VIVIDNESS OF WORDS AND LEARNING TO LEARN IN FREE-RECALL LEARNING[1]

ENDEL TULVING, JOHN A. McNULTY,[2] AND MARCIA OZIER
University of Toronto

ABSTRACT

Eighty-two words of approximately equal frequency-of-occurrence value were rated for vividness (V) and meaningfulness (M) by two independent groups of 100 raters. From this set of words three lists of 16 words each were then constructed. The lists varied in V, but were equal in M. They were used as learning materials in an experiment that was designed (1) to investigate the relation between V and free recall, and (2) to demonstrate learning to learn effects. The results showed that Ss learned lists of higher V more readily than the list of low V, and subsequent lists more readily than the first. Subjective organization (SO) scores were related to both vividness and learning to learning to learn effects.

IT HAS BEEN THOUGHT FOR A LONG TIME that the vividness of an experience is an important determinant of the ease with which the experience can be remembered. In the laboratory, investigations of this problem have usually taken the form of testing the recall for verbal materials varying in vividness. Haagen (1949) had subjects rate pairs of adjectives for several attributes, including vividness of connotation. Miller and Dost (1964) used Haagen's lists of high, medium, and low vivid adjectives in a memory task and found that recognition of previously seen words was a function of their vividness. Bowers (1931) found positive correlations between rated imagery of words and their frequency of recall by groups of subjects, and in another paper (Bowers, 1932) reported very reliable ratings of "distinctness of visual images" for single letters and groups of letters. These scattered experiments suggest that vividness of verbal units can be assessed experimentally and that it is related to retention of these units. The present experiment was designed to explore the relation between vividness and free-recall learning of English nouns.

The study consisted of two parts. In the first, vividness (V) of a set of 82 nouns was measured by a rating-scale procedure. The same set of words was also rated for meaningfulness (M). Rated meaningfulness of

[1]This study was first reported to the Canadian Psychological Association at Montreal, June, 1961. It has been supported by the National Research Council of Canada, Grant No. APA-39, and the National Science Foundation, Grant No. GB-810. It was prepared for publication while the senior author held a National Research Council of Canada Senior Research Fellowship at the Institute of Human Learning, University of California, Berkeley.

[2]Now at Dalhousie University, Halifax, Canada.

nonsense syllables has been shown to be closely related to meaningfulness measured by other kinds of scaling methods (Noble, Stockwell, & Pryer, 1957), and to such variables as judged speed of learning, familiarity, and pronunciability (Underwood & Schulz, 1960).

In the second part of the study, word lists at a fixed level of M and frequency of occurrence, but varying in V, were used to determine the relation between V and free-recall learning. The design of the experiment also permitted an evaluation of learning to learn effects in free-recall learning, a topic on which there exists conflicting evidence (Dallett, 1963; Murdock, 1960).

Measurement of Vividness

Method

A sample of 82 two-syllable nouns, 5 to 7 letters in length, was selected from among the words in the Thorndike-Lorge word book (1944) whose *G* count was between 15 and 19. These words were listed in a haphazard order and mimeographed on two sheets of paper. The two sheets (A and B), together with a face sheet containing instructions, were stapled into three-page booklets used in obtaining ratings of V and M. In half the booklets the sheets were in the order AB, in the other half they were in the order BA.

In the instructions for rating vividness, vividness was defined as "the ease with which you can picture something in your mind." The raters were to indicate how vivid each word was by indicating the position of the word on a seven-point scale on which the rating of 1 corresponded to "no image" and 7 corresponded to "extreme vividness." The instructions for rating meaningfulness were similar to those on vividness. The raters were asked to indicate on a seven-point scale "how meaningful" each word was to them. On the scale, 1 corresponded to "meaningless" and 7 to "extreme meaningfulness." In rating of both V and M, the raters were asked to use the full range of the scale and to treat each word independently.

The ratings were obtained from a total of 200 Ss, consisting of 100 psychology undergraduates (Group 1) and 100 premedical students (Group 2). In each of these groups, 50 Ss rated the 82 words for V (Groups 1V and 2V), and 50 Ss rated the words for M (Groups 1M and 2M).

Results

The rating procedure yielded 100 ratings of V and 100 ratings of M for each of the 82 words. The V score of a given word was defined as the mean of the V ratings from a given sample of raters, and the M score as the mean of the M ratings.[3] In Table I are shown mean V scores and mean M scores for all 82 words and their standard deviations for separate groups of 50 raters as well as the total sample of raters. The important fact in Table I concerns the variability of V and M scores. The variance is greater for V than for M. The *F* test for the testing of the differences

[3]The complete list of 82 words and their V and M scores can be obtained from the senior author.

TABLE I

MEANS AND VARIANCES OF V AND M SCORES FOR 82 WORDS BASED ON DATA FROM FOUR GROUPS OF 50 RATERS

	Variable rated	
Raters	V	M
Group 1	Mean = 4.23 Var = 1.12	Mean = 4.48 Var = .247
Group 2	Mean = 4.27 Var = 1.23	Mean = 4.31 Var = .347
Mean	Mean = 4.25 Var = 1.19	Mean = 4.39 Var = .295

between the two variances based on data from two groups of 100 Ss yields an F (81 & 81 df) = 4.03, $p < .001$.

The variability of M scores in the sample of 82 words is relatively small presumably because of the restricted range of Thorndike-Lorge frequencies of these words. To the extent that frequency and meaningfulness are correlated, restriction of the range of frequencies is expected to reduce the variability of M scores. The restriction of the frequency range seems to have much less effect on V, however. The observation that the variance of V scores is approximately four times greater than the variance of M scores can be regarded as evidence for the absence of a strong relation between V and M.

More important than the over-all means and variances of the V and M scores, however, is the evidence pertaining to their reliability and their intercorrelation across the sample of 82 words. The product-moment correlation coefficient between two sets of 82 V scores, one from Group 1V and the other from Group 2V, was 0.941. When the Spearman-Brown correction formula was applied to this statistic, the reliability coefficient of V was estimated as 0.970. The product-moment correlation coefficient between two sets of M scores, one from Group 1M and the other from Group 2M, was found to be .818. The application of the Spearman-Brown formula yielded an estimate of the reliability coefficient of 0.900. The lower coefficient of reliability for M than for V is attributable to the smaller variance of M scores than of V scores, as shown in Table I. We can estimate what the reliability coefficient of M scores would be if their variance were the same as that of V scores. When the appropriate correction is applied (Guilford, 1942, p. 281), the coefficient of reliability of M becomes 0.975, closely resembling that of V. The product-moment correlation between V and M scores was found to be 0.420 for Groups 1V and 1M, and 0.541 for Groups 2V and 2M. Although these statistics are sig-

nificantly different from zero, their magnitude indicates that V and M are only partly related and that each of the two variables taps underlying processes not tapped by the other.

In view of the high reliability of V and M scores and the fact that the variance of one accounts for only a part of the variance of the other, it seemed justifiable to construct word lists varying in V, but holding constant M, and to investigate ease of free-recall learning for such lists.

Vividness and Learning to Learn in Free-Recall Learning

Design

Three lists of 16 words used in this experiment differed in V of their constituent words, but were approximately equal in rated M as well as in Thorndike-Lorge frequency. Twenty-four *Ss* learned all three lists, each for eight study and recall trials. The order of lists was completely counterbalanced. Performance over successive trials was examined as a function of V and stage of practice. Subjective organization scores (Tulving, 1962) were also calculated and related to the two independent variables and to performance scores.

Method

Lists. From among the 82 words for which both V and M scores were available three lists of 16 words were selected. These lists are shown in Table II. Table II also

TABLE II

Experimental Lists

High V			Medium V			Low V		
Word	V	M	Word	V	M	Word	V	M
Apron	5.17	4.00	Abode	4.13	4.06	Buyer	2.94	3.93
Balloon	5.63	4.03	Bucket	4.50	3.88	Crisis	2.10	5.23
Bunny	5.23	3.93	Builder	4.00	4.28	Entry	2.08	3.71
Butler	5.79	3.79	Cargo	4.57	4.03	Founder	1.85	3.92
Cabbage	5.32	4.06	Fiber	3.96	3.50	Output	1.50	3.77
Camel	5.79	3.88	Hamlet	4.83	4.03	Patron	2.70	3.50
Chorus	5.63	4.43	Handful	4.36	3.43	Renown	1.33	3.94
Cigar	6.53	4.41	Madame	4.61	4.27	Routine	1.28	4.64
Circus	6.08	4.54	Pebble	4.76	4.15	Rover	2.11	3.30
Comet	5.81	3.96	Porter	4.64	3.79	Rumour	2.20	4.57
Granny	5.00	4.17	Pudding	4.61	4.05	Session	2.86	3.83
Jungle	6.00	4.36	Summit	4.77	4.46	Surplus	1.83	4.43
Lantern	5.43	4.32	Thicket	4.54	3.75	Tariff	1.50	3.61
Rainbow	6.13	4.94	Trainer	4.25	4.19	Topic	1.50	4.03
Runner	5.64	4.61	Veteran	4.07	4.40	Treason	1.50	4.70
Satin	5.42	3.62	Voter	4.32	5.15	Vigour	2.70	4.44
Mean	5.66	4.19	Mean	4.43	4.09	Mean	2.00	4.10

gives the V and M scores for each word, based on samples of 100 raters. In each list the words were first arranged into a random word order and then seven other word orders were systematically constructed such that no word occurred in a given serial position, preceded any other word, or followed any other word more than once in the total set of eight orders. The sequence of word orders on eight trials was different for each *S* learning a given list.

Counterbalancing of lists. With three lists (high, medium, and low) there are six possible orders in which they can be given to Ss: HML, HLM, MHL, MLH, LHM, LMH. Each of these list orders was given to four randomly determined Ss. This counterbalancing of lists means that each of the three lists (H, M, and L) was learned by an equal number of Ss at each of three stages of practice (first, second, or third list).

Subjects. The Ss were 24 female second-year general arts students at the University of Toronto who had not participated in any other verbal learning experiments and were naïve with respect to the nature and the purpose of the present experiment.

Procedure. All Ss were tested in individual sessions. Each S was told that she would be learning three lists of English words, one after another; that she would be given eight trials of practice on each list; that the words would be in a different order from trial to trial; and that her task was to recall as many words as possible from a given list after each trial, but that the order in which she recalled the words did not matter. The words were presened on a memory drum at the rate of one sec. per word. At the end of the input phase of each trial S was given 60 sec. to record her recall on a sheet of paper. The intertrial interval was approximately five sec. The *E* collected the recall sheet from S after each trial. When eight trials on a given list were concluded, the S was given two min. of rest before starting on the next list.

Results

The three learning curves associated with three levels of V are shown in Fig. 1. There is a clear separation between HV and LV curves over all eight trials indicating that HV words are recalled more readily than LV words. The recall of MV words falls between that of HV and LV words on most trials.

The three learning curves associated with three stages of practice are shown in Fig. 2. These curves are remarkably similar to the three curves shown in Fig. 1, with the curve of the first list comparable to the curve of the LV list, the second to MV, and the third to HV. This close correspondence is reflected in the mean recall scores over all eight trials: 12.03, 12.64, and 13.19 for the LV, MV, and HV lists, respectively, and 12.04, 12.77, and 13.05 for the first, second, and third lists, respectively. Thus variations in V in the present experiment had approximately the same effect on recall as did variations in stage of practice.

The arcsin transformed recall scores were evaluated statistically by an analysis of variance based on the cross-over design (Cochran & Cox, 1957). This analysis showed that the main effects associated with trials, V, and stage of practice were all significant at better than the .001 level, and that none of the interactions was significant.

In evaluating the recall data from the present experiment, possible "ceiling effects" must be kept in mind. In learning the HV list, 14 subjects out of 24 reached a criterion of two successive perfect trials during the total of eight trials given to all subjects, while only four subjects did so in learning the LV list. Recall performance on the HV list, therefore, is

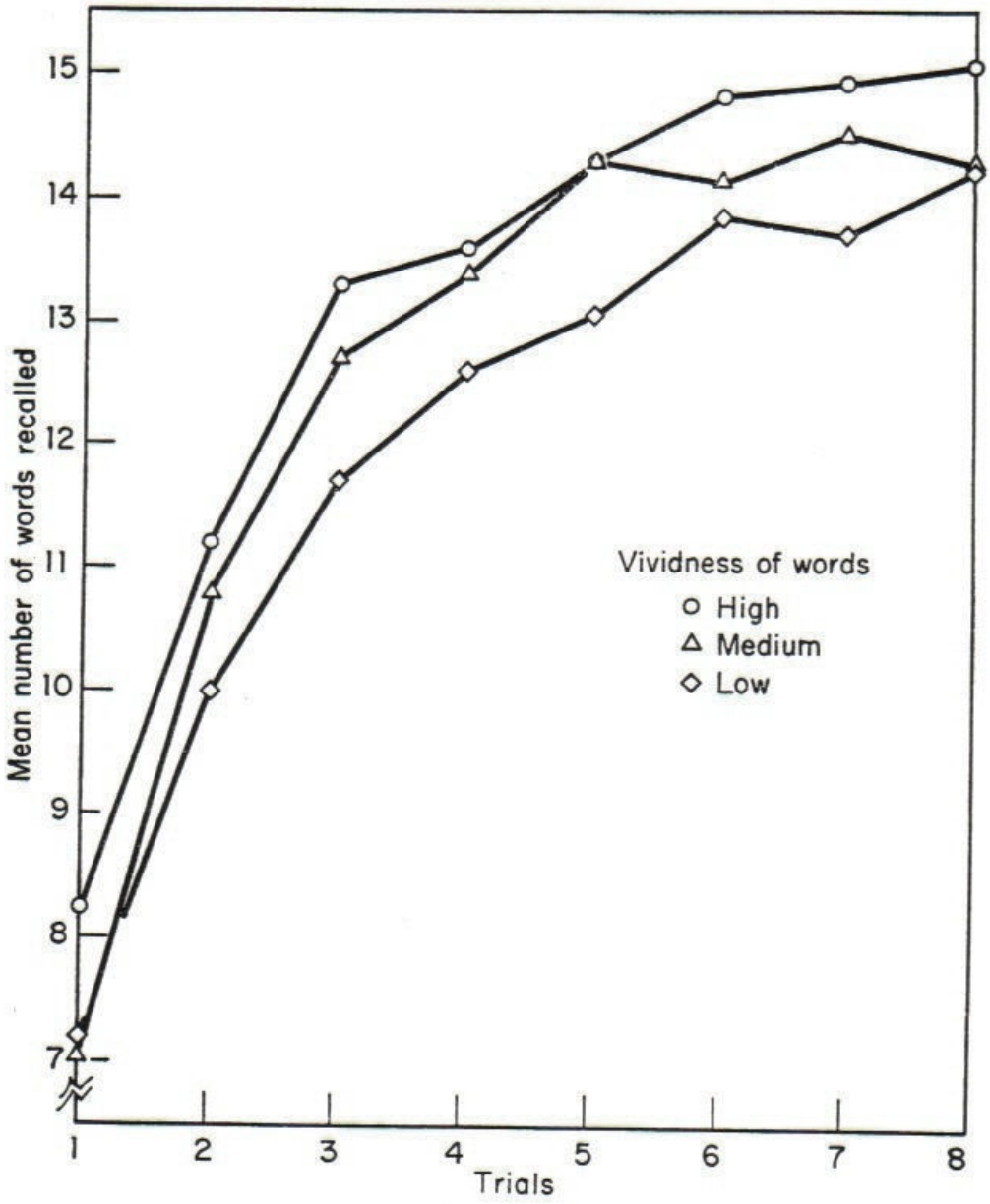

FIGURE 1. Mean number of words recalled as a function of trials for three lists differing in word vividness.

probably artifactually attenuated because of limited list length. The same consideration applies to learning to learn effects. Only five subjects reached the two-trial criterion in learning the first list, while 12 did so in learning the third list.

A subjective organization (SO, Lag 0) score, based on the order of recalled words on successive trials, was also calculated for each subject learning a given list (viz. Tulving, 1962). Each score was based on the data from the total block of eight trials. Thus each subject had three SO

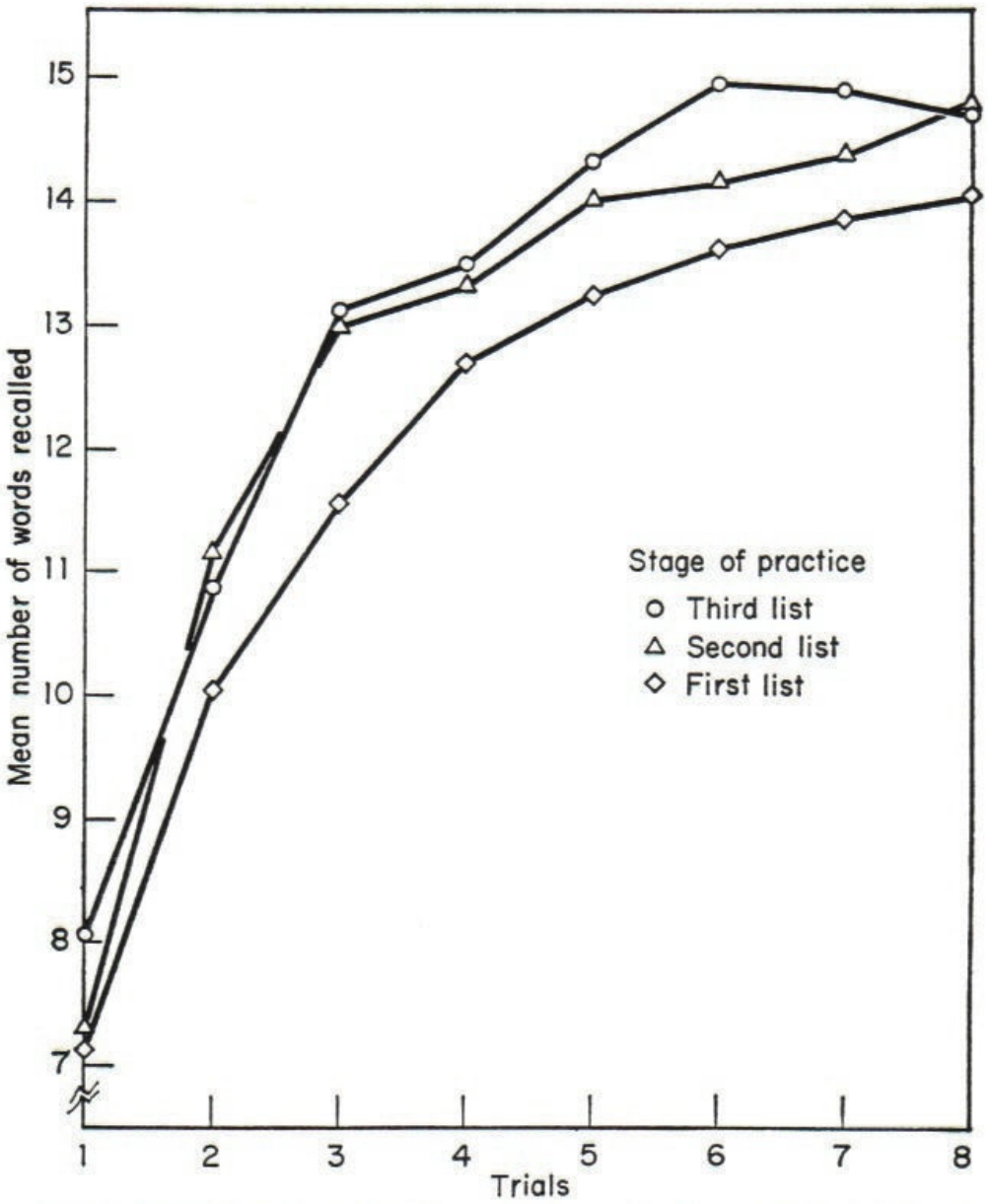

FIGURE 2. Mean number of words recalled as a function of trials for lists at three stages of practice.

scores, one for each of the three lists. The mean SO scores were .219, .277, and .264 for the LV, MV, and HV lists, respectively, and .208, .263, and .289 for the first, second, and third lists, respectively. Analysis of variance showed that the differences associated with stage of practice, F (2,44) = 6.82, and with V, F (2,44) = 3.80, were both significant, p values being .01 and .05, respectively.

Finally, we computed Spearman rank-order correlation coefficients (rho) between mean recall scores and SO scores for the 24 subjects. Rho

was .42, .86, and .76 in the LV, MV, and HV lists, respectively, and .50, .78, and .73 in the first, second, and third lists, respectively. All these correlation coefficients are significant at the .01 level, with the exception of a rho of .42 in the LV list, which is significant at the .05 level. Thus, there is evidence for a tendency for subjects' recall scores to parallel SO scores, this tendency being somewhat stronger in the case of MV and HV lists than in the LV list, and stronger in the second and third lists than in the first.

Conclusions and Discussion

The following conclusions seem to be justifiable on the basis of the findings of this study: (1) It is possible to obtain reliable ratings of vividness and meaningfulness of words from groups of raters. (2) Vividness and meaningfulness are only partly related to each other. (3) Vividness is directly related to the ease with which words equated for frequency and meaningfulness are learned in the free-recall situation. (4) More vivid words are not only learned more readily than less vivid words but their recall is also organized to a somewhat greater extent. (5) Learning to learn effects occur in the course of learning of three successive lists. (6) Increasing efficiency of performance over successive lists is accompanied by an increase in the degree of subjective organization.

That reliable ratings of vividness of words can be obtained is hardly surprising, despite the fact that apart from its operational definition the concept may appear vague to some. Both Haagen (1949) and Bowers (1931, 1932) have reported similar data for other types of verbal units. What is perhaps more interesting is the relatively weak relation between M and V. It is interesting, because there have been very few reports in the literature about characteristics of materials which can be measured reliably, which are related to ease of learning or recall, and which at the same time are relatively independent of the cluster of interrelated variables such as familiarity, frequency, and meaningfulness (Underwood & Schulz, 1960). Haagen (1949) did report even lower correlations between vividness of pairs of adjectives on the one hand and their similarity of meaning and closeness of associative connection on the other hand, but there has been little experimental work done with his materials on the problem of vividness. The study of Miller's and Dost's (1964), mentioned earlier, is the only one of which we are aware. The present data on the role of vividness in recall are in good agreement with those of Miller's and Dost's.

It is not immediately obvious from the present study why more vivid words are learned more easily in the free-recall learning situation. But

the fact that the recall of more vivid words was also organized by subjects to a greater extent than that of less vivid words is compatible with the hypothesis that vividness or picturability is an important component of meaning of words that affects the ease with which words can be grouped into higher-order memory units (Miller, 1956; Tulving, 1964). Even though the number of units that the subject can retrieve from the storage at any given time is limited at some fixed value, the number of words recalled varies directly with the average size of the units (cf. Tulving & Patkau, 1962). The higher word-recall in the HV list than in the LV list can thus be regarded as reflecting the existence of larger memory units in the HV list than in the LV list.

The main weakness of that part of the present experiment that deals with the relation between V and learning—a weakness shared by all other published experiments investigating the relation between some characteristics of the material and ease of learning or recall—lies in the fact that the data reveal only a correlation between vividness and learning. Although we held constant word frequency and rated meaningfulness, and although it appears reasonable that as a consequence other variables related to meaningfulness and frequency may also have been minimized, it is still possible that V was confounded with other variables.

One such variable, for instance, may have been abstractness (or concreteness) of words. Gorman (1961) used a two-point rating scale to assess a large number of words on an abstract-concrete dimension. She found short-term recognition to be better for concrete than for abstract words. In the sample of 82 words used in the present study, there were 30 that overlapped with Gorman's sample. Of these seven were abstract, as defined by Gorman, and 23 were concrete. On our V scale, the seven abstract words had the mean V score of 2.93 (SD = 1.12) and the mean V score of the 23 concrete words was 4.81 (SD = 1.01), a highly reliable difference. Thus it appears that V may be correlated with the abstractness-concreteness of words.

Another type of variable that might correlate with vividness is associative relatedness of words. This variable, measured in a number of ways, has also been shown to be related to the recallability of verbal materials (Marshall & Cofer, 1963). Quite possibly our three lists differed from one another with respect to this variable, but in absence of available free-association data for the words that we used, the relation between vividness and associative relatedness could not be assessed.

The reliable finding of learning to learn (LTL) effect in the present experiment corroborates Dallett's (1963) findings for a free-recall task in which multiple recall tests on the same list were given, but it appears to contradict Murdock's (1960) conclusion that neither LTL nor warm-up

occurs in multitrial free-recall learning. Since SO scores in the present experiment also increased over successive stages of practice, it may be that LTL effects occur only in situations in which subjects can learn how to organize the materials to be recalled. Murdock (1960) used a procedure in his experiments which may have minimized learning of subjective organization. Subjects in his experiments were given recall sheets provided with spaces for words beginning with different letters, and they were asked to write their recalled words in corresponding spaces on the recall sheets. This requirement may have induced subjects to adopt different strategies of memorizing than grouping of words according to various aspects of their meaning as it occurs in subjective organization (Tulving, 1962). To the extent that subjective organization may play only a minor role in one trial recall, the failure to demonstrate LTL effects in free-recall studies in which only one recall trial is given (e.g., Dallett, 1963) also seems to be consistent with the view that LTL may depend on interlist improvement in subjective organization.

Résumé

Deux groupes indépendants de 100 juges chacun évaluent le niveau de vividité (V) et d'intelligibilité (M) de 82 mots de fréquence d'usage approximativement égal. Trois sous-listes de 16 mots sont formées à partir de cet ensemble de mots. Ces listes varient quant à V, mais sont égales quant à M. Elles constituent le matériel d'apprentissage d'une expérience visant (1) à chercher la relation entre V et le simple rappel et (2) à illustrer des effets d'apprentissage à apprendre. Les résultats montrent que les listes à vividité élevée s'apprennent mieux que les listes à vividité inférieure, et que la première liste s'apprend plus difficilement que les suivantes. Mise en relation de cotes d'organisation subjective (SO) avec le niveau de vividité et avec les effets d'apprentissage à apprendre.

REFERENCES

Bowers, H. Memory and mental imagery. *Brit. J. Psychol.*, 1931, **21**, 271–82.

—— Factors influencing visual imagery for letter groups. *Amer. J. Psychol.*, 1932, **44**, 775–9.

Cochran, W. G., & Cox, Gertrude M. *Experimental designs.* Second edition. New York: Wiley, 1957.

Dallett, K. M. Practice effects in free and ordered recall. *J. exp. Psychol.*, 1963, **66**, 65–71.

Gorman, Aloysia M. Recognition memory for nouns as a function of abstractness and frequency. *J. exp. Psychol.*, 1961, **61**, 23–9.

Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942.

Haagen, C. H. Synonymity, vividness, familiarity, and associative value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–63.

Marshall, G. R., & Cofer, C. N. Associative indices as measures of word relatedness: a summary and comparison of ten methods. *J. verb. Learn. verb. Behav.*, 1963, **1**, 408–21.

MILLER, G. A. Human memory and the storage of information. *IRE Trans. Inform. Theory*, 1956, **2**, 129–37.

MILLER, MARILYN E., & DOST, JOYCE A. Stimulus vividness and anxiety level in intentional-incidental learning. *Psychol. Rep.*, 1964, **14**, 819–25.

MURDOCK, B. B., Jr. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–34.

NOBLE, C. E., STOCKWELL, F. E., & PRYER, M. W. Meaningfulness (m') and association value (a) in paired-associate syllable learning. *Psychol. Rep.*, 1957, **3**, 441–52.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.*, 1962, **69**, 344–54.

—— Intratrial and intertrial retention: notes towards a theory of free recall verbal learning. *Psychol. Rev.*, 1964, **71**, 219–37.

TULVING, E., & PATKAU, JEANNETTE E. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canad. J. Psychol.*, 1962, **16**, 83–95.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.

(*First received February 22, 1965*)

PRINTED IN CANADA