

## A Unit Analysis of Prose Memory

DAVID C. RUBIN

*Lawrence University*

Four stories were divided into function word units. These units were assigned dependent variable values determined by the scoring of subjects' recalls and independent variable values determined by measures of gist, imagery, repetition, frequency of occurrence, serial position, grammatical connectedness, centrality in a propositional net, and subjects' intuitions of which units would be remembered. The independent variables were all statistically significant predictors of recall. Subjects' intuitions and gist were the best predictors of the more structured stories, while repetition and serial position were the best predictors of the less structured stories. For each story, the underlying rank ordering of function word units from most to least likely to be remembered was the same for all subjects (i.e., scalable). While changes in the retention interval, subject population, and motivation level affected the amount recalled, these changes had little effect on the rank ordering of the units from most to least likely to be remembered. Changes in the retrieval task from free recall to prompted recall and recognition affected both the amount and rank ordering of units.

A multitude of independent variables and experimental conditions have, in isolation, been shown to affect prose memory (Kausler, 1974; Welborn & English, 1937). The purpose of this paper is not to add to this list, but rather to demonstrate a method of analysis which allows the many variables and conditions to be evaluated simultaneously. In proposing this method, it will be argued that not only does the method uncover enough regularity to allow for the prediction of individual subjects' recalls, but also that the method, or one like it is,

Requests for reprints should be sent to the author who is now at Department of Psychology, Duke University, Durham, NC 27706. The author wishes to thank Donald Broadbent, Roger Brown, Louis Guttman, Reid Hastie, Edwin Newman, Philip Stone, and Eric Wanner for their advice and encouragement; Thomas Baker, Francis Campos, Michael Goldstein, Edwin Olson, and Barbara Steinlauf for their comments on later drafts; Michael Hart and the Quincy Public Schools for the data from the eighth-graders; and John Thomas and the Normative Aging Study (HEW/NICHD Grant HD 05699.01) for the data from the older men. Part of this research constituted the author's PhD thesis (Harvard University, 1974). The remainder was sponsored by Lawrence University and NIMH Grant No. 1 R03 MH27024-01.

necessary for the understanding of prose memory.

Broadbent (1973) has argued that psychologists should try to formulate alternative hypotheses and perform experiments that compare them, rather than testing one hypothesis against the null hypothesis. In an area, such as prose memory, where many factors have been demonstrated to have statistically significant effects, this approach seems imperative. Where the variables are not mutually exclusive, not only their relative effectiveness, but also the interrelations among them should be considered. While large experiments crossing many experimental conditions with many hypotheses would be useful, with the standard experimental method, they would be impractical. The following variation in method is therefore offered.

In quantitative studies of prose memory, the scored recall protocols are represented as a subject by unit matrix of 1's (subject<sub>i</sub> remembers unit<sub>j</sub>) and 0's (subject<sub>i</sub> does not remember unit<sub>j</sub>). The standard method is to sum over units collapsing the matrix to a group mean

and standard deviation, and thereby a measure of how much the average group member remembers. Here, in addition to the sum over units, sums over subjects are obtained. The set of sums over subjects, one sum for each unit in the story, is referred to as the pattern of recall. This pattern of recall provides an estimate of the probability of each unit in the story being remembered. In this way a quantitative description of both *how many* and *which* units groups remember is obtained. This description will hopefully be able to quantify much of the richness which more qualitative studies, such as Bartlett's (1932), uncover without sacrificing experimental rigor.

Using this method hypotheses can be tested on their ability to predict which units are recalled. Any hypothesis which can assign values to the units of a story can be compared to the pattern of recall by a correlation coefficient calculated over the units of the story. Because of the correlational nature of the analysis, many hypotheses can be compared simultaneously. Similarly, the interrelations among the patterns of recall from various experimental conditions can be compared. Thus, the relative importance of the independent variables in a variety of experimental conditions can be evaluated.

The uncollapsed matrix of 1's and 0's can also be analyzed. It provides a quantitative description of which units individual subjects remember; that is, it indicates not only how many units individual subjects remember, but also which units they remember.

If all subjects have the same underlying ordering of units from most to least likely to be remembered, then the number of units that any subject recalls determines exactly which units that subject recalls. In particular, if a subject recalls  $n$  units, these  $n$  units will be the  $n$  units most often recalled by the group as a whole and all other units will fail to be recalled. Guttman's coefficient of reproducibility (Stouffer, Guttman, Suchman, Lazarfeld, Star, & Clausen, 1950) provides a measure of how closely actual data approaches this ideal of

scalability. As an example, assume we predict exactly which 24 units of a 50-unit story a subject recalls by assuming that that subject will recall the 24 units most often recalled by the group as a whole. By default, we are also predicting which 26 units will fail to be recalled. Correctly predicting 20 of the 24 units the subject recalls (and therefore 22 of the 26 units the subject failed to recall) would result in 42 correct predictions out of a possible 50, for a coefficient of reproducibility of  $42/50$  or  $.82$ . Chance levels of the coefficient vary from  $.50$  when all subjects recall half of the units to  $1.00$  for the trivial case of predicting exactly which units are recalled when subjects either recall, or fail to recall, all of the units (Kenny & Rubin, 1977).

Which units individual subjects recall can be predicted from independent variables as well as from their grouped recall data. In this case, if an individual subject remembers  $n$  units of a story then a hypothesis formed to test the predictive power of an independent variable would predict that the  $n$  units remembered should be the ones having the highest values on that independent variable.

It should be noted that the experimental procedures used to obtain and score the recall protocols are standard. What is novel is the attempt to demonstrate the power of a multivariate unit analysis for understanding prose memory, both for individual subject and grouped data.

The advantages of this simple change to a unit analysis should not be overlooked. In the standard method, a set of passages must be specifically prepared to vary with respect to a given variable, and hopefully only with respect to that variable. Using the unit analysis proposed here, naturally occurring prose can be sampled. In the standard method passages must be written to vary with respect to every variable to be investigated. Thus, if four variables are to be investigated simultaneously and if three levels of each variable are to be used (e.g.: a high, medium, and low level), then  $3^4$  or 81 versions of the same basic passage

must be written. Using the unit analysis one passage could be used to investigate all four variables. If with the unit analysis several stimulus passages were chosen to sample a wide range of writing, then generalizations about the relative importance and interrelationships of the independent variables could be made. No such goal is obtainable using the standard method, both because naturally occurring passages cannot be sampled, and because the number of passages needed to investigate many variables at once is prohibitive.

The first section of this paper describes the method in more detail and applies it to the recall of three stories by individual subjects. The second section groups the individual subjects' data to allow for more powerful statistical analyses. In the third section, Bransford and Johnson's (1972) demonstration that proper context can greatly affect the amount recalled from a passage is investigated in detail. The fourth section considers the role of retention interval and type of retrieval task. The fifth and sixth sections examine incidental vs intentional learning, and the effects of different subject populations, respectively. While the basic intent of these sections is to argue for the adoption of a method, the individual sections are not without content and as a whole present a novel and consistent picture of prose memory.

#### SECTION 1: THE METHOD AND ITS APPLICATION TO INDIVIDUAL SUBJECTS

Section 1 begins by considering three critical issues: (1) the choice of units, (2) the choice of stories, and (3) the choice of independent variables. Following this, Experiment 1 will be described.

##### *Units*

Stories were divided into units consisting of one or more grammatically defined content words (i.e., nouns, verbs, adverbs, and ad-

jectives) based on Fries' (1952) grammar.<sup>1</sup> Content words with a common referent or with meanings too similar to have reliably distinguishable synonyms were combined to form one unit.

The units for one of the stories, the Lincoln story, are shown in Fig. 1. The complete story can be read down the column. The first instance of each of the 47 units is set in upper case letters. If a unit is composed of more than one word, its other orthographic forms follow in parentheses. The word "way" appears as two separate units because its two occurrences have two different meanings and thus two sets of easily distinguishable synonyms.

The recall protocols were scored for the presence, or absence, of the units. Synonyms present in the recall protocols were accepted regardless of case, tense, or number, but only if they were synonyms for the particular sense of the word used in the original story. This method yielded reliable scoring. A correlation of 0.97 was obtained between the author's scoring of 16 recall protocols from the Lincoln story and that of an independent judge who was provided with both the general rules for scoring and the scored recall protocols of two other stories.

Proposition length units (e.g., Kintsch, 1974; Rumelhart, Lindsay, & Norman, 1972) were not used because they did not allow several hypotheses to be formulated in a reasonable way. As subjects almost always recall content words in their correct subject-verb-object or modifier-head relations, scoring content words produces units which could easily be translated into propositions-length units. Thus, although there may appear to be a significant difference between scoring words and scoring propositions; in fact, it appears from examination of the data that there would be little difference in the results obtained.

<sup>1</sup> The rules used to divide the stories into units, score the recall protocols, and translate the stories into Rumelhart, Lindsay, and Norman (1972) notation, the full tables of correlations among the variables, and copies of the stories are available from the author.

### Stories

Three stories, the Lincoln, Peanut, and Gull stories, were used as examples of prose which are coherent, about 150 words long, and do not present any strong views that might have affected the subject populations differentially.<sup>1</sup> Here, "coherent" is meant to imply only that subjects should be able to agree on paraphrases of the story while they have the story in front of them. Within these restrictions the stories were chosen to sample a range of prose.

The Lincoln story is from an early study of memory (Whipple, 1915). The story has a familiar type of anecdotal plot and "makes sense" in our culture. The Peanut story was used by Bartlett (1932). It also has a familiar kind of anecdotal plot but of an entirely different kind from the Lincoln story, one including elaborate semantic embedding. The Gull passage is from *The World of the Herring Gull* by Tinbergen (1967, p. 44). This excerpt was chosen because, in contrast to the two other stories, it lacked a standard plot structure. Although it does have a beginning and an end, it is much more a description of arbitrary aspects of a scene than a narration of a connected sequence of events.

### Independent Variables

Table 1 lists the variables selected for investigation. The names and abbreviations used in Table 1 will be referred to throughout the paper. To simplify the notation the same names will be used to refer to both the independent variables and the hypotheses which claim that units scored high on those variables will be remembered most often.

Three measures are used to evaluate the idea that the gist, or theme, of a passage is what is remembered. The two telegram variables follow from Gomulicki's (1956) hypothesis that recall is an abstractive process.

A simple imagery hypothesis would be that recalls are constructed from visual images evoked by a passage. It should be noted that the present sense of imagery is different from

Paivio's (e.g., Paivio, Yuille, & Madigan, 1968) in that the contribution to the overall visual image is considered and not the imagery value of each word in isolation.

The two Mnemonic Variables were developed in an attempt to measure the predictive value of people's intuitions about memory (Christie & Pettijohn, 1977). Naive was suggested by Eric Wanner (personal communication, 1972). Memo was an attempt to obtain the mnemonic core or outline of the story.

Miller and Selfridge (1950) claimed that redundancy is important in recall. Using orders of approximation to English, they found that "When short range contextual dependencies are preserved in nonsense material, the nonsense is as readily recalled as meaningful material" (p. 185). Based on this result they claimed that "contextual dependencies existing over five or six words permit positive transfer, and that it is these familiar dependencies, rather than meaning per se, that facilitate learning" (p. 185). CLOZE (i.e., closure) is included as a measure of item redundancy. The correlation of CLOZE and the Gist Variables as well as the value of CLOZE in predicting the pattern of recall provides a test of Miller's claim.

The Serial Position hypothesis holds that units near the beginning and end of the story will be remembered more often than units near the middle. This finding holds for lists of related sentences (Deese & Kaufman, 1957; Jersild, 1929) as well as for lists of words. Unstructured lists of sentences, however, are not stories, and thus, this study can be seen as an extension of this finding to prose (cf. Fox, Note 1).

The two Structural Variables were included to measure the effects on long-term memory of surface structure and of a somewhat "deeper" level of analysis, which is representative of much of the research using semantic nets (Kintsch, 1974; Rumelhart, Lindsay, & Norman, 1972). The Rumelhart, Lindsay, and Norman system was chosen because its prop-

TABLE 1  
VARIABLES

Name	Abbreviation	Method of formation
Gist Variables		
Gist	gist	Subjects' ratings of each unit's "contribution to the total meaning of the story"
20-word telegram	tele-20	The number of times each unit was used by subjects in writing a 10-20 word "telegram to communicate the major ideas of the story to a stranger"
5-word telegram	tele-5	The number of times each unit was used by subjects to "complete the following sentence using 5 words or less: The story is about ..."
Imagery Variable		
Imagery	image	Subjects' ratings of each unit's "contribution to the total visual image produced by the story"
Mnemonic Variables		
Naive	naive	Subjects' ratings of each unit on "the probability that it, or a synonym, will be recalled by other subjects in a situation just like the one you were just in"
Memo	memo	The number of times each unit was used by subjects to write a 15-word memo to serve as an aid to themselves in an imaginary six-month recall session
Redundancy Variable		
CLOZE	CLOZE	The number of times each unit was filled in by subjects in a standard CLOZE technique
List Learning Variables		
Serial position	<i>U</i>	Values determined by the location of each content word in the passage. Words near the beginning and end of the passage have a greater value than words near the middle
Repetition Frequency	rep. freq.	The number of times each unit appears in the passage. Values determined by the frequency of occurrence in English of the content words making up each unit
Structural Variables		
Number of grammatical connections	ngc	The number of grammatical connections to the unit in the surface structure of the story
Net	net	The number of connections to each unit (node) in a semantic network representation of the story
Dependent Variable		
Pattern of recall	recall	The number of subjects remembering each unit

ositions could be easily translated into the content word units used here.

#### Experiment 1

##### Method

*Subjects.* Thirty-six Harvard University undergraduates were paid \$2.50 each to

participate in a 1½ hour session. Twenty-eight other Harvard University students were paid \$1.50 each for a 30-min session run to obtain either additional recall or CLOZE data. All subjects were native speakers of English.

*Materials and procedure.* Each subject provided a recall protocol for one story and after-

wards several judgments on the two remaining stories.

The stimulus material was presented in the form of mimeographed booklets. The subjects were informed that the purpose of the experiment was to study various psychological properties of three stories, and were provided with some general information on the procedures and the construction of the booklet. It was indicated that all tasks were of equal interest to the experimenter. Examples of the Thurstone Letter Series Completion Test (e.g., a b a b a b \_ ) which would fill the 10-min interval between the reading and recall of a story were provided. The subjects were then informed that they would be asked to recall a story as accurately as possible and were instructed to read the story through twice at their normal reading speed. The subjects spent between 35 sec and 1 min 35 sec reading the story with a median time of about 1 min. The Thurstone Letter Series Completion Test proved very interesting to the subjects and seemed to distract their attention from the recall task. Twelve additional subjects were run at a later date to provide four additional recall protocols for each of the three stories.

The judgment tasks were self-paced. If the three stories are denoted as A, B, and C, then subjects performed the tasks in the following order: Recall (A), Naive (B), CLOZE (C), Memo (B), Gist (C), Imagery (B), and Telegrams (C). Six subjects were assigned to each of the six possible pairings of the Lincoln, Peanut, and Gull Stories with A, B, and C.

On the last page of the booklet, subjects were asked to rate the difficulty of the various tasks on a 5-point scale and to list any tasks on which they "gave up or did not put in any real effort." Subjects' responses were not scored for any task on which they indicated maximum difficulty or no effort (about 3% of the cases).

*Forms for obtaining judgments.* To obtain the rating needed for the Gist, Imagery, and Naive variables, each story was first displayed in paragraph form and then on the subsequent

One DAY  
MR. LINCOLN (he,I)  
was OUT  
RIDING (rode).  
As he PASSED  
along the ROAD  
he SAW  
a PIG(he,him,piggy)  
SINKING  
into a MUD-HOLE(mud).  
POOR  
piggy would CLIMB  
PART  
WAY  
up the SLIPPERY  
BANK  
then DOWN  
he would FALL  
AGAIN: "I  
SUPPOSE (think,thought)  
I should GET down  
and HELP that pig,"  
thought Mr. Lincoln  
"but I HAVE ON  
my NEW  
SUIT (clothes)  
and it will be QUITE  
SPOILED if I do so.  
I think I'll LET  
him get OUT  
the BEST  
WAY  
he CAN."  
He rode ON. (away)  
When NEARLY  
TWO  
MILES  
away he TURNED  
and CAME  
BACK.  
Not MINDING the  
clothes he STOOPED  
and TAKING (taken)  
the piggy in his ARMS  
he DRAGGED him out of  
the mud. The new suit  
was quite spoiled but  
he SAID  
he had taken a PAIN  
out of his MIND.

FIG. 1. The Lincoln Story divided into units.

page(s) in a one-unit-per-line basis in a manner similar to that shown in Fig. 1, but with a 5-point rating scale to the right of each unit. The Telegram and Memo forms were constructed by placing the instructions and blank lines on the same page as the story.

*Formation of the CLOZE Variable.* Redundancy can be measured in many ways (Rubin, 1976). Here, every third content word

was replaced by a blank space in order to measure the redundancy caused by all parts of the story. Three forms were prepared for each story: one form started with the first content word, one form started with the second content word, and one form started with the third. Each subject saw only one form. The subjects, who had never seen the complete story, were asked to read their form through before filling in any blanks, and were asked to guess in order to try to fill in every blank. After the subjects had completed the procedure, sets consisting of one of each of the three different CLOZE forms were compiled. Each of these sets were scored together as if they were a single recall protocol from a single subject.

Three times as many forms were needed with the CLOZE task, and so extra subjects were run. Sixteen naive subjects were given one CLOZE form from each of the three stories. The form and story order were randomized across subjects.

*Formation of the List Learning and Structural Variables.* The Serial Position, Frequency, and Number of Grammatical Connections variables were formed by assigning a value to each content word. These values were summed for units containing more than one content word. The values for the U-shaped serial position curve were obtained by assigning the first and last content words in an  $n$ -word story a value of  $n/2$  if  $n$  was even and  $(n + 1)/2$  if  $n$  was odd. The second and next to last content words were assigned a value smaller by 1, and so forth. The values for the Frequency variable were obtained by assigning each content word the logarithm of its frequency in the Kučera and Francis (1967) frequency count. Words not listed in the count were assigned a frequency of 1. For the Number of Grammatical Connections variable, each content word was assigned 1 point for each time it was modified by a content word, phrase, or clause, for each time it modified a word either by itself or as the head of a prepositional phrase, for each time it was a subject or object of a verb, for each time it

was a verb with a subject, and for each time it was a verb with an object.

In order to evaluate the Net Hypothesis, a set of rewrite rules was devised to convert text to a Rumelhart, Lindsay, and Norman (1972) type net. Alex A. Archbald, who helped devise these rewrite rules, and the experimenter agreed on over 80% of the relations independently translated from text. The net differed from a full representation in that it included only relations that were explicitly stated in the text. While the Rumelhart *et al.* (1972) system was inadequate for certain English constructions, especially those requiring reference to a group of relations (e.g.: the pronoun *that* in: And *that* is the whole story), in general it provided the basis for a reliable transformation of text into net representation.

### Results

Subjects recalled 56 ( $SD = 9$ ), 64 ( $SD = 9$ ), and 29% ( $SD = 12$ ) of the units of the Lincoln, Peanut and Gull stories, respectively. Each of the dozen variables of Table 1 can be used to predict exactly which of the units individual subjects remembered and failed to

TABLE 2

PROPORTION OF UNITS CORRECTLY PREDICTED BY THE VARIABLES OF TABLES 1

Variable	Story		
	Lincoln	Peanut	Gull
Gist	.630***	.676***	.651**
Tele-20	.638***	.684***	.646*
Tele-5	.580**	.691***	.660*
Naive	.676***	.718***	.681***
Memo	.715***	.668***	.660**
Image	.652***	.601**	.644*
CLOZE	.633***	.673***	.675***
U	.580**	.663***	.719***
Rep	.564	.676***	.708***
Freq	.548	.696***	.688***
Ngc	.580**	.630***	.677***
Net	.569*	.617***	.644*

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

remember by predicting that if a subject recalled  $n$  units then those  $n$  units would be the units that had the  $n$  highest values on the variable in question and that all other units would fail to be recalled. Table 2 lists the proportion of times that such predictions were correct. It should be stressed that these proportions are the average of predictions of individual subjects' recalls, and not of the grouped data.

A binomial model provides a measure of how accurate such predictions would be by chance; that is, without a hypothesis to indicate which  $n$  units to predict (Kenny & Rubin, 1977). For the Lincoln, Peanut, and Gull stories the by chance prediction would predict .528, .552, and .612 of its guesses correctly. The significance levels in Table 2 were obtained by calculating a matched  $t$ -test between the observed and the chance predictions for the 16 subjects who recalled each story.

Thus, all 12 variables are fairly successful at predicting which units individual subjects will recall, though none is very much better than its competitors. There is, however, regularity in the data for which none of the variables can account. If the number of times each unit is recalled by the group is used as a predictor, exactly which units each individual remembers and fails to remember can be predicted correctly 79.8, 84.9, and 77.8% of the time for the three stories, respectively. These values, which are known as the Guttman coefficient of reproducibility, are considerably higher than the predictions of any of the individual variables.

As the number of times each unit is recalled by the group is a predictor which is not independent of the individual subject's data it is predicting, the binomial by chance model cannot be used. Rather a statistic based on the number of units each subject recalls (Kenny & Rubin, 1977) provides appropriate by chance values of .605, .627, and .695 for the Lincoln, Peanut, and Gull stories, respectively. Using Green's (1956) method to obtain the standard

error in the by chance estimates (Chilton, 1969), the coefficients of reproducibility for the three stories are different than their by chance values (all  $p < .001$ ). It may be noted that the method of calculation of the coefficient of reproducibility used here is quite conservative (Kenny & Rubin, 1977) and that higher values would be obtained if other methods were used (Chilton, 1969).

#### *Discussion*

The ordering of units from most to least likely to be remembered is Guttman scalable. Given the grouped data and the number of units an individual subject remembered, exactly which units that subject remembered can be predicted with 80% accuracy. Thus, there is great regularity among subjects in which units are remembered relative to which are forgotten, even though the amount remembered varies. This finding becomes important when one considers that a major reason for using nonsense syllable lists in memory experiments was the lack of a suitable way of obtaining results with meaningful material that would not be affected by the idiosyncracies of the subjects' past histories.

A dozen variables were each moderately successful in predicting which units individual subjects recalled. Before these variables can be fully evaluated as predictors of recall, their relationship to each other must be assessed.

#### SECTION 2: A CORRELATIONAL ANALYSIS

No new data is introduced in this section. Rather, the data of the individual subjects of Experiment 1 is grouped in order to provide a more stable index and to allow the use of a greater variety of statistical techniques. This is accomplished by summing over subjects to provide a set of values, one for each unit in a story. This set of values, termed the pattern of recall, can be compared with the sets of values obtained from the independent variables by correlating over the units of a story. One caution is necessary in using the grouped

data. It cannot be assumed that the units of analysis (i.e., the units of the story) are statistically independent. Though false, this independence assumption is quite common in psychology (e.g., Johnson, 1970; Loftus & Suppes, 1972; Underwood & Schulz, 1960). For tests of unrelated items the independence assumption may not be a bad approximation. For prose it is more difficult to justify. Fortunately, the independence assumption is not needed in order to use correlation coefficients as descriptive statistics. It is needed only if the correlation coefficient is to be used as an inferential statistic. Therefore, the descriptive correlational analyses will be supported by proper inferential statistics.

correlations for the Lincoln, Peanut, and Gull stories are based on 47, 49, and 72 observations, respectively (i.e., on the number of units in each story).

The most outstanding result of Fig. 2 is the overall success of the variables in predicting which units are recalled. The second result to be drawn from Fig. 2 is that the patterns of recall of the two more structured stories, the Lincoln and the Peanut stories are predicted almost equally well by the independent variables, while the less structured Gull story has a different profile. The pattern of recall from the Gull story is predicted consistently lower by the Gist, Mnemonic, and Imagery Variables and consistently higher by the List Learning and Structural Variables. It should be noted that the Gull story not only appears a priori less structured and more list-like but it is also recalled to a lesser degree [ $t(30) = 6.86, p < .001$  compared to the Lincoln story  $t(30) = 8.78, p < .001$  compared to the Peanut story. The differences in the amount recalled between the Peanut story and the Lincoln story are not significant.]

*Correlations Between the Patterns of Recall and the Hypotheses*

Figure 2 displays the correlation coefficients of the pattern of recall with the variables tested for each of the three stories. The listing of the variables along the abscissa follows the order used in Table 1. Thus, the conceptual groupings indicated in Table 1 are maintained. The

A third result is that the Mnemonic

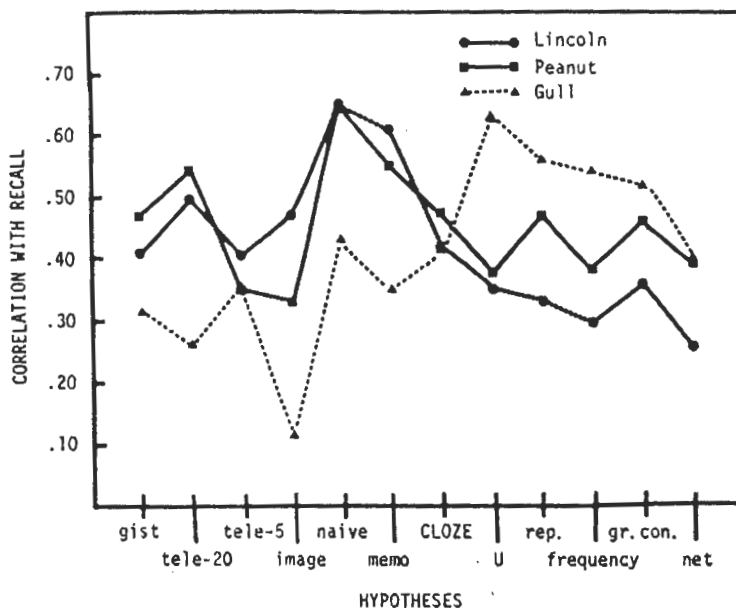


FIG. 2. Correlations of the variables listed in Table 1 with the patterns of recall for each of the three stories.

Hypotheses were better predictors of the pattern of recall than the Gist Hypotheses. Thus, it appears that people have knowledge about memory that goes beyond their ideas of "what's important." Methodologically this finding indicates that subjects providing gist ratings for memory experiments should not be aware of their ratings' ultimate use. Finally it should be noted that the correlations for one variable are not valid because examination of scatterplots revealed that that variable had a nonlinear relation to recall. The *U*-shaped function obtained with Frequency, however, is consistent with studies of the learning of mixed frequency word lists (Peters, 1936).

#### *Correlations Among Variables*

The average correlations among all the variables was quite high: .615, .647, and .421 for the Lincoln, Peanut, and Gull stories, respectively. This was true not only among variables that purport to measure the same thing, but also among variables with little a priori reason to be correlated. Examination of the correlation matrices and the resulting multidimensional scaling (Smallest Space Analysis: Guttman, 1968) revealed that the relationship among the independent variables was fairly constant for the three stories. There were two major clusters of variables, one composed of the Gist, Mnemonic, and Imagery Variables, and the other composed of the List Learning and Structural Variables. As would be expected from Fig. 2, however, the relation of the pattern of recall to the independent variables was not constant over the stories. The pattern of recall was close to the Gist cluster in the Lincoln and Peanut stories and close to the List Learning cluster in the Gull story.

The CLOZE variable was the only independent variable not closely related to either of the two major clusters of variables. Its independence argues against Miller and Selfridge's (1950) claim that familiar contextual dependence rather than meaning per se is what

facilitates the learning of text. Both facilitate learning—but they are distinct.

#### *Multiple Regression Analyses*

As the correlations among the independent variables are fairly high, multiple regression analyses do not add much to the amount of variance accounted for by the best individual variable. If all 12 independent variables are used, 62.2, 56.4, and 52.2% of the variance in the pattern of recall can be explained for the Lincoln, Peanut, and Gull stories, respectively, vs 46.8, 41.4, and 39.1% for the best individual independent variables for each story.

Using all 12 independent variables in the multiple regression analyses, however, is not good practice. First, some of the independent variables may be adding predictive power on the basis of chance variation rather than on the basis of some underlying relationship; that is, the weights of some of the variables in the multiple regression equations may not be statistically different from zero, or no contribution. Second, the pattern of recall is being predicted by two variables we know even less about than recall itself: the two Mnemonic Variables, Naive and Memo. These variables, while quite successful, are based on subjects intuitions rather than on psychological theory or objective properties of the stories. A more reasonable multiple regression would not include variables unless their weight was at least large enough to be statistically different from zero under the independence assumption, and would not include the two Mnemonic Variables. For the Lincoln, Peanut, and Gull stories such restrictions yield the following equations: Recall = .655 Tele-20 + 7.267 ( $r^2 = .268$ ), Recall = .418 Tele-20 + .494 CLOZE + 5.889 ( $r^2 = .341$ ), and Recall = .049 *U* + .265 CLOZE + 1.082 Gist - .313 ( $r^2 = .470$ ). As a psychologist it is humbling to note that except in the case of the relatively unstructured Gull story, where both the List Learning and the Gist Variables are good predictors, the combined force of most of psychology's major theories are no better than the

ratings of naive subjects. Even relaxing the statistical criterion and using all of the independent variables except for the Mnemonic Variables does not change this result. The amount of variance explained under these conditions is 41.3, 40.8, and 50.9% for the Lincoln, Peanut, and Gull stories, respectively, compared to 46.8, 41.4, and 18.8% for Naive alone.

### *Discussion*

Using a sample of three stories not specifically constructed to control for all but one variable, many variables were found to have statistically significant correlations with the pattern of recall and with each other. The correlations among the variables are quite regular over the sample of three stories, and the correlations of the hypotheses with the pattern of recall varied systematically. The high correlation among many of the variables formed to test conceptually different hypotheses points to both a weakness and a strength of the approach used in this study. It is impossible to sort out the relationships of highly correlated variables without using more controlled stimuli, specially prepared for that purpose. At the same time, the occurrence of a high correlation among variables is an empirical finding which might add understanding to the mechanisms sorted out experimentally.

The high correlation among the independent variables points to a weakness in the standard method as well. Studies claiming to use passages that vary one independent variable while holding all others constant probably do not really do so. The high correlation among independent variables indicates that, for example, passages written to vary in syntactic structure most likely vary gist, imagery, and redundancy variables as well. But these other confounded variables correlate with prose recall. Preparing passages that hold all but one independent variable constant may in practice be impossible.

Examining the relationships among the independent variables in more detail it was

found that for the two more structured stories the degree to which the various variables were predictors of the pattern of recall was fairly constant. For the less structured passage, List Learning Variables were more important. This difference in the types of variables which best predict the pattern of recall suggests an empirical scale similar to the senseless connections—meaningful connections; memorization—learning by understanding continuum proposed by Katona (1940). In more current terms this scale of stimulus structure would be seen as determining the maximum possible depth of processing. At the senseless connections extreme there can be, by definition, no gist. Decades of research at this extreme have shown serial position, repetition, and frequency of occurrence in English to be important. Here, we have found that elsewhere on the continuum these factors correlate highly with the Gist Variables, yet partial correlations would show that they still have some predictive power independent of Gist. What could account for these results? Suppose, as did early experimenters, that one can study the basic factors of memory by stripping away all structure and that adding the structure back does not destroy the basic factors but only serves to add new factors to them. This assumption could account for the simultaneous success of both the List Learning and the Gist Variables, but one further assumption is needed to account for their relatively high intercorrelation. It has been shown that naive subjects know more about what is remembered from a story than would be expected from their knowledge of gist alone. Assume that, among other factors, they know that repeating an idea or placing the idea in the first or last sentence of the passage or using a rather rare or a rather common word to express the idea would all tend to improve the recall of that idea independent of its importance in the overall plot of the story. Then, as authors, these naive subjects should use these and other such factors together to stress ideas they want remembered. True, they need not use these

factors if the plot is compelling enough, but they do so only at the risk of having their reader remember less of exactly what they want to convey.

### SECTION 3: WASHING CLOTHES

The last section ended with the claim that variables based on senseless connections play the same role in the learning of prose that they play in the learning of unrelated lists of words; the only difference being that in prose the effects of these variables are often masked by effects of variables based on meaningful connections. Stories used to study the effect of context on prose learning (e.g. Bransford & Johnson, 1972; Dooling & Lachman, 1971) provide ideal stimuli to test this claim. Bransford and Johnson's (1972) Washing Clothes passage is one example. With the title of washing clothes, the passage is an easily understood, coherent story; without the title it is a list of unrelated phrases. Adding the title does not affect the values of variables based on senseless connections because the wording of the passage does not change, but adding the title does make the formation of variables based on meaningful connections possible.

If variables based on senseless connections are operating in prose memory, then these variables, as exemplified by the List Learning Variable, Serial Position, should be better predictors when no title is given: that is, when the masking effects of variables based on meaningful connections are reduced. Likewise, variables based on meaningful connections, such as Gist, should be better predictors when the title is given. In other words, as the maximum possible depth of processing changes, the relative importance of the various processes operating should also change. Experiment 2 tests this notion.

#### *Experiment 2*

##### *Method*

*Subjects.* Thirty-two undergraduates at the

University of Wisconsin Fox Valley Center recalled the Washing Clothes passage, and 16 Lawrence University undergraduates provided judgments for the Gist variable.

*Materials.* Bransford and Johnson's (1972) Washing Clothes passage B (p. 722) was divided into 59 units and the Serial Position variable was formed using the rules described in Section 1.

Booklets were used to present the Washing Clothes passages in the recall experiment. Half of the booklets had titles inserted; the other half did not. All booklets contained a final questionnaire to allow the elimination of any subjects who were familiar with the story or who in the no title condition guessed what the story was about.

Booklets used to obtain the Gist judgments contained the story with a title displayed first in paragraph form and second in a one unit per line form as described in Experiment 1. Judges were asked to check the 16 units which they felt were the most important to the overall meaning of the story. A selection rather than a rating procedure was used because of the relative success of the variables formed by selection, rather than rating, in Experiment 1. Sixteen units were selected because this was the average number of units recalled in the recall condition.

*Procedure.* Subjects in the recall condition were asked to read their passage through twice at their normal reading speed, and were informed that they would be asked later to recall the passage as accurately as they could. A 10-min retention interval was filled with their scheduled introductory psychology lecture. The questionnaire followed the recall.

##### *Results and Discussion*

The 16 subjects who were given the title recalled more than the 16 who were not [34.1%,  $SD = 10.4$  vs 20.1%,  $SD = 8.4$  correct,  $t(30) = 4.18$ ,  $p < .001$ ]. The recall data from title and no title groups were orderly, with split half reliabilities as measured by Cronbach's  $\alpha$  equal to .877 and .911.

respectively, and Guttman coefficients of reproducibility of .797 and .881, respectively, as compared to .648 and .751 expected by chance (both  $p < .001$ ).

Serial Position predicted the pattern of recall better for the no title group than the title group ( $r = .743$  vs  $.671$ ), while Gist predicted the pattern of recall better for the title group than the no title group ( $r = .684$  vs  $.543$ ). As the 59 units of the story on which these correlations were calculated cannot be considered as independent observations, the success of these two variables in predicting exactly which units individual subjects recalled was used to test the statistical significance of the findings. As the amounts recalled in the two conditions, and thus the chance levels of the coefficient of reproducibility, varied greatly between the two groups, Green's (1956) Index of Consistency was used to adjust the individual coefficients of reproducibility. While the main effects of title-no title and Serial Position-Gist were not significant (both  $F < .50$ ), as predicted, the interaction was ( $F(1, 30) = 6.93, p = .01$ ). Thus, in going from the study of isolated words to the study of connected stories, psychologists need not, as some would claim (e.g., Jenkins, 1974), forget their old theories; they merely have to add in their new ones.

#### SECTION 4: RETENTION AND RETRIEVAL

The studies presented so far, like most in prose memory, deal with intentional learning by college students followed fairly quickly, though not immediately, by a free recall. The remaining three sections extend these results by varying the retention interval, retrieval task, subject population, and motivational level. Throughout these sections, common findings will be replicated using standard experimental designs. In addition, the unit analysis developed in the first three sections will be used to provide a measure of which units individual subjects remember.

Section 4 considers various retention inter-

vals ranging from 3 sec to 4 weeks. As Bartlett (1932) stressed, remembering is an active, constructive process. Retention, however, need not add to that activity. Some research indicates that retention does affect what is remembered in an active way; in particular, that details are forgotten more quickly than ideas central to a story's theme (e.g., Bower, 1976). However, other research indicated that retention affects only how much is remembered rather than what (e.g., Dawes, 1966). Here, this paradox can be redefined and resolved in a direct quantitative manner by examining the patterns of recall from different retention intervals.

In addition to using several retention intervals, retrieval tasks will be varied. If differences in what is remembered are apparent from examining the patterns of recall, then these differences can be examined by using the variables from Table 1. In this way, a better understanding of the processes of retention and retrieval can hopefully be obtained.

#### *Experiment 3*

##### *Method*

*Subjects.* Sixty-four Harvard University undergraduates enrolled in a large introductory psychology course were each paid \$1.50 for their participation. All were native speakers of English, unfamiliar with the story used.

*Procedure.* The subjects were assigned to either an immediate, 10-min, 1-week, or 1-month retention period. All subjects were instructed to read the Lincoln story through twice at their normal reading speed and, after the retention interval, to write the story down as accurately as they could. This free recall was followed by a prompted recall. The prompt consisted of the original story with all nouns, main verbs, adverbs, and adjectives (with the exception of articles) replaced by a blank space. The subjects were instructed to fill in each blank, guessing if necessary. The

retention interval for the subjects in the immediate group consisted of turning over a page. For the 10-min group, the retention interval was filled with a Thurstone Letter Series Completion Test. As it took about 3 sec for subjects in the immediate group to turn the page and start writing, the retention intervals are equivalent to approximately  $4 \times 10^{-2}$ , 10,  $10^4$ , and  $4 \times 10^4$  min, a range of six orders of magnitude.

### Results

Table 3 contains the percentage of units recalled correctly in each condition. As expected, there was a statistically significant decrease in the amount recalled with time for the free and prompted recalls [ $F(3,60) = 12.8$ ,  $p < .001$ ], with the free recall decreasing faster than prompted recall [ $F(3,60) = 3.59$ ,  $p < .05$ ]. Subjects recalled significantly more in the prompted than in the free recall situation for all the individual retention intervals except for the immediate (i.e., the  $p < .001$  confidence intervals for the means of prompted minus free recall did not include 0), as well as for all intervals combined [ $F(1,60) = 68.0$ ,  $p < .001$ ]. Thus, there was a significant percentage of units retained but not freely recalled, and this percentage increased as a function of the retention interval.

The relative values of the patterns of recall were surprisingly constant over time for both

the free and prompted recall. The average correlation among the patterns of recall from the four free recalls was .835 compared to an average split-half reliability (Cronbach's  $\alpha$ , Cronbach, 1951) of .890. The average correlation among the patterns of recall from the four prompted recalls was .876 compared to an average split-half reliability of .891. The full set of correlations are shown in Table 4. Thus, the relative degree to which units are remembered is only slightly affected by the passage of time.

An even stronger result is that this time insensitive ordering of units holds for individual subjects as well as for the group data. The coefficient of reproducibility for the free recalls from the four retention intervals combined is .786. The coefficient of reproducibility for the four free recall groups taken separately are .801, .795, .793, and .816 (.690, .630, .630, and .624 by chance, all  $p < .001$ ). This implies that there exists one underlying rank ordering of units from most- to least-likely to be remembered which is the same for all subjects and all retention intervals. That is, given the overall rank ordering of units from most to least likely to be remembered calculated from all 64 subjects and the number of units an individual subject remembered, we can predict exactly which units that subject remembered with an accuracy of almost 80%. Knowing the rank ordering for that subject's

TABLE 3  
PERCENTAGE OF UNITS RECALLED

Task	Retention intervals			
	Immediate	10 min	1 week	1 month
Free recall				
mean	70.0	60.0	54.0	42.8
SD	11.1	10.4	13.0	9.8
Prompted Recall				
mean	73.2	68.1	63.6	55.1
SD	11.3	8.7	12.8	10.2
Free-minus-prompted recall				
SD	9.2	6.5	7.6	9.8

TABLE 4  
CORRECTIONS AMONG THE VARIOUS RECALLS

Retention intervals	Retention intervals			
	Immediate	10 min	1 week	1 month
Free recall				
Immediate	.869 <sup>a</sup>	.885	.751	.732
10 min		.880 <sup>a</sup>	.882	.876
1 week			.890 <sup>a</sup>	.830
1 month				.915 <sup>a</sup>
Prompted recall				
Immediate	.888 <sup>a</sup>	.981	.772	.797
10 min		.905 <sup>a</sup>	.823	.822
1 week			.888 <sup>a</sup>	.883
1 month				.880 <sup>a</sup>
Free (horizontal) by prompted (vertical) recall				
Immediate	.801 <sup>b</sup>	.637	.493	.546
10 min	.845	.726 <sup>b</sup>	.620	.647
1 week	.807	.765	.731 <sup>b</sup>	.721
1 month	.736	.714	.630	.782 <sup>b</sup>

<sup>a</sup> Split-half reliabilities (Cronbach's  $\alpha$ ; Cronbach, 1951).

<sup>b</sup> Same subjects performed both tasks.

particular retention interval, instead of the rank ordering for the four intervals combined, does not greatly increase our accuracy. It may be noted that the coefficient of reproducibility for all four retention intervals combined must be somewhat smaller than the average of the values for the four intervals taken separately, except when the rank ordering of units from most to least likely to be recalled is identical for all intervals. This is because the empirically determined rank orderings maximize the coefficient of reproducibility for their respective intervals.

The average correlation among free prompted recalls (.712) is not as high as among either situation separately. Thus, while the ordering of units is constant over changes in the retention interval it is affected by changes in the retrieval task. The difference in the patterns of recall is, however, not great enough to allow for significant differences in the degree to which the various variables predict the two types of recall.

#### Experiment 4

##### Method

*Subjects.* Sixteen subjects drawn from the same population used in Experiment 3 were paid \$1.50 for their participation.

*Procedure.* The procedure was identical to the 10-min retention interval in Experiment 3, except that a recognition task replaced the free and prompted recalls. The recognition form consisted of a random ordered list of all the orthographically unique words that formed the 47 units of the story (57 words in all) plus an equal number of words, matched for frequency of occurrence in English, which did not occur in the story. The subjects were instructed to check the 57 words they thought actually occurred in the story, guessing if necessary to indicate exactly 57 words. In scoring the forms, a unit was considered to be recognized if any of its component words were recognized.

### Results

Subjects recognized 77.4% ( $SD = 3.4$ ) of the units. In order to compare the number of units recognized with the number of units freely recalled, a high-threshold model (Coombs, Dawes, & Tversky, 1970) was used to correct for guessing. It is assumed that if subjects knew  $X$  units and guessed at the remaining  $47 - X$  units, then they would be scored correct on all the units they knew plus half of the guessed units, or  $X + (47 - X)/2$  units. Using this method of correction 54.8% ( $SD = 6.8$ ) of the units were recognized compared to 60.0% ( $SD = 10.4$ ) recalled. This difference is not significant.

Examining the patterns of recall, we find that the recognition task correlates moderately with the 10-min free recall of the last experiment ( $r = .498$ , Cronbach's  $\alpha = .799$ ). Thus, a change in the retrieval task affects which units are remembered relative to which units are forgotten. In general the recognition task is not predicted as well by the variables. The differences for individual variables, however, again fail to reach significance.

### Discussion

The most theoretically interesting finding of this study is that the relative ordering of units from most to least likely to be remembered was insensitive to changes of six orders of magnitude in the retention interval, even though the total amount recalled decreased significantly. Mathematically this is equivalent to saying that the number of times a unit is remembered by a group is equal to the product of two independent factors. The first factor, which corresponds to the relative ordering of units from most to least likely to be remembered, is different for each unit. As this factor has been shown to be insensitive to the passage of time, it must be determined by encoding and/or retrieval. Changes in the retrieval task did, in fact, alter this factor.

The second factor is a function only of time, or the interference that accompanies it. As this

factor is the same for all units, it affects only the total amount remembered and not the relative ordering of units. Losses in retention are viewed as a passive fading (i.e., time-dependent decrease) in this second factor. Thus, this two-factor model provides a clear definition of what is meant by passive retention.

The paradox between studies that claim that retention is passive and those that claim that over time only the more important units are remembered can now be resolved. The two statements are consistent if, as found earlier, the more important units are also more likely to be remembered. In this case a passive fading of units would result in only the more important units being recalled.

The increase with time of items retained, but not recalled without prompting, suggests that at least part of the orderly fading is due to failures of retrieval as opposed to storage. Further attempts at testing memory under additional retrieval situations would be necessary to try to assess the relative causes of the loss. In this respect, the recognition task proved to have little advantage over free recall, as might have been expected from the mismatch between the encoding and retrieval situations (Tulving & Thompson, 1973).

### SECTION 5: MOTIVATION

The generalization of results over motivational factors is of great practical importance. Most experimental studies of remembering have been performed with highly motivated subjects who were aware that a test would follow. In contrast, most everyday learning does not occur under these circumstances. From laboratory studies specifically designed to study the effects of motivation, we know that motivation affects the amount remembered (Weiner, 1966), but we know less about its effect on what is remembered.

One area in which the effect of motivation on memory has received considerable atten-

tion is that of incidental vs intentional learning. For this distinction, evidence based on how much groups remember indicates a continuity in the two types of learning (Postman, 1964; Wolk, 1974). That is, intentionality seems to affect how much, but not what is remembered. Here, as in the last section, a direct evaluation of this claim will be made.

A second approach to varying motivation is to use two intentional learning groups which have different degrees of interest in remembering a prose passage. Fortunately two such groups were found; one, a group of the experimenter's students in a course on memory; the other, a group of students enrolled in a sociology course.

#### *Experiment 5*

##### *Method*

*Subjects.* The Incidental Learning group consisted of 16 paid Harvard University undergraduates enrolled in an introductory psychology course.

*Procedure.* Subjects were informed that the experiment was intended to test the reliability of a problem solving task. They were told the experiment therefore consisted of three parts: (1) an initial problem solving period, (2) a dummy task, consisting of reading a short story, which was included only to take their minds off the problem solving, and (3) a second problem solving period. Subjects spent 3 min on a set of Thurstone Letter Series Completion Test items, after which they were allowed 30 sec to read the Lincoln story. They were then given 10 min more of completion items before the surprise recall. The subjects were told that the recall was included to check that they did, in fact, take their attention off the problem solving task. They were asked, therefore, to remember the story as accurately as they could.

The subjects showed signs of surprise when asked for the recall, and in later questioning by the experimenter claimed not to have made any special effort to memorize the story. These

subjects will be compared with an Intentional Learning group: the 10 min recall group of Experiment 3, which was drawn from the same subject population.

##### *Results*

As might be expected, the number of units remembered in the Incidental Learning condition was less than in the Intentional Learning condition [43.2%,  $SD = 14.5$ , vs 60.0%,  $SD = 10.4$ ;  $t(30) = 4.50$ ,  $p < .001$ ]. The correlation between the patterns of recall of the two groups, however, is .822 (Cronbach's  $\alpha$  is .887 for the Incidental group and .880 for the Intentional group). The coefficient of reproducibility is .803 for the Incidental group (.639 by chance,  $p < .001$ ), .795 for the Intentional group, and .787 for the two groups combined. Thus, for individuals as well as for group data, the results obtained using the pattern of recall from the memory situation can be generalized to situations in which no special effort is made to memorize the passage.

#### *Experiment 6*

##### *Method*

*Subjects.* The High Interest group consisted of eight paid Harvard University undergraduates enrolled in a course on memory and cognition in which the experimenter taught. The Low Interest group consisted of eight paid Harvard University undergraduates enrolled in a course in sociology with which the experimenter had no contact outside of the experiment.

*Procedure.* The procedure was identical to that used in Experiment 3, except that an unfilled retention interval of 10 weeks was used to allow for possible long term effects.

##### *Results*

The High Interest group recalled almost twice as much as the Low Interest group (37.0%,  $SD = 7.1$  vs 22.9%,  $SD = 15.2$ ;  $t(14) = 2.39$ ,  $p < .05$ ). The correlation among the patterns of recall, however, was .799. This

value is close to the reliabilities of the groups (.778 and .838, respectively). The coefficients of reproducibility are .814 and .910 for the High and Low Interest groups and .848 for the two groups combined. The coefficients of reproducibility obtained by chance are .641, .778, and .685, respectively (all  $p < .001$ ). The high chance level for the Low Interest group is due to the low percentage of units remembered. Thus, again we have significant differences in the amount recalled but little difference in the pattern of recall.

#### *Discussion*

The most striking finding of this section is that the pattern of recall is insensitive to changes over large variations in motivation level, even when there are significant differences in the total amount recalled. Following the same argument used in the case of retention intervals, this is equivalent to saying that the number of times a unit is recalled by a group is equal to the product of two independent factors. The first, which is the same for all units, is a function only of motivational variables. The second, which is different for each unit, is not a function of motivation. While the finding that there is a continuity in processing across levels of motivation is not novel (e.g., Postman, 1964), the simplicity of the experimental method, and the accuracy with which the finding can be stated are.

Two limitations should be made explicit. First, the differences in amount recalled by the High and Low Interest Groups of Experiment 6 may be due to differences between the groups other than differences in motivation. Second, no attempt was made to investigate the effects of differences in motivation resulting from differences in specific attitudes (for instance, political attitudes, Levine & Murphy, 1943). That is, while the overall level of motivation varied, it was assumed that the belief structure and emotional involvement of all subjects was relatively constant with respect to the individual units of the story used.

#### SECTION 6: SUBJECT POPULATIONS

In attempting to find diverse subject populations, age was varied. The oldest available population was a group of men who were taking part in an aging study. A group of eight-graders provide a point around the age of the end of the period of rapid improvement in retention ability (Welborn & English, 1937). The undergraduates from the 10 min recall of Experiment 3 provide a comparison with studies using college students.

Age is clearly not the only difference among these groups. While all the groups were above average intelligence, there are obviously unmeasured differences in countless other dimensions. The problem of assigning differences in the pattern of recall to the various possible differences in the groups will be considered with the results.

#### *Experiment 7*

##### *Method*

All groups read the Lincoln story twice at their normal reading speed, had a 10-min filled retention period, and then recalled the story in writing.

*Eighth-graders.* This group consisted of 16 paid native speakers of English enrolled in an advanced eighth grade of the Quincy Public Schools. With the exception of one 12-year-old, the subjects were all 13. They had a mean I.Q. of 118 ( $SD = 8.5$ ) on the California Achievements Tests' Short Form Test of Academic Aptitude (1970 Edition).

The class was tested as a group by their teacher. The 16 subjects used here were selected at random from the recall protocols obtained. The retention interval was filled with a Thurstone Letter Series Completion Test.

*Older men.* This group consisted of 16 male native speakers of English who had volunteered to take a battery of psychological tests as part of the Normative Aging Study (Bell, Rose, & Damon, 1972). Their average age was 68 years ( $SD = 9.4$ ) with a range from 52 to

86 years. They were above average in intelligence with a mean IQ of 116 ( $SD = 9.6$ ) on the Quick Test (Ammons & Ammons, 1962) and had an average of 12.3 years of formal schooling ( $SD = 3.0$ ).

The subjects were run individually. The retention interval was filled with a description of the series of experiments in which the subject would later be participating.

### Results

The Eighth-Graders, Undergraduates and Older Men remembered 58.9% ( $SD = 10.8$ ), 60.0% ( $SD = 10.4$ ), and 45.3% ( $SD = 6.0$ ) of the units, respectively. The difference between Eighth-Graders and Undergraduates is not significant, while both the Eighth-Graders and the Undergraduates remembered more than the Older men [ $t(30) = 4.25$ ,  $p < .001$ , and  $t(30) = 4.74$ ,  $p < .001$ ].

More interesting than the differences in the amount recalled are correlations among the patterns of recall. The correlation between the Eighth-Graders and the Undergraduates is .893, between the Eighth-Graders and the Older Men is .889, and between the Undergraduates and the Older Men is .847. The split-half reliabilities calculated by using Cronbach's  $\alpha$  are .878, .880, and .930 for the Eighth-Graders, Undergraduates, and Older Men, respectively. Thus, the correlation among the various groups is roughly equivalent to their reliabilities. As there is a negligible probability that the high degree of correlation among the groups is due to a balancing of differences in group and procedural variables, a discussion of these differences becomes superfluous.

The high degree of agreement among the various patterns of recalls holds for the individual subjects as well as for the group data. That is, there appears to be one underlying scale for most- to least-likely to be recalled that holds for all subjects independent of which group they are from. The coefficient of reproducibility for the Eighth-Graders, Undergraduates, and Older Men are .787 (.636 by chance,  $p < .001$ ), .795, and .838

(.581 by chance,  $p < .001$ ), respectively. The coefficient of reproducibility for the three groups combined is .786. That is, given the number of units a subject remembers and the overall rank ordering of units from most- to least-likely to be recalled, exactly which units were remembered can be predicted with an accuracy of almost 80%. Knowing which group a subject came from does not greatly improve this accuracy.

### Discussion

The last three sections all indicate that the pattern of recall is affected by a wide range of variation in experimental conditions in a surprisingly simple way. Except when different retrieval tasks were introduced, the relative value of the units remained unchanged with all units changing by a common multiplier. By summing over units to obtain a score for each subject, many past experimenters have concentrated on the common multiplier and missed much of the regularity that is apparent when a sum is also made over subjects to obtain a score for each unit.

To put the findings mathematically, let  $R_i$  be the proportion of subjects recalling unit  $i$ ,  $f(t)$  be a function only a retention interval,  $f(mot)$  be a function only of motivation level,  $f(Ss)$  be a function only of the general prose memory ability of the subjects, and  $P_i$  be a probability dependent on  $i$  but not on retention interval, motivation level, or the prose learning ability of the subjects. Then a function consistent with the experimental data is  $R_i = f(t) \cdot f(mot) \cdot f(Ss) \cdot P_i$ .

### GENERAL DISCUSSION

By summing over subjects to provide a number correct value for each unit of a story, a rich quantitative description of recall was obtained. This minor modification of existing methods reveals overlooked regularities. Rather than being idiosyncratic, as those who opted for more controlled stimuli feared, the recalls are orderly enough to allow

for the prediction of which units individual subjects would remember. For a given set of conditions all individuals have the same basic ordering of units from most- to least-likely to be remembered, and over a wide range of conditions, including changes in retention interval, motivation level, and subject population this rank ordering is almost constant. The regularity exists in the data collected here and should exist in the data of other experimenters, if it is looked for.

A dozen variables were all found to be statistically significant predictors of recall. The predictive value of these variables and their interrelationships were lawful across a sample of four stories. For the more structured stories subjects' intuitions, gist, and redundancy were the best predictors; for the less structured stories, serial position, and repetition were. Thus, while these results support Bartlett's (1932) view that the overall meaning of a story greatly determines what will be remembered, the results also support Ebbinghaus' (1885) view that the basic processes of memory found where meaning is not important will also function where it is. As usual, the data lie between the theoretical extremes. Katona's (1940) stimulus dimension of senseless connections—meaningful connections, corresponding to a processing dimension of memorization—learning by understanding, provides the conceptual basis for the compromise.

The overall picture of prose memory obtained is one of a process determined by many factors operating on encoding and retrieval to produce recalls that are very similar for many different individuals and many different experimental conditions. While this view is not as simple as having one variable explain everything, it is not as complex as it could be: retention interval, subject population, and motivation level all affect recall in an extremely simple way.

There are several implications for current computer-based, propositional models of prose memory. The finding that the pattern of recall

does not change noticeably with changes in retention interval, and the finding that failures in recall are at least partially due to failures on retrieval, support the commonly used notion of a passive long term memory storage network that does not change over time unless new information is input. Another finding of interest is that over a wide range of subject populations and experimental conditions, properties of the story are important in determining which units will be recalled, but exactly who is recalling the story or exactly what the experimental conditions are is not. Thus, the current emphasis on studying properties of the text rather than properties of the experimental situation is supported. The finding that variables based on surface structure affect recall does not support the use of propositional systems that preserve meaning but lose all surface structure information; surface structure information affects recall. Perhaps more important than these particular findings is the method used to obtain them. If the propositions of a particular computer model are substituted for the units used here, a quantitative description of recall protocols could be obtained which would be worthy of the complexity of the computer model itself. Using this description the model could be modified to account for the data from one set of stories and then tested on other sets of stories.

While four stories is a larger sample than is usually used in studies of prose, it is a rather small sample on which to base generalizations. The regularity found across this sample, however, indicates that the results should hold for all stories of similar length and type. Unfortunately, a theoretical definition of "type" cannot be given. That is, the exact population from which the sample of four passages was drawn cannot be determined theoretically. It is clear that the Gettysburg Address and the Star Spangled Banner do not belong to the population (Rubin, 1977), but it is not clear what does. For this reason the four passages were treated individually and generalizations beyond the stories used here cannot be

supported statistically (Clark, 1973). Fortunately, the method does provide an empirical way of knowing when a story is from a different population; the interrelations among the independent variables and their relations to the pattern of recall will be different, indicating different underlying processes.

## REFERENCE

- AMMONS, R. B., & AMMONS, C. H. *The quick test*. Missoula, MT: Psychological Test Specialist, 1962.
- BARTLETT, F. C. *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press, 1932. (Republished, 1967).
- BELL, B., ROSE, C. L., & DAMON, A. The Normative Aging Study: An interdisciplinary and longitudinal study of health and aging. *Aging and Human Development*, 1972, 3, 5-17.
- BOWER, G. H. Experiments on story understanding and recall. *Quarterly Journal of Experimental Psychology*, 1976, 28, 511-534.
- BRANSFORD, J. D., & JOHNSON, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 717-726.
- BROADBENT, D. E. *In defense of empirical psychology*. London: Methuen & Co., 1973.
- CHILTON, R. J. A review and comparison of simple statistical tests for scalogram analysis. *American Sociological Review*, 1969, 34, 238-245.
- CHRISTIE, D. J., & PETTLOHN, T. F. Students' knowledge of recall-ability of sentences and strategies for recalling textual prose. *Psychological Reports*, 1977, 40, 532-534.
- CLARK, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 335-359.
- COOMBS, C. H., DAWES, R. M., & TVERSKY, A. *Mathematical psychology*. Englewood Cliffs, NJ: Prentice-Hall, 1970.
- CRONBACH, L. J. Coefficient Alpha and the internal structure of tests. *Psychometrika*, 1951, 16, 297-334.
- DAWES, R. M. Memory and distortion of meaningful written material. *British Journal of Psychology*, 1966, 57, 77-86.
- DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, 1957, 54, 180-187.
- DOOLING, J. D., & LACHMAN, R. Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, 1971, 88, 216-222.
- EBBINGHAUS, H. [*Memory: A contribution to experimental psychology*] (H. A. Ruger & C. E. Busenius, trans.). New York: Dover, 1964. (Originally published 1885.)
- FRIES, C. C. *The structure of English*. New York: Harcourt, Brace, & World, 1952.
- GOMULICKI, B. R. Recall as an abstractive process. *Acta Psychologica*, 1956, 12, 72-94.
- GREEN, B. F. A method of scalogram analysis using summary statistics. *Psychometrika*, 1956, 21, 79-88.
- GUTTMAN, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 1968, 33, 469-506.
- JENKINS, J. J. Remember that old theory of memory? Well, forget it! *American Psychologist*, 1974, 29, 785-795.
- JERSILD, A. Primacy, recency, frequency and vividness. *Journal of Experimental Psychology*, 1929, 12, 58-70.
- JOHNSON, R. E. Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 12-20.
- KATONA, G. *Organizing and memorizing: Studies in the psychology of learning and teaching*. New York: Columbia University Press, 1940.
- KAUSLER, D. H. *Psychology of verbal learning and memory*. New York: Academic Press, 1974.
- KENNY, D. A., & RUBIN, D. C. Estimating change reproducibility in Guttman scaling. *Social Science Research*, 1977, 6, 188-196.
- KINTSCH, W. *The representation of meaning in memory*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1974.
- KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, RI: Brown University Press, 1967.
- LEVINE, J. M., & MURPHY, G. The learning and forgetting of controversial materials. *Journal of Abnormal and Social Psychology*, 1943, 38, 507-517.
- LOFTUS, E. F., & SUPPES, P. Structural variables that determine the speed of retrieving words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 770-777.
- MILLER, G. A., & SELFRIDGE, J. A. Verbal context and recall of meaningful material. *American Journal of Psychology*, 1950, 63, 176-185.
- PAIVIO, A., YULLE, J. C., & MADIGAN, S. A. Concrete-ness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, 76 (1, Pt. 2).

- PETERS, H. N. The relationship between familiarity of words and their memory value. *American Journal of Psychology*, 1936, 48, 572-584.
- POSTMAN, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.
- RUBIN, D. C. The effectiveness of context before, after, and around a missing word. *Perception and Psychophysics*, 1976, 19, 214-216.
- RUBIN, D. C. Very long-term memory for prose and verse. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 611-621.
- RUMMELHART, D. E., LINDSAY, P. H., & NORMAN, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.
- STOFFER, S. A., GUTTMAN, L., SUCHMAN, E. A., LAZARFELD, P. F., STAR, S. A., & CLAUSEN, J. A. *Measurement and prediction*. Princeton: Princeton University Press, 1950.
- TINBERGEN, N. *The herring gull's world: A study of the social behavior of birds*. London: Collins, 1953. (Republished, Garden City, NY: Doubleday & Co., 1967).
- TULVING, E., & THOMSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, 80, 352-373.
- UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Chicago: J. B. Lippincott Co., 1960.
- WEINER, B. Effects of motivation on the availability and retrieval of memory traces. *Psychological Bulletin*, 1966, 65, 24-37.
- WELBORN, E. L., & ENGLISH, H. Logical learning and retention: A general review of experiments with meaningful verbal material. *Psychological Bulletin*, 1937, 34, 1-20.
- WHIPPLE, G. M. *Manual of mental and physical tests, part II: Complex processes*. Baltimore: Warwick & York, 1915.
- WOLK, S. The influence of meaningfulness upon intentional and incidental learning of verbal material. *Memory and Cognition*, 1974, 2, 189-193.

#### REFERENCE NOTE

1. FOX, R. *An analysis of memory for English prose*. Unpublished doctoral dissertation, Harvard University, 1968.

(Received November 14, 1977)