was essentially equal to that shown by 13 AO-trained rats (tested concurrently, also without prior magazine training) with a 30-sec CS-UCS interval and the 30-sec rats of the present study. This would seem to indicate little or no positive contribution of the autoshaping component in successful instances of AO-trained leverpressing, and from a practical standpoint, little advantage of AO over conventional operant procedures as a training technique, at least with relatively tame rats.

AO training may have more important uses as a criterion task than as a training method in any event. In studies concerned with the effects of prior exposure to stimulus and/or reward presentations on the initial acquisition of operant leverpressing (e.g., experiments on learned indolence, latent inhibition, incentive contrast, and classical-to-instrumental transfer), the use of a 30-sec AO procedure seems worth considering as a criterion task which is readily learnable and provides the additional control, unlike pure-operant procedures, of guaranteeing equal experience with the reinforcing stimulus at each stage of acquisition.

REFERENCES

- Brown, P. L., & Jenkins, H. M. Auto-shaping of the pigeon's key peck. *Journal of the Experimental Analysis of Behavior*, 1968, 11, 1-8.
- Engberg, L. A., Hansen, G., Welker, R. L., & Thomas, D. R. Acquisition of key-pecking via autoshaping as a function of prior experience: "Learned laziness"? *Science*, 1972, 178, 1002-1004.
- Hulse, S. Patterned reinforcement. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 7. New York: Academic Press, 1973.
- Peterson, G. B., Ackil, J. E., Frommer, G. P., & Hearst, E. S. Conditioned approach and contact behavior toward signals for food or brain-stimulation reinforcement. *Science*, 1972, 177, 1009-1011.
- Sidman, M., & Fletcher, F. G. A demonstration of autoshaping with monkeys. *Journal of the Experimental Analysis of Behavior*, 1969, 11, 307-309.
- Smith, G. S., Borgen, L. A., Davis, W. M., & Page, H. B. Automatic magazine and barpress training in the rat. *Journal of the Experimental Analysis of Behavior*, 1971, 15, 197-198.

(Received for publication February 25, 1974.)

Bulletin of the Psychonomic Society
1974, Vol. 3 (5B), 385-387

Does imagery enhance long-term retention*

LEO POSTMAN

University of California, Berkeley, California 94720

In a recent series of experiments, Begg and Robertson (1973) found a direct relationship between the concreteness-imagery value of words and long-term retention. It is suggested that the procedures used to equate degree of learning for concrete and abstract words were faulty, and the conclusion that imagery facilitates retention is questioned.

Is long-term retention, as well as speed of acquisition, related directly to the concreteness-imagery (C-I) value of words? The available empirical evidence appears to provide contradictory answers to this question. In a

recent study, Begg and Robertson (1973) reported a series of experiments in each of which long-term retention was found to be substantially higher for concrete (C) than for abstract (A) nouns. This difference was observed under conditions of both free-recall and paired-associate learning. The latter result contrasts

*The writer's research is supported by Grant MH-12006 from the National Institute of Mental Health.

sharply with that obtained previously by Postman and Burns (1973), who measured the retention of paired-associate lists as a function of the C-I value of the stimulus and of the response terms. With degree of learning equated by adjustment in the number of learning trials, there was a drastic change in the rank order of conditions between acquisition and long-term recall. In particular, recall was slightly lower for C-C than for A-A pairs. It is for these two types of pairs that large differences in retention were found by Begg and Robertson.

The purpose of the present note is to call attention to some serious methodological deficiencies in the experiments of Begg and Robertson. The points at issue concern the methods used to equate degree of learning prior to the assessment of long-term retention. This problem arises whenever recall is compared for materials acquired at different rates. The various procedures employed by Begg and Robertson illustrate within the confines of a single study the major sources of bias that are likely to vitiate the measurement of retention under such conditions.

Most of the methodological problems to be considered here were examined fully in two important papers by Underwood (1954, 1964). The present discussion will draw heavily on his analyses. The basic question is how the terminal degrees of learning are to be equated when there are differences in speed of acquisition. To carry fast and slow learners (or practice on easy and difficult materials) to the same criterion does not solve the problem because (a) the criterial trial produces a larger increment in degree of learning for fast than for slow Ss, and (b) the magnitude of the criterion fall is inversely related to speed of learning. Equally inadequate is the so-called "drop-out" procedure, under which items are removed from the list after being learned to some low criterion, e.g., one correct anticipation. This procedure is based on the demonstrably invalid assumption that the occurrence of a correct response reflects the same level of learning for fast and for slow Ss. It is a fact that the probability of the correct response being repeated on the next trial is considerably greater for fast than for slow learners. It is that probability, of course, that defines the expected level of retention. Thus, the use of the drop-out method virtually ensures inequality in the degree of learning and a corresponding bias in measures of retention.

The solution developed by Underwood is to adjust the number of training trials given to fast and slow Ss (or on easy and difficult materials) so as to equate the level of learning. The appropriate numbers of trials can be determined in a pilot experiment. In the retention study proper, expected recall scores can be calculated by means of a probability analysis, i.e., by projecting Ss' learning curves to the trial beyond the last one given during acquisition. The observed recall scores can then be expressed as losses from the projected levels. The correlations between predicted and obtained scores are typically quite high. A potentially important limitation of

this procedure is that it does not permit an assessment of the increments in strength produced by overlearning. Once an item reaches asymptotic strength, the predicted probability of recall levels off at, or near, unity, regardless of prior speed of acquisition. However, successive overlearning trials may be expected to result in greater gains for fast than for slow Ss. When the number of overlearning trials is equal, the expected retention scores of fast learners are, therefore, systematically underestimated relative to those of the slow ones. This bias may be counteracted or reversed when the number of overlearning trials is greater for slow than for fast Ss. There is no way to determine precisely the direction and magnitude of the error entailed by the presence of items at asymptotic strength. However, as will be shown presently, internal analyses can be carried out to assess the net effect on retention.

We return now to the experiment on paired-associate retention in the study of Begg and Robertson. These investigators objected to the method of adjusted trials used by Postman and Burns because it makes for a wide disparity in the number of correct anticipations during acquisition.¹ (Stated in these general terms, the objection obviously misses the logic of the adjustment, but it will be considered below in relation to the problem of asymptotic probabilities.) Begg and Robertson decided to resort to the drop-out procedure for the purpose of equating degree of learning, i.e., pairs were dropped from the list following one correct anticipation. Two list lengths and three criteria of learning (50%, 75%, and 100%) were used. On a 72-h test of retention, more C-C than A-A pairs were recalled by the 100% and the 75% criterion groups, but not by the 50% criterion groups. No conclusions about differences in retention can be drawn from these results because of the bias inherent in the drop-out procedure. In the present case, it is possible to provide an estimate of the magnitude of this bias. For the pilot Ss in the study of Postman and Burns, the probability of a correct response on the trial following the first correct anticipation was .81 for C-C pairs, and .52 for A-A pairs. A discrepancy in degree of learning of this order of magnitude would appear to be sufficient to account for the observed differences in retention.

Other features of the results reinforce the conclusion that the retention differences reflect a failure to control degree of learning. Recall of C-C and A-A pairs was equal for the 50% criterion groups, but an increasing difference emerged as the criterion level was raised. This trend points to the operation of a substantial item-selection bias. The higher the criterion level, the larger is the proportion of relatively difficult items that have to be learned by Ss. Thus, the discrepancy in degree of learning would be maximized at the highest criterion, and it is precisely under that condition that the largest difference in retention was found. The complete absence of a difference for the 50% criterion groups (Begg & Robertson, 1973, Fig. 2, p. 695) may, of course, be considered as *prima facie* evidence against the alleged

Table 1
Mean Measures of Acquisition and Retention for Fast and Slow Learners (Data from Postman & Burns, 1973)

	Antic/Item in Learning	Projected Score	Number Recalled	Percent Loss
Fast Learners				
C-C	2.37	13.52	6.63	50.78
A-A	7.71	14.27	7.50	47.89
Slow Learners				
C-C	1.45	11.41	3.13	72.91
A-A	3.86	10.69	3.50	67.08

relationship between C-I value and retention.

The next question to be considered is whether the method of adjusted trials used by Postman and Burns entailed a bias favoring higher retention of A-A than of C-C pairs.² The possible source of such a bias would be the greater overlearning of relatively easy pairs in the A-A than in the C-C list. Since there was a fixed number of trials for each list (4 for C-C and 13 for A-A), differential amounts of overlearning should have a greater effect on the retention scores of fast than of slow Ss. To determine whether such was the case, each group was evenly divided into fast and slow Ss on the basis of the numbers of correct responses in acquisition. As Table 1 shows, the discrepancy in the number of correct anticipations was indeed greater for fast than for slow Ss, and this fact is reflected in the projections. The losses were greater for the slow than the fast subgroups ($p < .01$), but the critical I-C Value by Ability interaction is clearly negligible ($F < 1$). The conclusion is that the comparison of retention losses was not distorted by biased projections for asymptotic items.

In this final section, the two remaining experiments of Begg and Robertson are discussed briefly. In one of these, the effect of C-I value on the long-term retention of free-recall lists was investigated. To control the degree of learning, the incremental method of learning originally introduced by Mandler and Dean (1969) was used. Under this arrangement, the first trial consists of the presentation of a single word, and the length of the list to be recalled is increased by one word on each successive trial. The authors state that under this procedure "there is no reason to expect differential acquisition as a function of concreteness [p. 690]." This statement is puzzling in view of their emphasis on the role of organizational processes in the mastery of this task. Furthermore, if concreteness (imagery?) does not influence acquisition, it is not clear why an effect on retention was predicted. Each S learned one concrete and one abstract list (with order balanced) in two sessions separated by 24 h. Retention for both lists was tested 48 h after the second session. Performance was virtually perfect for both lists on all learning trials, but retention was significantly higher for C than for A words. Again, no conclusions about retention differences are warranted because degree of learning cannot be

assessed when performance is nearly perfect throughout acquisition. It is quite likely that after a constant number of practice trials the C list was learned better than the A list. A similar error has been frequently made in analyses of short-term memory (cf. Underwood, 1964, pp. 126-128). Furthermore, the design confounds order of acquisition and retention interval; the complex retroactive and proactive effects that are bound to develop under these conditions cannot be evaluated. Thus, these measures of retention are in principle uninterpretable.

In the third experiment of Begg and Robertson, a mixed list of C-C and A-A pairs was presented for five trials. Recall was either cued or noncued. Under each condition, Ss were given both an immediate test and a delayed test (by mail) 6-10 days later. After a constant number of trials on a mixed list, immediate recall was, of course, higher for C than for A items. The relative difference was enhanced, particularly under the cued condition, on the delayed test. The basic error in this design is obvious, namely, the use of repeated tests of retention for materials learned to quite different degrees. The successive tests are not independent of each other. The long-term gains derived from the first test must be expected to vary directly with the initial level of performance. The correlation between initial scores and test-trial gains has been demonstrated previously (Postman, 1963, Figs. 8-2 and 8-3).

Begg and Robertson recognize that their procedures may be open to criticism. They go on to claim however, that the overall conclusion based on the very different techniques used by them—that imagery facilitates retention—is "unshakeable" (p. 699). Unfortunately, three consistently biased operations do not converge on an unbiased inference. The conclusion is far from unshakeable.

REFERENCES

- Begg, I., & Robertson, R. Imagery and long-term retention. *Journal of Verbal Learning & Verbal Behavior*, 1973, 12, 689-700.
- Mandler, G., & Dean, P. J. Seriation: Development of serial order in free recall. *Journal of Experimental Psychology*, 1969, 81, 207-215.
- Postman, L. One-trial learning. In C. N. Cofer and B. A. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963.
- Postman, L., & Burns, S. Experimental analysis of coding processes. *Memory & Cognition*, 1973, 1, 503-507.
- Underwood, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychological Bulletin*, 1954, 51, 276-282.
- Underwood, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning & Verbal Behavior*, 1964, 3, 112-129.

NOTES

1. Begg and Robertson state incorrectly that in Postman and Burns's experiment the lists were learned to criterion (p. 694). Fixed number of trials, as determined by the results of a pilot experiment, were given to the different groups.

2. Since the question of interest concerns the discrepancy between our results and those of Begg and Robertson, only the results for Lists C-C and A-A are presented.

(Received for publication March 8, 1974.)