relative probability of recall of units varied across conditions, a units by condition interaction would be expected. For the 12 studies that manipulated retention interval, an average of 21% of the variance was accounted for by the main effect of retention interval, 17% by the main effect of units, and only 2% by the retention interval by units interaction. Similarly, for the 12 studies that varied the age of the subjects, 6% of the variance was accounted for by the main effect of age, 32% by the main effect of units, and only 1% by the interaction of age by units. Moreover, although there was enough statistical power so that the unit main effects, which used the same error term as the interactions, were all significant at the .001 level, only half of the age by units interactions were significant at the .05 level. Unlike studies varying retention interval and the age of the subjects, studies varying the perspective of the reader at encoding often produced interactions larger than their main effects.

Similar results were obtained for the list learning and prose studies, suggesting a continuity between these two domains. In particular, the relative probability of recall changed little when retention interval, age of subject, or depth of processing changed in standard ways. This similarity among the tasks challenges some current views of retention and of depth of processing. In both list learning and prose studies, changes in retrieval task and some changes in initial encoding led to differences in the relative probability of recall, suggesting that these manipulations affect processing. It is argued that the descriptive unit analysis approach is an efficient way to approach theory development given our current state of knowledge about human memory.

This article is about human memory, but it is also about how psychology is done. At the lowest level, this article is an argument for including another dependent variable in memory studies. At the middle level, it is a review of the substantive findings that result from including this dependent variable and the changes they suggest in our current view of memory. At the top level, this article is an attempt to demonstrate that searching for quantitative regularity in complex cognitive behavior is a fruitful approach at psychology's current stage of development as a science.

On Unit Analyses

Let us start at the lowest level. Psychologists interested in memory rely on numerous quantitative dependent measures such as the amount, the order, and the latency of recall, yet some of our most influential studies have been qualitative descriptions of which aspects of stimuli are recalled and which are forgotten or distorted (e.g., Bartlett, 1932/1967; Freud, 1904/1951). The quantitative rigor of our usual quantity-oriented dependent measures and the richness of the more descriptive quality-oriented approach can, however, be combined. Table 1 represents the results of a hypothetical, but typical experiment in which 5 subjects recalled, or failed to recall, each of seven items. The items might be nonsense syllables, words, or propositions from a story. The bottom row is the number of items each subject recalled. The right-hand column is the number of subjects recalling each item.

The analyses performed on the bottom row

Table 1
Hypothetical Recall Data

Units	Subjects					Σ
	1	2	3	4	5	
1	1	1	1	1	1	5
2	1	1	0	0	1	3
3	0	1	0	0	1	2
4	1	0	1	0	1	3
5	0	0	0	0	0	0
6	1	1	1	0	1	4
7	1	1	0	1	1	4
Σ	5	5	3	2	6	

and the right-hand column are usually not symmetrical in a single study. In studies of individual differences the bottom row is maintained as an ordered set of five means and a reliability measure on that ordered set, and the right-hand column is ignored once a test is constructed. In contrast, in experimental psychology (Cronbach, 1957) the bottom row is usually collapsed into a single mean and a reliability measure on that mean (i.e., a standard error), whereas the right-hand column, when examined at all, is maintained as an ordered set of seven means and a reliability measure on that ordered set (e.g., Cronbach's alpha; Cronbach, 1951). The right-hand column provides a richer, more complex measure that allows a closer approach to purely descriptive techniques while maintaining quantitative rigor. Using correlations, this ordered set of seven means can be compared with any other sets of seven means obtained by presenting the same seven units to subjects under different conditions and to any independent variables that can provide values for the seven units. The resulting correlations can be submitted to multiple regression, factor analysis, or multidimensional scaling. That is, the techniques usually applied in individual differences research to predict the behavior of individual subjects can be applied in experimental psychology to predict the behavior of individual units. Moreover, none of these techniques interfere with the more common methods based on bottom row, amount recalled, analyses.

Although not common, using the number of subjects recalling each unit as a dependent variable is hardly without precedent. Rather, the method has appeared sporadically throughout the history of the study of verbal

Preparation of this article was supported in part by National Institute of Aging Grant RO1 AG04278 and National Science Foundation Grant BNS 8410124. Portions of this article were presented at the 23rd Annual Meeting of the Psychonomic Society, Minneapolis, November 1982.

I wish to thank Alan Baddeley, Gus Craik, Herbert Crovitz, Gregory Kimble, Walter Kintsch, Reed Hunt, Gregory Lockhead, John Staddon, Endel Tulving, and Michael Watkins for their comments, and Roger Dixon, Bruce Dunn, Gil Einstein, Mark Grabe, Art Graesser, Walter Kintsch, Michael Masson, Samuel Mathews II, Tom Petros, Stephen Schmidt, Eugene Schultz, and Donald Smith for providing me with their data.

Request for reprints should be sent to David C. Rubin, Department of Psychology, Duke University, Durham, North Carolina 27706.

materials. For example, with lists of isolated items, Underwood and Schultz (1960) used such a measure in their pioneering work on nonsense syllables, Paivio (1968) used the measure with isolated words, and Clark (1973) made use of this measure as a way of assessing the generalizability of findings to new material. In text research, the measure is more common (e.g., Johnson, 1970; Kintsch, 1974). In fact, Voss, Tyler, and Bisanz (1982), in summarizing methods of text research, grouped all studies using this dependent variable together under the title "unit recall."

Depth, Processing, and Depth of Processing

The proposed unit analysis dependent variable is especially well suited to our current views of memory. Assume that what people do during a task affects their later recall (e.g., Craik & Lockhart, 1972; Hyde & Jenkins, 1969; Meacham, 1972). Further assume that two factors affect what people do: (a) the general attitude, strategy, or cognitive processing prevailing during the task and (b) the ability of the particular items encountered to support that processing (Hunt & Elliott, 1980; Hunt, Elliott, & Spence, 1979; Jacoby, Bartz, & Evans, 1978; Rubin, 1980; Schultz, 1982; Seamon & Murray, 1976). It then follows that the application of the ongoing cognitive processing to the particular items encountered will in part determine both how many and which items are later recalled. Of course, it is possible that recall is at best a poor indicator of the processing of particular items. If this were the case, then results amassed by using both the standard depth of processing approach and the unit analyses presented here would remain as a substantial set of empirical findings in need of a theoretical framework. Such an alternative framework is not considered here because assuming that memorability indexes processing provides a reasonable account of the data and allows existing theoretical issues to be addressed. However, it should be made clear that the empirical descriptions of which items are recalled under various experimental conditions provided here are independent of the processing framework in which they are presented.

In order to avoid confusion, the phrase *the depth of processing approach* will be divided

into its two components. The first component is the *processing approach*, which holds that the cognitive processing present at encoding affects later recall. The second component is the inherently unidimensional concept of *depth*. Terms like *depth* and *level* imply a single vertical line along which objects, or processing, can be measured. One can even have spread perpendicular to the depth line at various specified levels. Even terms like *distinctiveness* and *elaborateness* imply a single, comparative dimension on which objects have a relative, more or less, location. The view presented in this article is that this unidimensional depth component is an unfortunate result of using the unidimensional *amount-recalled dependent measure* to assess the value of the processing approach. Once the division into a depth and a processing component is made, it is easier to view processing as varying in kind instead of along single dimension such as depth. The combined term *depth of processing* will be retained to refer to the standard view that combines the processing approach with the amount-recalled measure.

Craik and Tulving (1975), in describing the early depth of processing approach, provide the rationale for separating the processing approach from depth. Their basic argument makes predictions about the retention of particular items by individual subjects, rather than on how the data should be aggregated to form dependent measures. The latter issue is the focus of this article. Craik and Tulving state

Stimuli which do not receive full attention, and are analyzed only to a shallow sensory level, give rise to very transient memory traces. On the other hand, stimuli that are attended to, fully analyzed, and enriched by associations or images yield a deeper encoding of the event, and a long-lasting trace. (p. 270)

Of course, the concepts of shallow and deep have changed with a decade of research, but the basic argument has not. As the type of processing changes (e.g., from associations to images, to use Craik and Tulving's example), the extent to which particular stimuli that can support processing changes and, therefore, so must the relative probability of recall of those stimuli. That is, when all subjects in a group use the same processing and the items vary in how well they can support that pro-

cessing, the processing approach, and for that matter the classical depth of processing approach, makes clear predictions about the retention of particular items. For instance, on the basis of Paivio, Yuille, and Madigan's (1968) norms, a person using an imagery technique to learn a list might remember *circle* more often than *soul*, whereas that same person using an associative technique might remember *soul* more often than *circle*. This is because *circle* can support more imagery processing than *soul*, and *soul* can support more associative processing than *circle*. Similarly, if context changes the relative amount or kind of processing a proposition in a prose passage can support, then the relative probability of recall of that proposition should also change (Bransford & Johnson, 1972; Einstein, McDaniel, Bowers, & Stevens, 1984; Grabe, 1979, 1981; Meyer, 1975, 1977; Pichert & Anderson, 1977; Rubin, 1978).

Unlike the standard depth of processing approach, in the processing approach a change in the total amount recalled, but not in the relative probability of recall of individual items, would not imply a change in the depth or kind of processing. Such a change could be due simply to a change in the gross amount of processing being applied to all items. For example, if the time per item or the attention directed to the items as opposed to a secondary task were changed, the total amount recalled would change. There would be no need to assume that the kind of processing, or its relative allocation among items, changed. That is, the commonly used amount-recalled dependent variable may be less clearly related to the basic processing claim than the alternative relative probability dependent variable proposed here. At least the alternative dependent measure presented here is as plausible as the more common amount-recalled dependent measure.

If the alternative put forth here is correct, it should be possible to identify empirically those tasks that use similar cognitive processing and to separate them from other tasks that use different underlying processing. That is, a similarity space of tasks in terms of the underlying cognitive processing should be possible: Tasks using similar processing should tend to have the same particular units re-

called, and tasks using different processing should tend to have different units recalled.

The approach being put forward here will first be applied to the learning of lists of unrelated words. A parallel analysis will then be applied to prose and supported by a review of the prose literature. On the basis of these analyses, the value of the unit analysis dependent variable and the findings that result from its use will be evaluated.

A Unit Analysis of List Learning

The Analysis

The first 10 entries in Table 2 are from one study (Rubin, 1980) in which the same 125 words were learned by different groups of subjects under different conditions. The last three entries of Table 2 were added to this collection in order to provide a standard set of depth of processing tasks. In order to obtain these entries, 116 Duke University undergraduates participated as part of an introductory psychology course requirement. Each subject was presented with a booklet that asked that 125 words be rated on a scale of 1 to 7 for either *orthographic distinctiveness*, *emotionality*, or *imagery*. Before beginning, the subjects, who were tested in groups, were told the importance of such ratings and were instructed to rate one word every 5 s, as signaled by an audible click. The rating scales were chosen (a) because they represented a range of depth of processing (Craig & Lockhart, 1972; Craig & Tulving, 1975; Hyde & Jenkins, 1969) with orthographic distinctiveness representing shallow processing and emotionality and imagery representing deep processing and (b) because the ratings were essentially uncorrelated (all r 's < .12) for the sample of 125 words used here (Rubin, 1980). Thus, for example, words that could support more imagery processing would not, on the average, be able to support more or less emotionality processing, and similarities in recall could not be attributed to such confoundings.

A total of 20 random orders and their reverse orders were prepared, each order being used in one booklet in each rating scale. There were 39, 38, and 39 subjects in the orthographic distinctiveness, emotionality, and

imagery conditions, respectively. A surprise free recall followed 30 s after the ratings were completed.

The basic amount-recalled depth of processing prediction was confirmed with 13, 22, and 24% of the words being recalled in the orthographic distinctiveness, emotionality, and imagery conditions, $F(2, 112) = 33.71$, $p < .001$. The level of recall in the orthographic distinctiveness condition was lower than that of the other two conditions, which did not differ significantly (Tukey's (*a*) test, $p < .05$). The number of subjects recalling each unit was calculated for the unit analysis.

The reliabilities for these recall values from the orthographic distinctiveness, emotionality, and imagery conditions were .78, .90, and .87, respectively (Cronbach's alpha; Cronbach, 1951).

Table 3 presents the 13 tasks of Table 2 ranked by the amount-recalled dependent variable. Table 3 is included to make graphic and explicit the unidimensional nature of the amount-recalled dependent variable: Tasks can vary only in amount, not in kind. If a priori claims about the depth of various forms of processing were made, if minor variations in presentation rate and subject

Table 2
List Learning Tasks

Name	Label	Description
Free recall with overt rehearsal	FROR	36 subjects, list length of 16, 5 s/word, 30-s filled retention interval
Free recall with overt rehearsal of only the word showing	FR1	Same as FROR, but only word showing could be rehearsed
Free recall: Fast presentation	FRFP	30 subjects, list length of 125, 1 s/word, 30-s filled retention interval
Free recall lexical decision	FRLD	36 subjects, list length 125, surprise recall after lexical decision task
Paired-associate stimulus	PAS	76 subjects, list length of 21 pairs, standard paired-associate task with number of pairs completed correctly tallied by stimulus member
Paired-associate response	PAR	Same as PAS but tallied by response member
Free recall from paired-associate stimulus	FRPAS	Surprise recall following the complete associate task. Stimulus words scored separately
Free recall from paired-associate response	FRPAR	Same as FRPAS but response words scored
Recall using a mnemonic pathway	RMN	34 subjects, 10 s/word, list length of 42, learned by using the previously mastered method of loci
Recognition	Recg	5-item forced choice, 4 s/choice following FRFP
Free recall: Orthographic	FROD	39 subjects 5 s/word, list length of 125, surprise recall following rating of orthographic distinctiveness
Free recall: Emotionality	FRE	38 subjects, same as FROD but rating emotionality
Free recall: Imagery	FR1	39 subjects, same as FROD but rating imagery

Table 3
Tasks Listed on Table 2 Ordered by Depth Under the Assumption that Percentage Recalled is an Indicator of Depth

Task	Percentage recalled	Incidental learning	Free recall
Free recall: Fast presentation	11		x
Free recall: Orthographic distinctiveness	14	x	x
Free recall: Lexical decision	15	x	x
Free recall: Emotionality	22	x	x
Free recall: Imagery	25	x	x
Recognition	48		
Free recall from paired-associate stimulus	48	x	x
Free recall from paired-associate response	49	x	x
Free recall with overt rehearsal of only the word showing	56		x
Free recall with overt rehearsal	58		x
Paired-associate stimulus	70		
Paired-associate response	70		
Recall using mnemonic pathway	82		

population were overlooked, and if only the tasks involving incidental learning were included, Table 3 would be typical of the results of standard amount-recalled depth of processing studies. The tasks using deeper processing at the time of encoding lead to a higher percentage of recall and, therefore, appear deeper down on the table. All the free-recall tasks could be included, if intentional learning of various kinds is considered as a type of encoding process to be dealt with by the depth of processing approach. In either case, the empirical results agree with what would be expected: Orienting tasks, including intentional learning, which are commonly considered to use deeper processing, lead to more recall. This result is uninteresting only because it has been previously demonstrated. It is limited in principle, however, because it must, by the nature of the amount-recalled dependent variable and not by the processing framework itself, be unidimensional.

So far, the analysis of the 13 tasks of Table 2 has been in terms of amount recalled. That is, using the Subject \times Unit matrix for each task analogous to Table 1, only the bottom row has been used. A different view results if the same data for each task are analyzed in terms of how many subjects recalled each of the 125 words, that is, if the analysis is performed in terms of a measure for each task analogous to the right-hand column of Table 1. A simple correlation coefficient pro-

vides an adequate measure of similarity between the number, or percentage, of subjects recalling each word in a pair of tasks. In order to compare the complete set of 13 tasks listed in Table 2, 13 ordered sets of 125 numbers, one for each of the 13 tasks, were correlated, leading to a 13×13 symmetrical matrix of correlation coefficients. This matrix was in turn submitted to multidimensional scaling.

Here the multidimensional scaling program, ALSCAL (Young & Lewycky, 1979), was used with a simple euclidean nonmetric model. ALSCAL provided a one-, two-, three- and four-dimensional space in which the rank order of the closeness among all 13 tasks and the rank order of the correlation coefficients produced by those tasks were as close as possible. As no sharp improvement in the fit was noted above two dimensions moving from one to four dimensions ($r^2 = .90, .95, .97, \text{ and } .98$, respectively), and as the three- and four-dimension solutions presented no additional insights, the two-dimensional solution was chosen for convenience and is presented as Figure 1. Had the solution clearly been three or four dimensional, however, the techniques used here would have allowed this to be noted and displayed.

There appear to be two basic clusters in Figure 1 and two tasks that fail to cluster with other tasks. The largest cluster includes all of the free-recall tasks. Within the cluster, the free-recall tasks with the lowest levels of

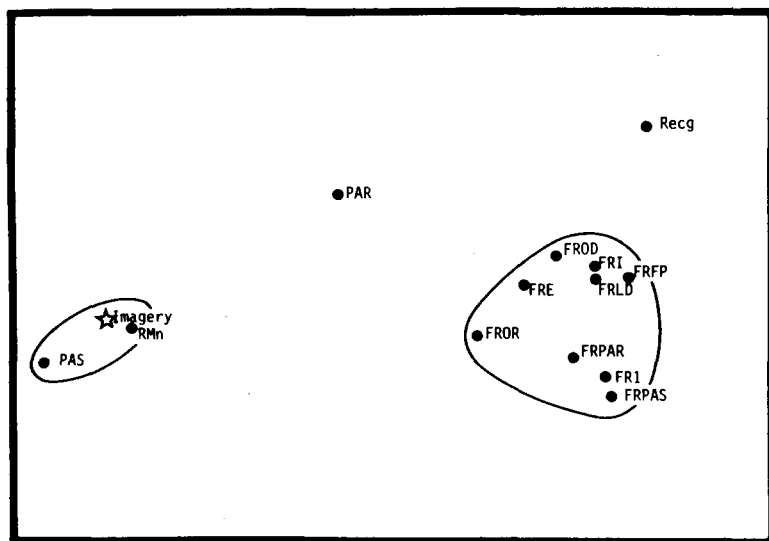


Figure 1. Multidimensional scaling solution for the list learning tasks shown in Table 2 and the independent variable of rated imagery. (Tasks close to each other have similar rank orders of words from most to least likely to be recalled. See Table 2 for definitions of abbreviations.)

recall appear to be at the top. The small cluster at the left contains two tasks known to depend on imagery (Bower, 1970; Paivio, 1968), paired-associate stimulus and recall using a mnemonic pathway. The third point, indicated by a star, is not one of the 13 tasks of Table 2 but rather is the imagery rating of the 125 words from Paivio et al. (1968). This point was included to demonstrate that the analysis used here allows independent, as well as dependent, variables to be included as a way of labeling the empirically derived space. Recognition and paired-associate response do not cluster with the other tasks as might be expected from the literature (G. Mandler, 1980; Paivio, 1968). Thus, the unit analysis of Figure 1 provides a reasonable, interpretable view of the data, but one that is quite different from that of the amount-recalled analysis of Table 3.

The Free-Recall Cluster

The free-recall cluster of Figure 1 deserves additional attention. First, the observation that different initial encodings lead to different amounts of recall, but still lead to the same items being recalled most often, presents complications for the classical depth of processing approach, but not for the processing

approach itself. Let us concentrate on the three free-recall tasks added specifically to test this possibility: surprise free recalls following ratings of orthographic distinctiveness, emotionality, and imagery. The shallow encoding task of rating orthographic distinctiveness leads to 57% of the recall level of the deep encoding tasks of rating emotionality or imagery. It follows, from the standard way in which the depth of processing hypothesis is made operational, that this difference in amount recalled indicates a difference in depth of processing. If, however, the depth of processing hypothesis is made operational in the way suggested here, it would be expected that different items would tend to be recalled if processing changes in kind and not just amount. This does not appear to be the case. The average correlation among the three tasks is .76 compared with an average reliability of .86. That is, if the correlations were corrected for attenuation, the average correlation among the three tasks would be .92. Thus, the two different ways to measure the depth of processing hypothesis lead to two different conclusions. This result is supported by the rest of the free-recall tasks used here as well as those reported by Hunt et al. (1979) and Postman and Kruesi (1977). In contrast, the processing approach would judge the similar-

ity of processing only on the basis of which units tended to be recalled. Differences in the overall amount of recall would indicate only that the overall amount of processing varied, not that the processing had changed in kind or in depth.

The finding that the free-recall tasks used here tend to lead to the same items being recalled needs some qualifications. If more extreme forms of processing manipulations were used, differences would probably be noted. For example, previous studies have shown that when people were given the task of rating the position of their lips while they vocalized words as their shallow task (Seamon & Murray, 1976), and when people were given CVC nonsense syllables of varying levels of meaningfulness instead of words (Jacoby, Bartz, & Evans, 1978), interactions between the type of processing and the units recalled were observed. Similarly, changes from oral to visual presentation lead to different words being recalled if the words are chosen to vary along orthographic distinctiveness (Hunt & Elliott, 1980). That is, the fact that a null hypothesis can be accepted for the free-recall tasks used here does not imply that it holds for all free-recall tasks. Rather, the fact that the null hypothesis holds for so many superficially different free-recall tasks that have quite different levels of recall, although not holding for other memory tasks, such as recognition, provides a good starting point for research. Filling in the free-recall tasks that fail to cluster with the others would be instructive, especially if their position on a plot such as Figure 1 could be interpreted in terms of other tasks or independent variables. The direction in which such research could be expanded is limited only by the patience of the investigator. Another qualification is that there are small differences in the relative probability of the recall of items among the free-recall tasks used here, and some of these differences would probably be shown to be nonrandom if experiments with enough power were used. Such small differences can be very important theoretically. The only reason they are being glossed over here is so that the general relations among tasks can be outlined. In fact, once sets of probabilities for recall tasks are established, small differences (i.e., residuals) are easier to observe.

Free recall following imagery rating was included, because it was expected to correlate with the other imagery tasks and not with the free-recall tasks. Its failure to do so may provide a clue as to why the intentional and incidental free-recall tasks used here cluster. The rating scale orienting tasks used here, as well as the intentional learning instructions, implicitly encourage the subject to make comparisons among all items in a list, thereby allowing associations among items to function as a major retrieval device in all tasks (Postman & Kruesi, 1977; Shaughnessy, 1979). Two of the depth of processing studies cited that showed different items being recalled under different conditions (Jacoby et al., 1978; Seamon & Murray, 1976) did not encourage comparison among items in the lists. This common aspect of the processing may overshadow the differences in processing. That is, the shallow and deep processing orienting tasks used here may share some key elements that lead to the same items being recalled. Comparisons among items in a list may be a sufficient condition for producing similar recall, though it appears not to be a necessary one (Postman & Kruesi, 1977). Further mapping, such as that started in Figure 1, would be needed to try to tease apart such component processes.

An alternative explanation was offered by Hunt et al. (1979).

Assuming that encoding involves matching the incoming information against a featural representation of that information . . . , the orienting task may control the activation of a subset of the features. Activation of this subset, however, may spread to other features, and to the extent that these activated features are not congruent with the orienting task, structural effects will appear independent of process. (p. 345)

Under this view, the particular words recalled with different orienting tasks will be the same to the extent that the processing invoked specifically by the orienting task does not add much to the processing that results from spread to other features (e.g., Nelson, Reed, & McEvoy, 1977).

A rating task that was assumed to involve imagery processing failed to affect high- and low-imagery words differently than did a non-imagery processing task. In contrast, two tasks that previously have been shown to involve imagery processing did have a differ-

ential effect. Paivio (1971) described a similar experience in his search for "the elusive interaction" (p. 366). After reporting failures to find an interaction between the processing induced by mediating tasks and the imagery rating of the stimuli, Paivio stated:

The most probable explanation of the failure remained the difficulty of establishing strong and persistent sets to use mediating devices that are incongruent with the associative processes most readily aroused by the nouns themselves. The problem was finally solved by a procedure that literally forced subjects to generate a particular kind of mediator. (p. 366)

Paivio's explanation is similar to that devised by Hunt et al. (1979) for a different set of tasks.

Independent of the explanation favored, what is clear for processing studies is that orienting tasks involving ratings are not as effective controllers of processing as might be expected. Having subjects recall items by using imagery on a mnemonic pathway constrains processing to imagery factors to a much greater extent than having subjects rate imagery.

Figure 1 is not presented as the ultimate answer to the way in which tasks are related when the unit analysis dependent variable is used. Rather, Figure 1 is an example of a way to proceed. Figure 1 represents one answer for a particular 125 words and a particular set of tasks, though other data indicate that the correlations obtained are representative (Rubin, 1980). By using Figure 1 as a starting point, tasks and independent variables can be added to begin answering some of the questions about processing that have been raised. Just as Figure 1 provides a partial tentative answer, so does the claim that all of the free-recall tasks presented here cluster and yet are different from other memory tasks. The null hypothesis that several tasks fail to differ in terms of which units are recalled is useful enough to deserve a name. The *fixed rank order hypothesis* holds that for a specified domain of tasks there is a single fixed rank order of units from most to least likely to be recalled. In Figure 1, one such domain exists for the free-recall tasks, and another different domain for the tasks that correlate with imagery. The fixed rank order hypothesis is seen as both a sophisticated null hypothesis from which deviations can be

noted and as an empirical statement about free recall which allows results from one particular condition to be generalized to many others.

Before providing a general discussion of the methods and results presented here, a second domain of verbal material will be considered.

A Unit Analysis of Text Recall

This section presents an analysis parallel in form and content to the one just given for list learning. Here the units are not words presented in random orders, but the 47 content units of the Lincoln story (Whipple, 1915), as analyzed by Rubin (1974, 1978). In the section following these analyses, results from numerous other prose studies in the literature will be reanalyzed and tabulated.

Table 4 presents the 13 memory tasks considered here. More details on these tasks and their relation to a dozen predictor variables can be found in Rubin's (1974, 1978) studies.

The same ALSICAL statistical procedures that led to Figure 1 were applied here. Given no marked improvement in the fit above two dimensions going from one to four dimensions ($r^2 = .75, .92, .96, \text{ and } .98$, respectively) and no insights from the higher dimensional solutions, a two-dimensional solution was again used and is presented as Figure 2. As with the list learning task, the free-recall memory tasks cluster and recognition appears as a separate task. The prompted-recall tasks included here, but not in the list learning study, also cluster. Within the free-recall and the prompted-recall clusters there appears to be a time or amount-recalled organization going from the immediate retention interval in the lower left to the 1-month retention interval in the upper right.

An independent variable, imagery, was added to Figure 1 to demonstrate how theoretically relevant predictor variables could be used to label the space obtained. The same technique could be used here. For instance, it was found that with the college free-recall task and the structured story used here, the recall task appeared near to independent variables such as rated meaning, whereas for a less structured story, the recall task appeared

Table 4
Text Recall Tasks

Name	Label	Description ^a	% recalled
College free recall	College FR	—	56%
Eight grade free recall	8th Gr FR	average age of 12	59%
Older men free recall	Older FR	average age of 68	45%
Incidental learning free recall	IL FR	recall was unexpected	43%
Immediate free recall	Immed FR	no retention interval, just turn page and recall	70%
Ten-minute free recall	10-Min FR	—	60%
One-week free recall	1-Wk FR	—	54%
One-month free recall	1-Mnth FR	—	43%
Immediate prompted recall	Immed PR	The four prompted recall tasks followed immediately after their respective free-recall tasks using the same subjects. The prompts were the story with all content words removed.	73%
Ten-minute prompted recall	10-Min PR		68%
One-week prompted recall	1-Wk PR		64%
One-month prompted recall	1-Mnth PR		55%
Recognition	Recog	Units had to be selected from a list containing an equal number of foils.	77%

^a All groups had 16 subjects and unless noted in the title had a 10-min filled retention interval. Different subjects took part in each task except that the subjects who took part in the free recall at one retention interval also took part in the prompted recall for the retention interval.

closer to independent variables such as serial position (Rubin, 1974). Similarly, if a text was divided into units using a system such as Kintsch and van Dijk's (1978), then theoretical recall values for each of the propositions could be obtained by fitting a few parameters, and these values could lead to points in the space in the same way that imagery did. Modifications in the recall theory could be attempted to see if it could also account for the recognition and prompted-recall points. This could be done without running additional experiments. Alternatively, parameter values in the existing theory could be changed to see if this resulted in the movement of theoretical points across the space of Figure 2 in ways interpretable by the theory. That is, although Figure 2 is presented here as a description of the relation among memory tasks, it could also function as a highly efficient way of testing theories. Once adequate fits are found, a plot using a new story and some new tasks could be used to provide confirmation.

The results of the text-recall analysis are remarkably like those of the list learning analysis, and the discussion of the list learning analysis could be repeated here. However,

unlike the list learning literature, there exists for text recall a multitude of studies that provide enough statistical analysis of their units to allow a direct comparison with the results shown in Figure 2. Therefore, before discussing Figures 1 and 2, additional studies will be considered.

A Reanalysis of Text-Recall Studies

Study Selection and Analysis

Included in the following summary are all studies that could be found that: (a) manipulated some experimental condition or used different subject populations; (b) included an analysis of the effects of the manipulation on the individual units or on groups of units; and (c) presented enough data so that a quantitative comparison with other studies was possible. Where appropriate, the memory tasks from Figure 2 are included. In comparing many studies, justice cannot be done to the special contributions of each. The many procedural differences among the studies makes it difficult to specify the causes of differences in results, but also makes it possible to generalize any observed similarities.

Up to this point correlational analyses

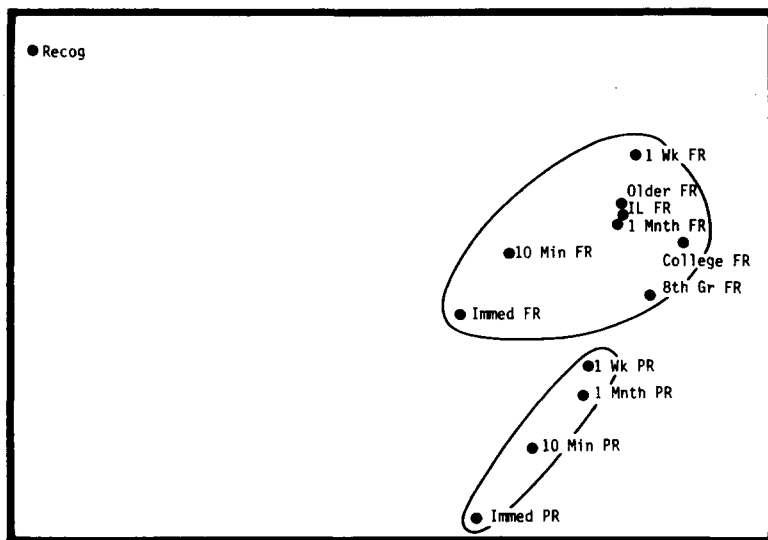


Figure 2. Multidimensional scaling solution for the prose memory tasks shown in Table 4. (Tasks close to each other have similar rank orders of units from most to least likely to be recalled. See Table 4 for definition of abbreviations.)

have been used exclusively. However, because the most common statistical technique used in experimental psychology is analysis of variance, both statistical procedures will be included here. For each analysis of variance considered, the amount of variance accounted for by a manipulation, by the units, and by their interaction will be reported. The units effect is always within subjects; the manipulation is, with a few exceptions, between subjects; and when the manipulation is between subjects, the interaction uses the same mean square error term as the units effect. As the analyses vary greatly in their degrees of freedom, the amount of variance accounted for by each effect provides an easier to understand and more appropriate metric of the size of effects than levels of significance (Hyde, 1981). In addition, the statistical significance of the effects is not of major importance as most effects reported in the literature are significant. Those that are not significant are noted.

Many different units have been used in the study of text recall. Those reported in Figure 2 were based on content words. Older techniques just divided text using slashes (e.g., Wechsler, 1945), whereas more recent techniques have used breath pauses (e.g., Johnson, 1970) as well as propositional systems (e.g.,

Kintsch, 1974). If the units are considered individually, the results are usually reported using correlations. When an analysis of variance is used, the individual units are usually collapsed into a few categories such as those based on rated importance (e.g., Johnson, 1970) level in a structural hierarchy (e.g., Kintsch, 1974) or a story grammar (e.g., J. M. Mandler, 1978; Mandler & Johnson, 1977). Here, all possible unit and category of unit schemes are included.

Grouping units into categories allows hypotheses concerning those categories to be tested directly using analysis of variance. However, it should be noted that grouping units imposes the theoretical disposition of the researcher on the data before they can be analyzed in a different way. For example, the results reported for a story grammar grouping may not hold, if the same units and data were analyzed using a structural hierarchy grouping. Thus, results determined from different groupings of units into categories are less general than results from studies which do not group units. By including many different types of groupings, the problem of generality will be minimized.

An example of the same data being analyzed by both correlational and analysis of variance techniques may make the relation

among the statistical techniques clearer. Consider the immediate, 10 minute, 1 week, and 1 month free-recall tasks from Figure 2. Correlations among the four retention intervals can be computed using the number of subjects recalling each of the 47 units as data. In the corresponding analysis of variance, there would be 4 levels of a retention interval main effect, and 47 levels of a units main effect.

The actual values obtained in this example provide a similar view of the relation among the four tasks, although they do so in slightly different ways. The average correlations among the four tasks is .84, whereas the average reliability is .89. Thus, most of the variance that can be accounted for in which particular items are recalled in one retention interval can be accounted for by knowing which items are recalled at another retention interval. In terms of the analysis of variance, the effect of retention interval, units, and their interaction are $F(3, 60) = 15.66$, $F(46, 2760) = 32.30$, and $F(138, 2760) = 1.67$, respectively (all $ps < .001$). The percentage of variances accounted for by the retention interval, units, and their interaction are 5%, 29%, and 2%, respectively. Thus, corresponding to the correlational conclusion, the units differ on their relative level of recall (i.e., there is a main effect of units), but this difference is not greatly affected by retention interval (i.e., the amount of variance accounted for by the interaction is relatively small). In addition, for the analysis of variance there is an effect of retention interval.

For studies in which the units themselves, and not categories of units, are considered, correlations are usually used partly because analyses of variance with 50 or 100 degrees of freedom in the numerator are too powerful to be interpreted easily (e.g., with 100, 200 degrees of freedom, an F of 1.32 is significant at the .05 level). However, when the units are collapsed into categories, analysis of variance provides a more complete report, yielding both main effects and interactions in one analysis.

Analysis of variance, however, can yield results different from correlations. For example, consider two conditions that lead to a perfect correlation and also lead to a zero interaction term in the analysis of variance.

If all the data in one condition are multiplied by a positive constant, then the correlation will remain perfect, but the interaction term will no longer be zero. This is because for the interaction term between two conditions to be zero, they can differ only by an additive constant, whereas for a correlation to be perfect, two conditions can differ by both an additive and multiplicative constant (i.e., a linear transformation). Therefore, there may be cases where a correlation approaches 1.00 and yet the interaction is significant. If the underlying conceptual measure is the rank order of units from most to least likely to be recalled, and the probability of recall for each unit can be changed by a multiplier as tasks change, then analysis of variance could overestimate deviations from the model that a correlation would not.

Tables 5, 6, 7, 8, and 9 contain studies that vary retention interval, the age of the subjects, aspects of the subject population, perspective or depth of processing at encoding, and other factors, respectively. Studies may appear in more than one table if they include more than one experiment or if an experiment manipulates more than one factor in addition to units. For each study, some general indication of the manipulation and the units used is given as well as any correlational or analysis of variance values that could be obtained. Each variance-accounted-for figure is the average of all individual experiments in the study cited. As most of the variance-accounted-for figures were not included in the original articles, most include some error introduced by my calculating values from the reported, rounded-off F and mean values, from estimated means obtained from figures, or from assumed values for incompletely reported nonsignificant F ratios. Where possible, the original analyses of variance provided by the researchers were used instead of published values.

Studies Varying the Retention Interval

Retention interval, according to Figure 2, does not play a large role in determining which units get recalled. Table 5 supports this finding. On the average, 21% of the variance in the results reported can be accounted for by retention interval, 17% by

Table 5
Retention Interval

Experiment	Condition	Units	\bar{r}	$\bar{\alpha}$	Variance accounted for ^a		
					Condition	Units	Interaction
Baggett (1979)	immed vs. 1 week	story grammar (3)			.11	.01	.00 ^b
Dixon, Simon, Nowak, and Hultsch (1982)	immed vs. 1 week	proposition level (4)			.06	.45	.00 ^b
Graesser, Woll, Kowalski, and Smith (1980)	30 min vs. 1 week	typicality (2)			.44	.09	.10
Johnson (1970)	15 min vs. 1-9 weeks	breath pauses	.74				
	immed vs. 3 weeks	breath pauses	.65				
	immed vs. 3 weeks	breath pauses	.76				
		importance rating (6)			.38	.24	.00
Johnson (1974)	immed vs. 1 week	breath pauses	.82				
Kintsch, Kozminsky, Streby, McKeon, and Keenan (1975)	immed vs. 1 day	proposition level (3)			.13	.07	.01
Kintsch and van Dijk (1978)	immed—3 month: data	propositions	.90				
	immed—3 month: theory	propositions	.99				
Meyer (1977)	immed vs. 1 week	propositions	.81				
Rubin (1978)	immed—1 month	content words	.84	.89	.05	.29	.02
Singer (1982)	immed—6 weeks	micro-macro			.24	.12	.04
Smith and Graesser (1981)	30 min—3 weeks	typicality (2)			.11	.09	.03
Stein and Glenn (1979)	immed vs. 1 week	propositions	.97 ^c				
Average			.84		.21	.17	.02

Notes. Immed = immediate. ^a Where possible ω^2 was used. For three or more factor designs, η^2 was usually used. ^b This effect was not significant at the .05 level. ^c This is a rank order correlation.

units, and only 2% by their interactions. As this finding is of special theoretical importance, it provides an opportunity to pursue in depth the kind of arguments that can be made from Figures 1 and 2 and Tables 5 through 9.

It is commonly claimed that, with the passage of time, the gist of a story is recalled and the details are forgotten (e.g., Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975; Kintsch & van Dijk, 1978; Meyer, 1975; Miller, Perry, & Cunningham, 1977). This claim suggests that important and unimportant items have different decay rates and has been the basis for considerable theoretical speculation. For example, Smith and Graesser (1981) contrasted schema models that predicted equal decay rates for typical and atypical information from schema models that did not predict equal decay rates.

Consider a simple alternative, the fixed rank order hypothesis, which is consistent with Figure 2 and Table 5. Assume that if an immediate-recall test were given, some units would be more likely to be recalled than others; that is, under some specified conditions at immediate recall there would be a rank ordering of units from most to least likely to be recalled. Further, assume that as time passes this rank ordering remains the same, the only change being in how far down this rank order most subjects recall units. In other words, assume a strength model of the simplest kind. Such a model would be consistent with Underwood's (1966) claim that the initial degree of learning is the only factor that affects retention. If the model worked, it would certainly be simpler than models based on differential decay rates because in either case, some specific theory is needed to predict what makes units differentially recallable at immediate recall.

What predictions would be made for Figure 2 and Table 5, assuming the fixed rank order hypothesis? Depending on the distribution of where on the rank order various subjects stop recalling units in the different retention intervals, almost any outcome is possible including decay being relatively uniform at all units, or decay being higher at either the high or low probability of recall units. However, if floor and ceiling effects are minimized under most distributions of subjects' positions on

the rank order, it can reasonably be expected that correlations among various retention intervals would be high, and interactions between units and retention factors in an analysis of variance would be low. Such is the case in most of the studies reported here.

It should be noted, however, that the fixed rank order hypothesis technically does not contradict the differential decay rate hypothesis. The fixed rank order hypothesis is just a simpler and more general statement of the same observation. In fact, if the fixed rank order hypothesis holds, any variable that correlates with the rank order at immediate recall (or, for that matter, with recall at any retention interval) must produce "differential decay" rates. The clearest example is Kintsch and van Dijk's (1978) study. They fit their data with a complex model. The three parameters that are free to vary with retention interval take on different values at the immediate, 1-month, and 3-month delay. As shown in Table 5, using Kintsch and van Dijk's (1978) own published values for the parameters and the recall levels for each unit, the average correlation among their theoretical values for their three times intervals is .99. Thus, a complex model shows statistically significant differential loss as measured by three parameters and, at the same time, is consistent with the fixed rank order hypothesis.

The paradox can be resolved as follows. One of Kintsch and van Dijk's three parameters is used only in important propositions (i.e., macropropositions), whereas the other two are used mostly in less important propositions (i.e., micropropositions). As Kintsch and van Dijk reported, the parameter that is used with important propositions changed very little with retention interval, whereas the other two parameters changed by approximately a factor of 5. Thus Kintsch and van Dijk's theory specifies which propositions are important. These important propositions were not forgotten over the retention intervals used, whereas the less important propositions were. That is, they found that the fixed rank order hypothesis held for their data ($r = .90$) and even more so for their theory ($r = .99$).

The lack of a substantial Unit \times Retention interval interaction in the Kintsch and van Dijk data and theory, although technically

consistent with their analysis, poses a serious problem for the spirit of their retention interval argument. If you were asked to predict which n propositions a particular subject recalled in Kintsch and van Dijk's study, the original analysis would lead you to consider the subject's retention interval before making your prediction. The reanalysis presented here indicates that, if you were using Kintsch and van Dijk's theory to make predictions, information about the subject's retention interval would be of no practical value. The change in parameters that Kintsch and van Dijk observed had to occur, given the fixed rank order hypothesis and the success of Kintsch and van Dijk's theory at any one retention interval.

The contribution of Kintsch and van Dijk's and others' theories should not be minimized by these comments. The fixed rank order hypothesis does not specify what the rank will be at any one time; the hypothesis states only that the rank order will not change. Kintsch and van Dijk, however, specified the rank order. All that is being argued here is that their theory makes changes in recall with retention more complex than they need be, and that it draws too much attention to differential decay rates, when these differential rates appear to be mostly an artifact of how far down the rank order a single decay rate causes subjects to be at different retention intervals.

In Table 5, there are two entries from the same laboratory that contradict the fixed rank order hypothesis: Graesser, Woll, Kowalski, and Smith (1980) and Smith and Graesser (1981). The interaction between retention interval and units accounts for 10% of the variance in the Graesser et al. study and 3% of the variance in the Smith and Graesser study. The only other study with over 2% of the variance accounted for by the interaction term (Singer, 1982) has a floor effect that accounts for much of its interaction. The stories in the Graesser et al. and Smith and Graesser studies contained randomly selected scripted activities (e.g., going to a restaurant), and they noted the recall of typical (e.g., Jack sat down at the table) and atypical (e.g., Jack put a pen in his pocket) actions within the script. In a recognition experiment, people knew whether or not they

had seen an atypical action, but could not choose correctly between typical actions they had or had not seen. If the correct recalls introduced by guessing were subtracted from the recall scores in the Graesser et al. recall experiment analyzed here (which was not done in Table 5 to allow comparisons to be made with other studies), there would be more atypical actions recalled at the 1/2-hr retention interval and more typical actions recalled at the 1-week interval. Thus, the fixed rank order hypothesis appears not to hold for passages that present a haphazard sample of subjects' highly organized world knowledge. Given the range of studies over which the fixed rank order hypothesis appears to hold, finding principled exceptions becomes more interesting. That is, if figures like Figure 2 are the rule for most types of text, then finding specific classes of material for which different patterns arise becomes quite informative. When there exists knowledge that can guide retrieval independent of a passage itself, the interaction of that knowledge and the passage rather than just the passage itself, guides recall. In the Graesser et al. and Smith and Graesser examples, it appears that the weighting of the two sources of information changes with time, as might be suspected from Jost's law (Simon, 1966), violating the fixed rank order hypothesis.

What can we say about retention given the data presented in Figure 2 and Table 5? Retention, as opposed to encoding or retrieval, appears to be a relatively passive process. The same units that tend to be recalled most often immediately tend to be recalled most often at later times. It is difficult to find a situation in which one unit is recalled more often than a second at one time delay and less often at another time delay. The major exception to the fixed rank order hypothesis comes from a situation in which world knowledge and the particulars of the text to be recalled change their relative importance to recall over time. For the most part, the processes that determine what will be remembered and what will be forgotten of a prose passage must be determined by encoding, retrieval, and/or their interaction, not by the length of the retention interval.

The retention-interval data should make the relation between Figures 1 and 2 and

Tables 5 through 9 clearer. Tasks that cluster in Figures 1 and 2 produce small Unit \times Task interactions in the tables. Tasks that do not cluster in Figures 1 and 2 when included in the same experiment yield large Unit \times Task interactions. That is, when the fixed rank order hypothesis holds, there should be considerable similarity in multidimensional scaling and negligible interactions in analysis of variance. Figures 1 and 2 and Tables 5 through 9 are thus two ways of looking at the same phenomenon. Inferential statistics, which is central to the analysis of variance but not to multidimensional scaling, forces a distinction to be made between the figures and the tables, a distinction which is not desirable for our present purposes. Similarity is basically a description of the data which is continuous; tasks are more or less related. Analysis of variance is basically a description of the data which is dichotomous; tasks differ significantly or they do not. The use of the amount-of-variance-accounted-for statistics in Tables 5 through 9 is an attempt to make analysis of variance more like the descriptive correlations on which the figures are based.

Studies Varying the Age of the Subjects

The argument just given for the effects of the retention interval can be repeated in even stronger terms for development across the life span. In Figure 2 and in Table 6 there is little evidence that the age of the subject from first grade to age 80 has an effect on which units are recalled, though there are the normal differences in how many units are recalled by the various groups. That is, the fixed rank order hypothesis can be extended to development. In a way this is a disappointment as varying age has been a useful technique in finding out about cognition. Here, however, from the time the subjects are old enough to be tested easily with a given text, that text no longer shows age-related changes in which units of the text are recalled. Only half of the dozen studies that provided analyses of variance for Table 6 had a significant Age \times Unit interaction. This is noteworthy. First, the studies were done by people interested in a field in which interactions with age are a standard basis from which to make inferences. Second, considerable power existed to note

such interactions. All of the unit main effects, which use the same error terms as the interactions, were significant at the .001 level (median $F = 87.52$, range = 16.50 to 794.44).

Studies Varying the Subject Population

Tables 5 and 6 support both Figure 2 and the fixed rank order hypothesis. Table 7 contains manipulations not represented in Figure 2 and initially seems to offer a domain for which the fixed rank order hypothesis fails to be a good approximation. However, most of the variance accounted for by the interaction terms is due to floor effects, and more careful correlational and Guttman scaling analyses in some of the studies yielding interaction terms (e.g., Bacon & Rubin, 1983; Rubin, Olson, Richter, & Butters, 1981) support the fixed rank order hypothesis.

Consider the Spilich (1983) study, which produces the largest variance-accounted-for value for an interaction. Spilich compares three groups of subjects: young-normal, elderly-normal, and elderly-impaired. The young-normal group of four propositional level categories remembered approximately 62, 27, 20, and 17% of the units. The elderly-normal group remembered 52, 21, 7, and 7%. Thus, for these two groups there is little interaction between the group and propositional level factors, and the interaction that does exist need not be an exception to the fixed rank order effect. The third group, the elderly-impaired, recalled approximately 5, 2, 1, and 2% of the units at the four propositional level categories. The curve of the elderly-impaired is not parallel to the other two groups' curves, leading to the large interaction in Table 7. However, it is not clear whether this interaction is due to a violation of the fixed rank order hypothesis or merely to a floor effect. Similar problems exist in varying degrees with the other groups presented in Table 7 that have substantial memory deficits.

Studies Varying the Subjects' Perspective or Depth of Processing at Encoding

Although changes in retention interval, in subject age, and in various other subject population variables do not lead to violations of the fixed rank order hypothesis, Table 8

(text continued on page 232)

Table 6
Age of Subjects

Experiment	Condition	Units	\bar{r}	$\bar{\alpha}$	Variance accounted for ^a		
					Condition	Units	Interaction
Brown, Smiley, Day, Townsend, and Lawton (1977)	2nd-7th grade	importance rating (4)			.02	.62	.01
Brown and Smiley (1977)	8-18 years	importance rating (4)			.04	.51	<.01 ^b
Dixon, Simon, Nowak, and Hultsch (1982)	18-81 years	proposition level (4)			.06	.45	.00 ^b
Dixon, Hultsch, Simon, and von Eye (1984)	32-68 years	proposition level (4)			.08	.32	.00 ^b
Dunn, Mathews, and Biegar (1979)	4th-6th grade	proposition level (5-7)			.05	.14	.02
J. M. Mandler (1978)	2nd grade-college	story grammar (6)			.05	.38	.02 ^b
Mandler and Johnson (1977)	1st-college	story grammar (6)			.28	.15	.02
Meyer and Rice (1981)	23-68 years	proposition level (3)			<.04	<.33	<.04
Meyer and Rice (1983)	25-69 years	proposition level (2)			.01 ^b	.06	.00 ^b
Petros, Tabor, Cooney, and Chabot (1983)	24-72 years	importance rating (3-4)			.04	.41	.00
Rubin (1978)	13-68 years	content words	.88	.90	.02	.33	.02
Simon, Dixon, Nowak, and Hultsch (1982)	23-67 years	proposition level (4)			.02	.15	.00 ^b
Stein and Glenn (1979)	1st-5th grade	propositions	.94 ^c				
Average			.91		.06	.32	.01

^a Where possible ω^2 was used. For three or more factor designs, η^2 was usually used. ^b This effect was not significant at the .05 level. ^c This is a rank order correlation.

Table 7
Subject Populations

Experiment	Condition	Units	\bar{r}	$\bar{\alpha}$	Variance accounted for ^a		
					Condition	Units	Interaction
Bacon and Rubin (1983)	retarded and M.A. matched controls	phrases	.74	.81	.00 ^b	.16	.02
Dixon, Hultsch, Simon, and von Eye (1984)	verbal ability in normal range	proposition level (4)			.04	.32	.00 ^b
Dunn, Mathews, and Bieger (1979)	average and below average reading ability	proposition level (5-7)			.06	.14	.01
Graybeal (1981)	language impaired and controls	story grammar (6)			.06	.05	.02 ^b
Luftig and Johnson (1982)	retarded and M.A. matched controls with 5 orienting tasks	importance rating (3)			.17	.19	.00 ^b
Mandler, Scribner, Cole, and DeForest (1980)	literacy, schooling, culture	story grammar (6)			.12	.28	.01 ^b
Petros, Tabor, Cooney, and Chabot (1983)	12 vs. 18 years of school	importance rating (3-4)			.04	.41	.00 ^b
Rubin, Olson, Richter, and Butters (1981)	Korsakoff and controls	content words	.72	.86	.12	.24	.05
Smiley, Oakley, Worthen, Campione, and Brown (1977)	schizophrenics and controls	content words	.81	.89	.14	.22	.04
Spilich (1983)	poor and good readers	importance rating (4)			.45	.20	.03
Average	memory impaired elderly, and young adults	proposition level (4)	.76	.86	.11	.21	.03

^a Where possible ω^2 was used. For three or more factor designs, η^2 was usually used. ^b This effect was not significant at the .05 level.

Table 8
Change in Perspective or Depth of Processing at Encoding

Experiment	Condition	Units	\bar{r}	$\bar{\alpha}$	Variance accounted for ^a		
					Condition	Units	Interaction
Brown, Smiley, Day, Townsend, and Lawton (1977)	orienting paragraph	importance rating (4)			.03	.62	<.01 ^b
Einstein, McDaniel, Bowers, and Stevens (1984)	orienting title	importance rating (3)			.12	.31	.02
Grabe (1979)	proposition specific vs. relational processing				.00 ^b	.31	.02
Grabe (1981)	read vs. fill in letters				.06	.08	.01
Meyer (1977)	augmented perspective change in perspective signaling or emphasis of same propositions	importance rating (2) importance rating (2) propositions	.89		.00 ^b	.00 ^b	.05
	same structure-different content:				.00 ^b	.04	.06
	for the whole passage for a single paragraph high in importance for that paragraph but low in importance		.55 .76				
Meyer and Rice (1983)	emphasis: story 1 emphasis: story 2	proposition level (2)			.03	.10	.01
Rubin (1978)	orienting title	content words	.80	.89	.03	.03	.04
Schultz (1982)	depth of processing line by line blocked	rated importance (3)			.04	.29	.03
Simon, Dixon, Nowak, and Hultsch (1982)	depth of processing	proposition level (4)			.00 ^b	.22	.00 ^b
Average					.10	.14	.03
					.03	.15	.00 ^b
					.04	.19	.02

^a Where possible ω^2 was used. For three or more factor designs, η^2 was used. ^b This effect was not significant at the .05 level.

offers one domain in which careful manipulations provide clear violations of the fixed rank order hypothesis. For instance, in one of Meyer and Rice's (1983) stories, and in the Grabe studies (1979, 1981), more variance is accounted for by the interaction term than by either of the main effects related to it. Moreover, propositions that are recalled frequently under one perspective are recalled infrequently under another, directly violating the fixed rank order hypothesis. A person reading about a house from the perspective of a home buyer just remembers different things than a person reading the same passage from the perspective of a burglar (Grabe, 1979, 1981; Pichert & Anderson, 1977). Although themselves providing values that can be added to Table 8, studies in which a passage is made coherent by the addition of a title (Bransford & Johnson, 1972), and studies which present a passage in a normal of randomized order (Thorndyke, 1977), surely would also lead to violations of the fixed rank order hypothesis.

Other more subtle manipulations of the reader's perspective during the processing of a passage also show clear violations of the fixed rank order hypothesis. In a clever manipulation, Meyer (1975, 1977) demonstrated that placement of a paragraph in the overall structure of a larger segment of text is very important to what is remembered. Meyer constructed two paragraphs that had the same organizational structure, but different content. When the paragraphs were placed high in the organizational structure of a larger passage, the structure of the individual paragraph largely determined what was recalled. However, when the paragraph was placed low in the organizational structure of a larger passage, the structure of the individual paragraph did not determine what was recalled. That is, which units were recalled changed as a function of the paragraph's position within a larger passage.

Standard depth of processing manipulations, similar to the ones used in the studies with lists cited earlier, were employed with prose. In certain, but not all, conditions the depth of processing shallow-deep manipulation appears to affect which items are recalled (Schultz, 1982; Simon, Dixon, Nowak, & Hultsch, 1982). Thus, the fixed rank order

hypothesis often does not hold when the initial encoding of a text is manipulated.

Studies Varying Other Factors

Table 9 presents several manipulations that fail to challenge the fixed rank order hypothesis. Manipulating the rate and mode of presentation or interleaving episodes of the story does not have much effect on which units are recalled. As with list learning, the overall motivational level and the intention on the part of the subject to learn also have little effect on which units are recalled.

Changing the retrieval task, for example, from free recall to prompted recall or recognition is a manipulation that changes which units are remembered, as shown in Figures 1 and 2. However, this manipulation is not represented in Table 9 except for the data already presented in Figure 2. Although it is easy to find quantitative support for this idea in the literature (e.g., Graesser et al., 1980), experimenters typically present different retrieval tasks as different experiments and, therefore, analyses of variance including retrieval tasks as a factor are not reported. Although it is clear that the fixed rank order hypothesis does not hold over changes in retrieval tasks, more work is needed varying the kinds of prompts used in cued recall and the types of foils used in recognition before it will be known whether the fixed rank order hypothesis holds within specified ranges of prompted recall or recognition tasks.

General Discussion

Two major unit analyses were presented, one for isolated words and one for prose. The implications of these analyses will be considered first as methodological techniques and, second, in terms of the actual results obtained.

Methodological Advantages of Unit Analyses

The most important methodological point to note is that interesting, interpretable results were obtained with the minor change in the dependent variable made here. At times these results lead to views that are in conflict with views that follow from more conventional analyses of the same data. Little more could be asked of a method. This change occurs

Table 9
Other Variations

Experiment	Condition	Units	\bar{r}	α	Variance accounted for ^a		
					Condition	Units	Interaction
Baggett (1979)	movie vs. text	story grammar (3)			.04	.01	.00 ^b
Johnson (1970)	paced vs. nonpaced presentation	breath pauses	.91		.01	.38	<.01 ^b
J. M. Mandler (1978)	interleaving of episodes	story grammar (6)			.13	.34	.02
Masson (1982)	rate of presentation	proposition level (2)			.01	.32	.00 ^b
Petros, Tabor, Cooney, and Chabot (1983)	rate of presentation	importance rating (3)					
Rubin (1978)	incidental-intentional	content words	.82	.88	.04	.28	.03
	motivational level		.80	.81	.04	.31	.02 ^b
	recall-recognition		.50	.84	.06	.18	.11
	free-prompted recall		.71	.89	.01	.28	.03
Schmidt (1983)	expecting recall vs. expecting recognition	importance rating (3)			.03	.08	.00 ^b

^a Where possible ω^2 was used. For three or more factor designs, η^2 was used. ^b This effect was not significant at the .05 level.

from asking which rather than how many units are remembered. That is, the change occurs from asking a quantitative question that is closer to descriptive studies of memory.

In terms of efficiency, the analyses performed here have several advantages over the more conventional analysis of variance approach. If, for a theoretical reason or just as an exploratory hunch, an experimenter wanted to add another recall condition to either Figure 1 or 2, only that condition would need to be run; the existing values could still be used. Similarly, more independent variables could be added, or, as discussed in the prose section, modifications of existing theories could be given an initial test. Because it is correlational, the unit analysis, does not require the existing data to be modified when a new measure is added and allows that new measure to be compared with all other measures in the existing data. This efficiency would make it possible to test the many different kinds of manipulations needed to begin to explore memory fully (e.g., Jenkins, 1979).

Therefore, the unit analysis is ideally suited for exploratory work. If a new theory is constructed, it can be given an initial quantitative test on an existing data set collected years earlier as long as the theory is capable of assigning a value to each unit. Not only would one get an idea of whether the new theory worked, one could also determine how well it worked in comparison to existing independent variables, which units it predicted best, and to which existing independent variables it is most and least similar. This is an impressive collection of information to get without running a new experiment. If the theory or hunch proved successful, new materials could be prepared for a unit analysis in which the theory could be better contrasted with its near competitors.

Four Technical Points

Before leaving methodological consideration, four technical points should be made. First, the units of a unit analysis, unlike the subjects of an analogous individual differences study, are not independent. Therefore, the use of inferential statistics becomes more difficult. Several alternative analyses, such as

Guttman scaling (see Rubin, 1977, for an application) or repeated measures analysis of variance as used in Tables 5 through 9, can supplement the correlational analyses, where it is important to reject a null hypothesis rather than describe the magnitude of an effect. This point becomes less important in the work done here, because the correlations obtained have been sufficiently large and lawful to negate the possibility that they occurred by chance.

The second technical point concerns the relation of the theoretical view of recall and its measurement. For a particular set of conditions, recall is considered to be the result of a strength continuum. For an individual, recall strength is measured by a dichotomous scale: The unit is either recalled or not. For a group, the number-of-subjects-recalling-each-unit measure is a more continuous, but not necessarily a linear, function of the underlying strength dimension. This is because not only the underlying strength, but also the distribution of the amounts recalled by the subjects, determine the shape of the number-of-subjects-recalling-each-unit function. Therefore, a rank order measure is the strongest measure that should have been used in this manuscript. To be consistent with and to make use of existing correlations and analyses of variance, interval scale measurements were used throughout this article, whereas only ordinal level measurements were indicated by theoretical considerations. For the most part, this resulted in no practical difference; however, there are many hypothetical cases and a few actual cases where there are noticeable differences. For example, several studies were noted where floor effects caused large interactions between units and a condition, even though it was possible that the same rank ordering of units existed in all conditions. It follows that using correlations and analyses of variance to test the fixed rank order hypothesis is a conservative strategy.

The third technical point that should be stressed is that the fixed rank order hypothesis is an empirical regularity that probably represents no more than a good first approximation. For example, if 1,000 subjects were tested in each condition shown in Figures 1 and 2 raising the reliabilities to over .99, there is little doubt that the correlations among the tasks would not all also be over .99. Although

the tasks would still be very similar, small but clear differences would probably be found. Similarly, in the analyses of variance of Tables 5 through 9, statistically significant interactions accounting for little variance were often noted. As a first approximation, the fixed rank order hypothesis allows us to generalize across situations, to make practical applications, and to develop theories that will not put the main burden of predicting which units are recalled on factors that have been shown to have little effect on that measure. As a sophisticated null hypothesis, the fixed rank order hypothesis allows deviations from the hypothesis to be seen more easily and to have greater theoretical weight. Deviations can be seen more easily by explicitly using the fixed rank order hypothesis as a predictor and noting deviations, or residuals, from it. Given the evidence collected in support of the fixed rank order hypothesis, specifiable, repeatable deviations become the next step in refining the data base and the theories it can support.

The fourth technical point concerns the assumptions that small Unit \times Manipulation interactions and high correlations between tasks imply similarity of processing. In the processing approach, a large Unit \times Manipulation interaction or a clear violation of the fixed rank order hypothesis demonstrates that different processes are operating, but what should be made of the consistent small interactions and high correlations? Their size cannot be attributed to chance. Some small interactions and high correlations could be due to a coincidental balancing out of competing processes, but this interpretation cannot account for the numerous, orderly cases of small interactions and high correlations found with different materials from different laboratories. Some other small interactions could be due to a lack of statistical power, but not in the cases where there are strong main effects or high reliabilities. The small interactions and high correlations are a reliable finding. To the extent that memorability is a measure of processing, the small interactions and high correlations must indicate similarity in processing.

Conclusions

Several substantive results emerged from the unit analyses performed here. Over a wide range of tasks, the rank order of units

from most to least likely to be recalled was relatively fixed over variation in conditions. Changes in the amount of units recalled without changes in the kind of units recalled were reported in lists, in prose, or in both for variation in depth of processing, retention interval, subjects' age, subjects' clinical population, and the incidental-intentional learning distinction. Each of these findings could have been given more individual attention. For example, though it was simply stated as an empirical regularity, deciding whether incidental and intentional learning were basically the same or not once involved a considerable research effort and much theoretical debate (e.g., Postman, 1964; Wolk, 1974).

Where does this fixed rank order hypothesis leave us? If we take the view of science we are often led to by psychology's emphasis on inferential statistics, then this article uncovers a massive failure in our enterprise. Large significant differences in which units were recalled were rarely found. Under this view, Tables 5 through 9 are nothing more than evidence of too many replications of the same dull finding. If we take the more widespread view that science is about uncovering, describing, and understanding regularities, then this article outlines an area where advances have and can be made. For a diverse set of materials, procedures, and subjects, a regularity has been uncovered and has been tentatively described. The rank order of units from most to least likely to be recalled from a list of words, either in random or in textual form, is the same under many conditions and fails to be the same under other clearly defined conditions.

Lacking in this article is a clear understanding of how or why this regularity occurs. Serious beginnings of such an understanding, however, do exist. For instance, Kintsch and van Dijk (1978) offer a quantitatively successful model of why particular units of text are remembered or forgotten. Not as much formal theoretical work has been undertaken with lists, but empirically it appears that imagery and associative frequency are the major factors in predicting which words are recalled (Rubin, 1980, 1983). Still lacking, however, is a formal theory of why the rank order is stable over some changes in conditions, but not others. Even here our existing knowledge of human memory provides some

idea of how to formulate such a theory based on differences in retrieval and encoding processes. The question of why the rank order would change or remain stable, however, had been premature until the regularity was described.

Thus, rather than summarizing a massive failure, the fixed rank order hypothesis organizes our observations and points to the need for more theory development and testing. A theory that can predict the rank ordering of units under one set of recall conditions becomes more important, because it will predict over many sets of conditions. A theory that can predict when the rank ordering will change can be begun, because we know something of when such changes occur.

Jacoby et al. (1978) summarize the present state of affairs.

We feel that the time is past when simple demonstrations of effects of levels of processing and effects of context are really helpful. What is now needed is some means of incorporating those effects into a larger theory. It seems to us that a reasonable first step toward developing such a theory is to search for regularity in the functions relating task variables to retention performances and to specify how those functions change across situations. If we can devise a classification of situations in terms of the form of functional relationships they produce, we will then be in a better position to speculate about the nature of the beast that can give rise to such diversity. (p. 345)

With luck, the fixed rank order hypothesis and its failures (as expressed in Figures 1 and 2 and Tables 5 through 9) are a step in the direction to which Jacoby et al. point.

Depth of processing has been a major research interest in cognitive psychology over the last decade. Depth of processing's basic processing claim that the type of processing occurring at encoding affects later recall was taken as an assumption here. Instead of pursuing this claim solely in terms of amount recalled, a unit analysis was also conducted. If the standard amount-recalled measure is used, the standard shallow-deep dimension occurs. If the unit analysis measure is used, a different picture (Figure 1) occurs in which free-recall tasks are quite similar regardless of their depth of encoding. It is disconcerting to have one basic approach and one set of data lead to two different conclusions, when different dependent measures are used. If a *how many* question is asked, depth, as opposed to kind, of processing is important for

the manipulations used here. If a *which one* question is asked, it typically is not. The results of both the standard amount-recalled and the unit analyses, however, are consistent with the following view of processing (cf. Hunt & Elliott, 1980; Jacoby et al., 1978; Postman & Kruesi, 1977). Over a wide range of encoding tasks, the encoding task mainly affects the total amount of processing allocated with the individual units determining the relative allocation of that processing. Thus, standard depth of processing manipulations have more effect on overall task measures, such as amount recalled, than on the detailed unit analysis. This view offers a different picture of processing than studies based solely on the amount-recalled dependent measure, and the supporting data collected here offer a more general measure of where differences in processing occur than previous studies that manipulated only one property of words, such as meaningfulness or orthographic distinctiveness.

One of the most interesting results of the analyses performed here was the similarity among the findings of the list learning and the prose analyses. Wherever comparable data existed, the same manipulation either led to or failed to lead to a change in both domains. When psychologists recently moved their emphasis of study from lists of isolated words to lists of words organized into text, they tended to ignore their earlier out-of-fashion efforts (e.g., Jenkins, 1974). Extending studies of list learning directly into studies of prose is not inconsistent with the original goal of verbal learning (e.g., Ebbinghaus, 1885/1964) and was, in fact, the research strategy used by some psychologists (Musgrave & Cohen, 1971). The similarity obtained here between the two domains, as well as the prose research itself (Rubin, 1978), suggests that such a strategy may not have been misguided. People are remembering verbal material in both domains; increasing the level of organization as one goes from lists to prose need not change all of the processes involved. Research in cognitive psychology may be more cumulative than we realize.

References

Bacon, E. H., & Rubin, D. C. (1983). Story recall in mentally retarded children. *Psychological Reports, 53*, 791-796.

- Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning and Verbal Behavior, 18*, 333-356.
- Bartlett, F. C. (1967). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press. (Original work published 1932)
- Bower, G. H. (1970). Analysis of a mnemonic device. *American Scientist, 58*, 496-510.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.
- Brown, A. L., & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem in metacognitive development. *Child Development, 48*, 1-8.
- Brown, A. L., Smiley, S. S., Day, J. D., Townsend, M. A. R., & Lawton, S. C. (1977). Intrusion of a thematic idea in children's comprehension and retention of stories. *Child Development, 48*, 1454-1466.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671-684.
- Dixon, R. A., Hultsch, D. F., Simon, E. W., & von Eye, A. (1984). Verbal ability and text structure effects on adult age differences in text recall. *Journal of Verbal Learning and Verbal Behavior, 23*, 569-578.
- Dixon, R. A., Simon, E. W., Nowak, C. A., & Hultsch, D. F. (1982). Text recall in adulthood as a function of level of information, input modality, and delay interval. *Journal of Gerontology, 37*, 358-364.
- Dunn, B. R., Mathews, S. R., & Bieger, G. (1979). *Individual differences in the recall of lower-level textual information* (Tech. Rep. No. 150). Urbana-Champaign: University of Illinois, Center for the Study of Reading.
- Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885)
- Einstein, G. O., McDaniel, M. A., Bowers, C. A., & Stevens, D. T. (1984). Memory for prose: The influence of relational and proposition-specific processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 133-143.
- Freud, S. (1951). *Psychopathology of everyday life* (A. A. Brill, Trans.). New York: Mentor. (Original work published 1904)
- Grabe, M. D. (1979). Reader imposed structure and prose retention. *Contemporary Educational Psychology, 4*, 162-171.
- Grabe, M. D. (1981). Variable inspection time as an indicator of cognitive reading behaviors. *Contemporary Educational Psychology, 6*, 334-343.

- Graesser, A. C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 503-515.
- Graybeal, C. M. (1981). Memory for stories in language-impaired children. *Applied Psycholinguistics*, 2, 269-283.
- Hunt, R. R., & Elliott, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects in retention. *Journal of Experimental Psychology: General*, 109, 49-74.
- Hunt, R. R., Elliott, J. M., & Spence, M. J. (1979). Independent effects of process and structure on encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 339-347.
- Hyde, J. S. (1981). How large are cognitive gender differences? A metaanalysis using ω^2 and d . *American Psychologist*, 36, 892-901.
- Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, 82, 472-481.
- Jacoby, L. L., Bartz, W. H., & Evans, J. D. (1978). A functional approach to levels of processing. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 331-346.
- Jenkins, J. J. (1974). Remember that old theory of memory? Well, forget it! *American Psychologist*, 29, 785-795.
- Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 429-446). Hillsdale, NJ: Erlbaum.
- Johnson, R. E. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior*, 9, 12-20.
- Johnson, R. E. (1974). Learners' predictions of the recallability of prose. *Journal of Reading Behavior*, 6, 41-52.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*, 14, 196-214.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Luftig, R. L., & Johnson, R. E. (1982). Identification and recall of structurally important units in prose by mentally retarded learners. *American Journal of Mental Deficiency*, 86, 495-502.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252-271.
- Mandler, J. M. (1978). A code in the node: The use of a story schema in retrieval. *Discourse Processes*, 1, 14-35.
- Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111-151.
- Mandler, J. M., Scribner, S., Cole, M., & DeForest, M. (1980). Cross-cultural invariance in story recall. *Child Development*, 51, 19-26.
- Masson, M. E. J. (1982). Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 400-417.
- Meacham, J. A. (1972). The development of memory abilities in the individual and society. *Human Development*, 15, 205-228.
- Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Amsterdam: North-Holland.
- Meyer, B. J. F. (1977). What is remembered from prose: A function of passage structure. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 307-333). Norwood, NJ: Ablex.
- Meyer, B. J. F., & Rice, G. E. (1981). Information recalled from prose by young, middle, and old adult readers. *Experimental Aging Research*, 7, 253-268.
- Meyer, B. J. F., & Rice, G. E. (1983). *Interaction of text variables and processing strategies for young, middle-aged, and older expert readers* (Research Rep. No. 12, Prose Learning Series). Tempe: Arizona State University, College of Education.
- Miller, R. B., Perry, F. L., & Cunningham, D. J. (1977). Differential forgetting of superordinate and subordinate information acquired from prose material. *Journal of Educational Psychology*, 69, 730-735.
- Musgrave, B. S., & Cohen, J. C. (1971). Relationships between prose and list learning. In E. Z. Rothkopf & P. E. Johnson (Eds.), *Verbal learning research and the technology of written instruction* (pp. 79-109). New York: Teachers College Press.
- Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model for sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 485-497.
- Paivio, A. (1968). A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 7, 41-49.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, 76(1, Pt. 2).
- Petros, T., Tabor, L., Cooney, T., & Chabot, R. J. (1983). Adult age differences in sensitivity to semantic structure of prose. *Developmental Psychology*, 19, 907-914.
- Pichert, J. W., & Anderson, R. C. (1977). Taking different perspectives on a story. *Journal of Educational Psychology*, 69, 309-315.
- Postman, L. (1964). Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 145-201). New York: Academic Press.
- Postman, L., & Kruesi, E. (1977). The influence of orienting tasks on the encoding and recall of words. *Journal of Verbal Learning and Verbal Behavior*, 16, 353-369.
- Rubin, D. C. (1974). *Remembering prose*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Rubin, D. C. (1977). Very long-term memory for prose and verse. *Journal of Verbal Learning and Verbal Behavior*, 16, 611-621.
- Rubin, D. C. (1978). A unit analysis of prose memory. *Journal of Verbal Learning and Verbal Behavior*, 17, 599-620.
- Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, 19, 736-755.

- Rubin, D. C. (1983). Associative asymmetry, availability and retrieval. *Memory & Cognition*, 11, 83-92.
- Rubin, D. C., Olson, E. H., Richter, M., & Butters, N. (1981). Memory for prose in Korsakoff and schizophrenic populations. *International Journal of Neuroscience*, 13, 81-85.
- Schmidt, S. R. (1983). The effects of recall and recognition test expectancies on the retention of prose. *Memory & Cognition*, 11, 172-180.
- Schultz, E. E., Jr. (1982). *On the relationship between structure and processing in prose memory*. Unpublished manuscript.
- Seamon, J. G., & Murray, P. (1976). Depth of processing in recall and recognition memory: Differential effects of stimulus meaningfulness and serial position. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 680-687.
- Shaughnessy, J. J. (1979). Subjective rating scales and the control of encoding in incidental learning. *Bulletin of the Psychonomic Society*, 14, 205-208.
- Simon, E. W., Dixon, R. A., Nowak, C. A., & Hultsch, D. F. (1982). Orienting task effects on text recall in adulthood. *Journal of Gerontology*, 37, 575-580.
- Simon, H. A. (1966). A note on Jost's law and exponential forgetting. *Psychometrika*, 31, 505-506.
- Singer, M. (1982). Comparing memory for natural and laboratory reading. *Journal of Experimental Psychology: General*, 111, 331-347.
- Smiley, S. S., Oakley, D. D., Worthem, D., Campione, J. C., & Brown, A. L. (1977). Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. *Journal of Educational Psychology*, 69, 381-387.
- Smith, D. A., & Graesser, A. C. (1981). Memory for actions in scripted activities as a function of typicality, retention interval, and retrieval tasks. *Memory & Cognition*, 9, 550-559.
- Spilich, G. J. (1983). Life-span components of text processing: Structural and procedural differences. *Journal of Verbal Learning and Verbal Behavior*, 22, 231-244.
- Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53-120). Norwood, NJ: Ablex.
- Thorndyke, P. W. (1977). Cognitive structure in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110.
- Underwood, B. J. (1966). *Experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.
- Underwood, B. J., & Schulz, R. W. (1960). *Meaningfulness and verbal learning*. Chicago: Lippincott.
- Voss, J. F., Tyler, S. W., & Bisanz, G. L. (1982). Prose comprehension and memory. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 349-393). New York: Academic Press.
- Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology*, 19, 87-95.
- Whipple, G. M. (1915). *Manual of mental and physical tests: Part II. Complex processes*. Baltimore, MD: Warwick & York.
- Wolk, S. (1974). The influence of meaningfulness upon intentional and incidental learning of verbal material. *Memory & Cognition*, 2, 189-193.
- Young, F. W., & Lewycky, R. (1979). *ALSCAL-4 user's guide* (2nd ed.). Carrboro, NC: Data Analysis and Theory Associates.

Received July 20, 1984

Revision received December 11, 1984 ■