Imagery and meaningfulness of right and wrong items in verbal discrimination learning*

JAMES R. ULLRICH and BELA A. BALOGH†
University of Montana, Missoula, Mont. 59801

The effects of variations in the rated imagery (I) and the meaningfulness (m) of the right word (R) and the wrong word (W) on a verbal discrimination (VD) task were examined by factorial manipulations of the I and m values of the R and W words within VD pairs. Results indicated that I was a more potent variable than m in the VD task.

The rated image-evoking capacity of words has been shown to be a potent variable in a wide variety of verbal-learning tasks (Paivio, 1969). In particular, Paivio & Rowe (1970) and Rowe & Paivio (1971) have recently demonstrated that a verbal discrimination (VD) task is easier to learn when both the right (R) and the wrong (W) words are high in imagery (I) than when both words are low in I. By extending the concepts of representational responses and implicit associative responses to include imagerial as well as verbal responses, these results were interpreted as favoring a frequency theoretic (Ekstrand, Wallace, & Underwood, 1966) explanation for VD learning.

Since frequency theory assumes that performance is a function of the difference in frequency units attached to the R and W items, the question arises whether speed of learning is influenced by the imagery of the R word or by the imagery of the W word within a specific pair. Paivio and Rowe's extension of frequency theory would predict that words high in I would have a greater tendency than words low in I to evoke responses analogous to the verbal representational and pronunciation responses of the original statement of frequency theory. These imagerial reactions would be included as contributors to the number of frequency units accrued to each word and, hence, would affect the rate of VD acquisition. The present study was designed to investigate this extension by means of a 2 by 2 factorial design in which the imagery of the R words and the imagery of the W words were independently varied. In contrast, Paivio & Rowe (1970) and Rowe & Paivio (1971) maintained either high or low imagery within a VD pair.

*This research was supported by Grant MH 18695-01 from the National Institute of Mental Health, United States Public Health Service, and a grant from the University of Montana College of Arts and Sciences.

†Requests for reprints should be sent to James R. Ullrich, Department of Psychology, University of Montana, Missoula, Mont. 59801.

This study was also designed to determine if the meaningfulness (m) of the R and W words within a VD pair was an important determinant of performance. Runquist & Freeman (1960) found that when the VD pairs consisted of one high and one low m value CVC syllable, the lists in which the correct item was high in m were easier to learn than the lists in which the correct item was low in m. Runquist and Freeman with CVC syllables and Schulz & Hopkins (1968) with dissyllables found that there were no differences in the rates of acquisition of visually presented items when the items were either both high or both low in m. The latter authors found that aurally presented lists of dissyllables that were low in m were more difficult than lists that were high in m. Paivio & Rowe (1970) found no differences in the rate of learning between lists consisting of VD pairs high in m and VD pairs low in m when I value was held constant for both lists. The present interpretation of frequency theory would predict that m has an effect within a VD pair, since the m value would affect the ease with which representational and rehearsal responses could be made to the individual R and W words. A 2 by 2 factorial design, in which the m value of the R and W words was varied when I value was held constant, was used to assess the effects of the meaningfulness dimension on VD acquisition.

METHOD

The present experiment was a VD task in which each of eight groups of 18 Ss were required to differentiate between the R and W items of a 16-pair VD list. In contrast to previous VD learning experiments employing the conventional memory drum for the presentation of stimuli, this experiment used a method of presentation in which the items to be discriminated and the feedback were all printed on individual Hollerith data cards and presented via an IBM 029 keypunch to the S. The R and W terms appeared in the card bed for each presentation, and S pushed a button corresponding to the position of the

correct item of a particular VD pair; following this, the trailing edge of the data card was exposed, and the correct item of the VD pair was shown to S to inform him of whether or not he made the correct choice. [See Ullrich (1972) for a more complete description of the apparatus.] There was one card for each of the 16 pairs, and a blank card separated each of the 25 trials; consequently, each S was required to proceed through a deck of $425 = (16 + 1) \times 25$ data cards during the experiment. Each S proceeded at his own pace; almost all Ss completed the task within a 30-min period. Each S received a practice list of 4 pairs of words intermediate in I and m values before starting the experiment.

Stimulus items were obtained from a list of words previously scaled for I and m (Paivio, Yuille, & Madigan, 1968) by selecting a pool of 32 high-I and a pool of 32 low-I words for which the scaled m values were held approximately constant at an intermediate level. Items were chosen from these two pools of words and randomly paired so as to form four possible combinations of high and low lists. The groups are denoted as HIR-HIW, HIR-LIW, LIR-HIW, and LIR-LIW, where H and L refer to high or low, I refers to imagery, and R and W refer to the right and wrong words. Thus, the LIR-HIW group had R words that were low in I and W words that were high in I. The two groups HIR-HIW and HIR-LIW had identical R words, as did the groups LIR-HIW and LIR-LIW. Identical controls existed for the W words.

In order to investigate questions pertinent to the meaningfulness dimension, pools of 32 high- and 32 low-scaled m value items were selected from the Paivio et al (1968) norms; however, this time the scaled I value was held approximately constant at an intermediate level. As was done with words varying in terms of imagery, these pools of items were employed to form random pairings of four possible combinations in terms of meaningfulness (HmR-HmW, HmR-LmW, LmR-HmW, and

Table 1
Mean I Values of the R and W Words (IR and IW) and Mean m Values for the R and W Words (mR and mW) for Each of the Eight Experimental Groups

Group	IR	mR	IW	mW
HIR-HIW	6.38	5.74	6.39	5.66
HIR-LIW	6.38	5.74	3.01	5.45
LIR-HIW	3.08	5.43	6.39	5.66
LIR-LIW	3.08	5.43	3.01	5.45
HmR-HmW	4.41	6.47	4.26	6.53
HmR-LmW	4.41	6.47	4.20	4.78
LmR-HmW	3.82	4.74	4.26	6.53
LmR-LmW	3.82	4.74	4.20	4.78

Table 2
Means and Standard Deviations of the Total Number of Errors Made on Trials 2-25
for Each of the Experimental Conditions

Group	Mean	SD	Group	Mean	SD
HIR-HIW	14.94	15.75	HmR-HmW	22.00	17.56
HIR-LIW	16.78	16.98	HmR-LmW	14.67	10.16
LIR-HIW	11.28	14.49	LmR-HmW	16.94	11.55
LIR-LIW	27.17	15.46	LmR-LmW	20.56	17.56

LmR-LmW). Thus, the same types of lists were obtained in terms of meaningfulness as was done with lists for imagery. Table 1 contains the mean I and m values for all lists used in this experiment.

Once the four lists dealing with the variable of R and W word imagery and the four lists dealing with the variable of R and W word meaningfulness were constructed, a computer program generated a random order of presentation for the pairs from trial to trial and for the position of the R and W words on the stimulus card. All eight groups used the same randomizations; no attempt was made to counterbalance order within a trial or the order of the R and W items within a pair.

RESULTS AND DISCUSSION

Since the data from both experiments were collected simultaneously, the number of errors made by each S on blocks of three trials (T) was analyzed as a 2 by 2 by 2 by 8 analysis of variance with repeated measures on the last factor. The main effects were denoted as I-m, R, W, and T. The results indicated a significant interaction between R and W, $F(1,136) = 6.09$, $p < .05$, and a significant interaction between R, W, and T, $F(7,952) = 6.28$, $p < .01$. The interaction between I-m and W was significant, $F(1,136) = 4.50$, $p < .05$, as was the interaction of I-m by W by T, $F(7,952) = 6.95$, $p < .01$. The T main effect was significant, $F(7,952) = 154.03$, $p < .01$, but the tests for neither W nor the W by T interaction reached the required levels, $F(1,136) = 1.92$ and $F(7,952) = 1.93$, respectively. The Fs for each of the other tests were less than unity. Table 2 contains the means and standard deviations for the total number of errors made on Trials 2-25 for each of the eight experimental conditions. Examination of the means as a function of blocks of three trials revealed that the form of the interactions with T were such that the groups were different during the initial blocks and converged to a common value as trials progressed.

The significant interaction between R and W can be summarized by noting that when one item of a VD pair is high and the other item is low in either I or m, there are fewer errors than when both items are of the same rated

I or m value. The absence of a triple interaction indicated this to be true for both I and m; however, this interpretation is obscured by the significant I-m by W interaction. The form of this latter interaction is specified by noting that when the I value of the W word increases or when the m value decreases, the number of errors decreases.

An a priori t test contrasting HIR-HIW with LIR-LIW was found to be significant, $t = 2.43$, $df = 136$, $p < .05$, but a test comparing HmR-HmW with LmR-LmW was not significant, $t = 0.28$. The results of these tests are in accord with comparisons for equivalently constructed lists reported by Paivio & Rowe (1970). Their conclusion that I was a more potent stimulus attribute than m is supported by these two tests; however, the possibility does exist that this result may be due to a more complex relationship between I-m and W. The finding of a significant R by W interaction would generally support an extended version of frequency theory were it not for the fact that the LR-HW conditions had a low number of errors. To account for this finding under the tenets of frequency theory, it would be necessary to assume that Ss are applying Ekstrand et al's Rule 2 and that rehearsal under these conditions would be for the W items. Since it is unlikely that these assumptions are realized, it can only be concluded that frequency theory is not supported by the particular form of this interaction. Furthermore, the significant I-m by W interaction does not seem to follow from any reasonable extensions of frequency theory.

An alternative explanation is that the Ss in the VD task are able to utilize whatever cues that may exist in the list and that these cues may be imaginal or frequency-based in nature. If one assumes that imaginal cues are more salient than those that are frequency-based, it is not unreasonable to assume that an item very high in I may be coded either as a R item or as a W item. Thus, it would be predicted that the LIR-LIW list would be more difficult than the HIR-HIW, HIR-LIW, or the LIR-HIW lists, since I could not be effectively utilized as a cue in the former list. When imaginal properties are not sufficiently high, frequency

cues may be utilized to form the basis for the discrimination. Since meaningfulness would affect the number of implicit associative responses and representational responses, it would be expected that items higher in m would produce a larger number of frequency units. Thus, whenever there is a difference in meaningfulness of the R and W words, there would be a difference in the number of frequency units and discrimination would be easier. Hence, it would be predicted that the HmR-LmW and LmR-HmW pairs would be easier than the HmR-HmW and LmR-LmW pairs. The hierarchical nature of cue utilization was suggested by Rowe & Paivio (1971) to account for the fact that Thorndike-Lorge frequency was utilized only when the I value was low. The results of the present investigation are entirely in accord with this interpretation.

There exists a procedural artifact that potentially may have influenced the direction of the results of the present investigation. The basis for this artifact may be seen by noting that for those VD pairs involving high and low intrapair values, for either the imagery or meaningfulness dimension, Ss may have tended to select the R items on the basis of the accrual of frequency units or on the basis of imaginal coding. Thus, for these lists, the VD task may have become one of concept identification and Ss may have derived additional cues from the conceptual differences among intrapair relationships in selecting the correct items. An examination of the number of errors on the second half of the initial guessing trial revealed that the 252 errors were 3.00 standard deviations below the expected value of 288 as calculated by the binomial distribution; this indicated that Ss may have utilized the information on the first half of the guessing trial to aid performance on the second half. On the other hand, this information did

Table 3
Total Number of Errors Made on the First Eight Pairs (First Half) and Second Eight Pairs (Second Half) of the First and Second Trials*

Group	Trial 1		Trial 2	
	1st Half	2nd Half	1st Half	2nd Half
HIR-HIW	69	84	16	32
HIR-LIW	63	58	29	32
LIR-HIW	70	64	14	18
LIR-LIW	68	57	35	44
HmR-HmW	63	75	31	43
HmR-LmW	61	56	23	30
LmR-HmW	76	74	19	42
LmR-LmW	68	75	26	39

*Under the binomial distribution with $p = \frac{1}{2}$, the expected number of errors is 72 with a standard deviation of 6.

not seem to be utilized after the first trial insofar as the Pearson correlation between the number of errors made on the first half of the second trial by each S in the groups in which one item was H and those in which the other was L was -0.03 . Further evidence to support the position that the artifact was not a potent variable in this task may be gained by noting that the improvement in performance from the last half of the first trial to the first half of the second trial was greater for those groups in which the artifact was absent than for the groups with heterogeneous lists in which the artifact may have been operating. Table 3 presents the total number of errors for the first and second halves

of the first two trials for each of the experimental groups. This problem of the unknown effects of differential guessing rates on initial trials when alternatives are heterogeneous has been reported previously by both Runquist & Freeman (1960) and by Schulz & Hopkins (1968).

REFERENCES

- EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal discrimination learning. *Psychological Review*, 1966, 73, 556-578.
- PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, 76, 241-263.
- PAIVIO, A., & ROWE, E. J. Noun imagery, frequency, and meaningfulness in verbal discrimination. *Journal of Experimental Psychology*, 1970, 85, 264-269.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 1968, 76, 1-25.
- ROWE, E. J., & PAIVIO, A. Word frequency and imagery effects in verbal discrimination learning. *Journal of Experimental Psychology*, 1971, 88, 319-326.
- RUNQUIST, W. N., & FREEMAN, M. Roles of association value and syllable familiarization in verbal discrimination learning. *Journal of Experimental Psychology*, 1960, 59, 396-401.
- SCHULZ, R. W., & HOPKINS, R. H. Presentation mode and meaningfulness as variables in several verbal-learning tasks. *Journal of Verbal Learning & Verbal Behavior*, 1968, 7, 1-13.
- ULLRICH, J. R. Re-pairing, number of alternates, and meaningfulness in verbal discrimination learning. *Journal of Experimental Psychology*, 1972, in press.