

# Ratings of stimulus encoding similarity as related to concreteness and meaningfulness

JOHN W. CRAGHEAD and FRANK W. WICKER  
*University of Texas, Austin, Texas 78712*

Subjects rated nouns varying in both meaningfulness-frequency (M) and concreteness (C) on the within-occasion similarity, the between-occasion similarity, and the "uniqueness" of images produced to them, and learned paired associates using rated nouns as stimulus items. In a second study subjects were given recognition training on the previously rated nouns. Within-item encoding similarity (defined by ratings of successive images to the same word) was found to be positively related to both M and C, but between-item similarity of encodings (defined by the uniqueness ratings) was affected positively by C and negatively by M. Recall was related to C, to M, and to their interaction. There was some evidence that between-item similarity of encodings affects recognition, but no clear retention effect for within-item similarity. Results were discussed in terms of certain components and extensions of the encoding variability hypothesis.

The encoding-variability hypothesis, as developed by Martin (1968, 1972), contains several component ideas, including: (1) encodings for a particular stimulus can vary over occasions, (2) this variation can have either a negative or a positive effect on learning, depending on the particular task context, and (3) variability is an inverse function of stimulus meaningfulness (M). The negative effect on learning was predicted for any task having recognition as a component, because change of encoding from study to test will disrupt recognition. A positive effect was predicted for the second phase of an A-D, A-B paradigm, where a stimulus-coding change can reduce negative transfer or interference.

Some studies have provided data supporting the hypothesis (e.g., Butler & Merikle, 1970; Hasher & Johnson, 1975; Merryman & Merryman, 1971), but other data have been negative or mixed (e.g., Davidson, Schwenn, & Adams, 1970; Hashtroudi & Johnson, 1976; Winograd & Geis, 1974). Evaluation of the literature is difficult because studies differ on the directness of relation between theory and data. In the present research, rating indices of encoding similarity were obtained for stimulus items.

Until recently, most researchers defined encoding variability in terms of stimulus selection. The present study dealt with visual encodings, defined by ratings about visual images to the stimuli. Also, variability has usually been defined by number of different encodings or probability-density functions over encodings. In this

Parts of this paper are based on a PhD dissertation by the senior author, supervised by the junior author, in the Department of Educational Psychology, University of Texas at Austin. The authors are grateful for the invaluable assistance of the other Committee members: Robert K. Young, Wilson A. Judd, and Jackson B. Reid. Requests for reprints should be sent to Frank W. Wicker.

paper a factor is considered which provides another operationalization of the general notion of variability among encodings: the similarity of different encodings to the same stimulus. It may not be precise to speak of "the probability of the same encoding" occurring on two occasions; it may be more informative to consider any two encoding events on a continuum of similarity. Consideration of a related variable—the similarity of encodings produced to different stimuli on the list—allows integration with the literature on encoding uniqueness (Lesgold & Goldman, 1973; Moscovitch & Craik, 1976). It was predicted that high within-item similarity would, through a process analogous to stimulus generalization, enhance reliable responding to a stimulus, while high between-item similarity would, through interitem interference, reduce retention.

We were also concerned with other stimulus attributes, such as concreteness (C). It can be claimed that the superiority of concrete over abstract stimuli results from greater encoding variability with the latter (Paivio, 1971, p. 287). Experiment 1 was designed to look at both within- and between-item encoding similarity for stimuli varying factorially in both C and M.

## EXPERIMENT 1

### Method

**Subjects.** Ninety-six undergraduate students of educational psychology participated in the study for course credit.

**Word selection.** Four types of words were chosen from the Paivio, Yuille, and Madigan (1968) norms for 925 nouns. The C and M scores from these norms were converted to standard (*z*) scores, and high-concrete/high-meaningful (HH) words were chosen from those having the highest sum of the two *z* scores, with the restrictions that the two scores differ by less than one unit and that no two words chosen be synonymous. Low-concrete/low-meaningful (LL) words were chosen from those having the smallest sum of *z* scores, with the same restrictions.

Low-concrete/high-meaningful (LH) and high-concrete/low-meaningful (HL) words were chosen from those with the greatest difference between the two z scores, subject to the conditions that one be positive and the other negative and that no two words chosen be synonymous. In all cases, highly unusual, difficultly defined words were avoided, as were a few words whose M obviously varies with geographical location. Otherwise, words were chosen entirely on the basis of their position on the objective indices described above.

**Tasks.** All subjects performed the following five tasks.

**Describe-rate task.** Subjects were given a 32-page booklet with one word at the top of each page. Eight words of each of the four types (HH, HL, LH, and LL) were presented in random order in the booklet. For each word presented, subjects were asked to form three "mental images" or "mental pictures" and to write a very brief description of each image. They then rated the similarity of the two images of each pairwise comparison among the three: the first and second, first and third, and the second and third images produced. Ratings were circled on a 9-point scale from very low difference to very high difference between the two images. Some pilot subjects had reported that occasionally several images had come to them simultaneously in an initial cluster; to evaluate this phenomenon, subjects were asked to circle the word YES after their description of each image which was one of such an initial cluster and to circle NO otherwise. Number of yeses thus provided a measure of initial-cluster size. The pilot study had suggested that many subjects had difficulty with instructions like those used here; therefore, with the instructions of the present study extensive stress was given to the difference between word associations and descriptions of visual images, and to the need to avoid visually or verbally mediated chain associations.

Subjects were given 60 sec to form and describe three images, circle YES or NO for each, and give a difference rating to each of the three image pairs for a given word. Every 60 sec the experimenter called out the number of the following page, whereupon subjects were required to turn immediately to the next page and repeat the procedure.

**Recall-rate task.** The 32 word stimuli were presented again, in a new random order, one at the top of each page of a rating booklet. Subjects were asked to form and describe three images for each word, then to rate the similarity of each image to the images they had produced to that word in the describe-rate task, from IDENTICAL to an image they had produced before to completely DIFFERENT from any they had produced before. Ratings were circled on a 9-point scale. Subjects had 30 sec to form, describe, and rate three images on each page of the booklet.

**Uniqueness-rating task.** Each subject was given a booklet with 32 words printed per page. The words were new but had the same factorial combination of M and C values as those used in the two earlier rating tasks. Subjects were asked to form a single image in response to each word, to write a brief description of it, and to rate the "uniqueness" of that image in relation to all images they had produced during the previous rating tasks. Uniqueness was defined in terms of difference from all previous images; ratings were on a 9-point scale ranging from low to high uniqueness. Subjects were given 20 sec to look at each word, form and describe an image, and rate its uniqueness.

**Filler task.** Subjects rated the pleasantness of 56 line drawings and 56 words for 4 sec each. None of these items corresponded to the words used in other rating tasks. The function of this task was to consume time between two of the image-rating tasks (see Design).

**Paired-associate learning.** Four trials of PAL were given by the study-test method. The words used in the describe-rate and recall-rate tasks were stimulus items for the 32 word pairs presented. For each type of stimulus word, half of the response words were HH and half were LL. Items were presented 3 sec

for study and 5 sec for recall, by a Carousel slide projector automatically controlled by an electric timer.

**Design.** There were two randomly determined orders of presentation of items for each rating task; these were different from the two random orders for other tasks. Half of the subjects worked with each task in one of its two orders and the other subjects worked with the second set of orders. The rating tasks were always presented in the order: describe rate, filler, recall rate, uniqueness rating. For half of the subjects of each item-order group, PAL preceded this rating sequence; for the other half, PAL followed it. With subjects divided by both task order and item order within tasks, there were four groups of subjects differing in terms of order of presentation of materials. There were 24 subjects in each of these groups.

## Results

Analyses were performed on rating scores as a function of C and M, and on recall as a function of C, M, and rating scores. Mean rating scores and PAL scores are given for each type of stimulus in Table 1. With the describe-rate tasks, an inverse relationship was found between difference ratings and both C [ $F(1,92) = 9.61$ ,  $MSe = .73$ ,  $p < .01$ ] and M [ $F(1,92) = 12.99$ ,  $MSe = 1.33$ ,  $p < .001$ ]. That is, successive encodings were more similar with concrete and with meaningful stimuli. The same pattern of results was obtained with the recall-rate task for both C [ $F(1,92) = 12.29$ ,  $MSe = .89$ ,  $p < .001$ ] and M [ $F(1,92) = 8.14$ ,  $MSe = 1.01$ ,  $p < .01$ ]. For both tasks, high M was associated with greater encoding similarity, which can be interpreted as less encoding variability [as postulated by Martin (1968, 1972)]; this is also true of high C.

On the uniqueness task, both C [ $F(1,92) = 19.34$ ,  $MSe = 1.14$ ,  $p < .001$ ] and M [ $F(1,92) = 4.44$ ,  $MSe = 1.61$ ,  $p < .05$ ] produced significant effects, but in this case the two variables operated in opposite directions. Uniqueness was related to high C but to low M. That is, high-concrete but low-meaningful stimuli were rated as producing encodings which were most distinctively different from encodings to other stimuli. Also, there was an interaction of C by M such that the effect of M on uniqueness ratings was greater with concrete than with abstract items [ $F(1,92) = 4.76$ ,  $MSe = .79$ ,  $p < .05$ ].

High-M stimuli evoked larger clusters of reported initial images than low-M stimuli [ $F(1,92) = 64.97$ ,  $MSe = .02$ ,  $p < .001$ ], but this variable was not affected

**Table 1**  
Mean Difference Ratings, Uniqueness Ratings, Initial-Cluster Size, and Number Correct Per Item in Four Trials of PAL for Four Types of Stimulus Item

	High C		Low C	
	High M	Low M	High M	Low M
Describe-Rate Difference	4.92	5.38	5.23	5.62
Recall-Rate Difference	3.66	3.96	4.00	4.29
Uniqueness Rating	5.99	6.46	5.71	5.79
Initial-Cluster Size	.44	.32	.44	.30
PAL	2.55	2.36	2.22	1.66

by C. There is a suggestion here of a direct relationship between number of verbal associations and number of immediate "visual" associations to a stimulus.

Number correct in PAL was affected by C [ $F(1,92) = 166.74$ ,  $MSe = .15$ ,  $p < .001$ ] and M [ $F(1,92) = 90.01$ ,  $MSe = .14$ ,  $p < .001$ ], and there was a C by M interaction [ $F(1,92) = 21.66$ ,  $MSe = .15$ ,  $p < .001$ ]. As Table 1 indicates, the effect of each of the stimulus attributes was more evident when stimuli had low values on the other.

In these analyses only two significant effects involving order of presentation were obtained. M had a smaller effect on PAL performance when PAL was last [ $F(1,92) = 3.39$ ,  $p < .05$ ], perhaps reflecting acquired M for low-M items from the prior rating task. Also, giving PAL first led to greater difference ratings on the recall-rate task than giving PAL last [ $F(1,92) = 6.12$ ,  $MSe = 6.45$ ,  $p < .001$ ]. The latter effect may simply reflect the fact that subjects had more previous encodings to compare present images to when the PAL task had preceded these ratings. At any rate, PAL-rating order did not interact with C or M effects; that is, these effects were not dependent on whether ratings preceded or followed PAL (cf. Stoff & Eagle, 1971).

Intercorrelations among the measures were obtained, correlating across mean values for each stimulus item. Correlations were computed separately for each PAL-rating order, but there were no significant differences between corresponding values of  $r$  in the PAL-first and PAL-last groups, so all subjects were used to obtain the correlations presented in Table 2.

Correlations taken over homogeneous subgroups of stimuli were smaller than those in Table 2; for example, the correlation between recall-rate similarity and PAL was only  $-.145$  for concrete stimuli and  $-.215$  for abstract stimuli. Ross (1968) pointed out that correlations can occur between means taken across subjects which are larger than the corresponding value for any given subject. Therefore, a separate intercorrelation matrix was computed across stimulus items for each subject; the  $\chi^2$  statistic was used to evaluate whether particular correlations were significantly more often positive than negative in the group of subjects. By this analysis, initial-cluster size was found to be significantly related to describe-rate similarity [ $\chi^2(1) = 26.32$ ,  $p < .001$ ] and to recall-rate similarity [ $\chi^2(1) = 4.59$ ,  $p < .05$ ]. The smaller the size of the original cluster for a given item, the greater the difference ratings tended to be for that item. Describe-rate similarity and recall-rate similarity were also significantly correlated [ $\chi^2(1) =$

$12.17$ ,  $p < .001$ ]. Total number correct in PAL was only marginally related, however, to recall-rate similarity [ $\chi^2(1) = 3.76$ ,  $p = .05$ ], to initial-cluster size [ $\chi^2(1) = 3.01$ ,  $p = .08$ ], or to describe-rate similarity [ $\chi^2(1) = 3.41$ ,  $p = .06$ ].

## EXPERIMENT 2

Although rated similarity of successive encodings related strongly to M and C in Experiment 1, the ratings were not powerful predictors of learning. It may be that the PAL procedure employed was not very sensitive to similarity effects. Perhaps a recognition task would provide a more sensitive measure, because variability is assumed to effect cued recall through a component recognition process. Also, variability might have a greater effect at faster rates of presentation.

### Method

**Subjects.** Sixty-nine undergraduate students of educational psychology served as subjects.

**Materials and procedure.** The 32 words were projected at a 1.3-sec rate during each of three study trials. On each test trial, subjects had 3 sec to rate their recognition of each of 64 words projected. They circled Y (yes) on an answer sheet to indicate that they had seen the word on the target list during the study trials or N to indicate that they had not. The 32 words on the target list were those rated and used as PAL stimuli in Experiment 1; the 32 distractors were words that had received uniqueness ratings in Experiment 1.

### Results

Scores were number of Ys circled for target items and number of Ns circled for distractors. A three-way ANOVA treated C, M, and item type (targets vs. distractors) as factors. There was significance for C [ $F(1,68) = 90.81$ ,  $MSe = 4.13$ ,  $p < .001$ ] but not for M ( $F < 1$ ) or C by M ( $F < 1$ ). An effect for item type suggested a bias toward Y ratings [ $F(1,68) = 10.50$ ,  $MSe = 14.81$ ,  $p < .01$ ].

The findings of primary interest were correlations over words, between performance in this study and ratings of Experiment 1. Recognition ratings for target words correlated .01 with initial-cluster size, .08 with describe-rate similarity, and .21 with recall-rate similarity. None of the correlations was significantly different from zero. The recognition ratings for a single word reflect bias as well as recognition sensitivity for that word, but there is no apparent reason why bias would operate systematically to reduce a relationship with prior similarity ratings; thus, there is no evidence in this study that rated similarity of successive images affects retention. Correct rejection of distractors, on the other hand, correlated .44 with their uniqueness ratings ( $p < .01$ ). The correlation was positive with 51 of 69 individual subjects [ $\chi^2(1) = 14.84$ ,  $p < .001$ ]. Thus, there is evidence that between-item similarity, or uniqueness, affects recognition ratings but no corresponding evidence for within-item similarity in Experiment 2.

Table 2  
Intercorrelations Among Dependent Variables Over  
Stimulus Items for All Subjects

	Describe-Rate Similarity	Recall-Rate Similarity	PAL
Cluster Size	.777	.535	.364
Describe-Rate Similarity		.638	.250
Recall-Rate Similarity			.322

## DISCUSSION

These data support the hypothesis that stimulus M is inversely related to encoding variability, as defined by rated differences among successive visual encodings to the same stimulus. This result occurred despite the relatively narrow range of M (i.e., only easily defined words were employed) and is consistent with the findings of Butler and Merikle (1970), despite the different definition of M and different dependent variables employed. Also, effects were of comparable magnitude with the C variable, which was found to relate to similarity of successive encodings in the same way as M.

Little evidence was obtained, however, for a relationship between similarity of encodings and recall. In contrast, both C and M had a large effect on recall, which might imply that consideration of other factors is necessary to explain the effects of these two attributes (Galbraith, 1975). Among likely alternate factors are between-item similarity (suggested by Experiment 2) and other nonrelational properties of encodings.

Interpretation must be highly tentative, however, because (1) there are previous data which indicate that retention can be affected by similarity of encoding between study and test (e.g., Bower & Holyoak, 1973), (2) only visual coding was considered in the present study, thus tapping only one aspect of similarity of representation, and (3) it is not clear to what extent similarity of encoding under instructions to produce several different images will reflect similarity of encoding under standard learning conditions.

Although within-item encoding similarity related in the same way to C and M, encoding uniqueness related in opposite ways. High-C and low-M stimuli were associated with more unique encodings, a result which may relate to the fact that superior recognition has been found with concrete stimuli and with low-frequency stimuli (Gorman, 1961). Note that the relation of M to uniqueness is opposite that predicted by Martin (1968, p. 428), but he was talking about nonsense trigrams, where M is typically measured by the Archer rather than the Noble procedure; also, high-M trigrams are easier to recognize than low-M trigrams.

The uniqueness effect with C is consistent with Salz's hypothesis that abstract concepts are forgotten more rapidly than concrete concepts because of their diffuse nature (Klein & Salz, 1976), but it appears to contrast with Paivio's (1965) finding of no difference in verbal associative overlap as a function of noun C. The ratings used here could have given a different result because they emphasized encoding rather than association or because they emphasized visual rather than verbal processing.

In summary, the rating measures developed relate to stimulus attributes such as M and C in ways that are consistent with theoretical predictions, but, except for uniqueness ratings, they will apparently require modification before they clearly predict learning and retention.

## REFERENCES

- BOWER, G. H., & HOLYOAK, K. Encoding and recognition memory for naturalistic sounds. *Journal of Experimental Psychology*, 1973, **101**, 360-366.

- BUTLER, B. E., & MERIKLE, P. M. Uncertainty and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 634-641.
- DAVIDSON, R. E., SCHWENN, E. A., & ADAMS, J. F. Semantic effects in transfer. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 212-217.
- GALBRAITH, R. C. Encoding variability and the concreteness effect in paired-associate learning. *Memory & Cognition*, 1975, **3**, 282-286.
- GORMAN, A. M. Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, 1961, **61**, 23-29.
- HASHER, L., & JOHNSON, M. K. Interpretive factors in forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 567-575.
- HASHTROUDI, S., & JOHNSON, M. K. Transfer and forgetting: Interpretive shifts and stimulus reinstatement. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 262-272.
- KLEIN, K., & SALZ, E. Specifying the mechanisms in a levels-of-processing approach to memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 671-679.
- LESGOLD, A. M., & GOLDMAN, S. R. Encoding uniqueness and the imagery mnemonic in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 193-202.
- MARTIN, E. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review*, 1968, **75**, 421-441.
- MARTIN, E. Stimulus encoding in learning and transfer. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. New York: Wiley, 1972. Pp. 59-84.
- MERRYMAN, C. T., & MERRYMAN, S. S. Stimulus encoding in the A-B', AX-B' and the A-Br', AX-B paradigms. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 681-685.
- MOSCOVITCH, M., & CRAIK, F. I. M. Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 447-458.
- PAIVIO, A. Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 32-38.
- PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Part 2).
- ROSS, J. The verbal loop hypothesis: A statistical audit. *Australian Journal of Psychology*, 1968, **20**, 233-235.
- STOFF, D. M., & EAGLE, M. N. The relationship among reported strategies, presentation rate, and verbal ability and their effects on free recall learning. *Journal of Experimental Psychology*, 1971, **87**, 423-428.
- WINOGRAD, E., & GEIS, M. F. Semantic encoding and recognition memory: A test of encoding variability theory. *Journal of Experimental Psychology*, 1974, **102**, 1061-1068.

(Received for publication July 18, 1977.)