JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOUR 7, 864–868 (1968)

The Effects of Stimulus Uncertainty and S-R Transitional Probability on Paired-Associate Learning[1]

IRVING BIEDERMAN, JOSEPH S. DUMAS, AND ROY LACHMAN

State University of New York at Buffalo 14214

The Associative Probability Theory asserts that the greater the number of associates elicited by a stimulus, the greater the probability that one of these will be an appropriate mediator for new learning. The present study tested the adequacy of this theory in describing the effect of natural-language-sequential habits on PA learning. The transitional probability between pairs of words and the uncertainty of the probability distribution of words following a given word were computed from 985 student essays. On the basis of these norms, 3 PA lists were constructed with the stimulus terms either high or low in uncertainty and with either high or low transitional probabilities between stimulus and response terms. In all three lists fewer errors were made in learning pairs of high transitional probability but, contrary to expectations from the extension of Associative Probability Theory, stimuli of high uncertainty resulted in more errors.

It is widely accepted that associations to a presented stimulus can serve to mediate and thus facilitate S-R learning. Underwood and Schulz (1960) describe an "Associative Probability Theory" (APT) to explain how such learning varies as a function of the number of associations.

The idea of associative probability is simple: the greater the number of associates elicited by a stimulus, the greater is the probability that one of these will link up with another item. Thus an already established association, albeit weak, perhaps, can be used as the basic association for new learning (Underwood and Schulz, 1960, p. 295–296). Several authors (e.g., Mandler and Huttenlocher, 1956; Noble, 1963) have accepted this conception.

Criticisms of APT derive from three considerations. First, some associates might not be employable for mediation of a given response. From an information-processing viewpoint (Field and Lachman, 1966; Treisman, 1965; Yntema and Trask, 1963), an increase in the proportion of inappropriate associates might increase the duration of implicit scanning processes by which associations are selected for mediation. Second, is the problem of the "interference paradox" (Underwood and Schulz, 1960). The existence of alternative associations to the same stimulus, even if they were all appropriate, could result in response conflict if these associations were incompatible with each other. Both of these criticisms imply that increases in the number of associations to a stimulus might result in increased interference, as well as facilitation, effects. If learning rates are to be predicted from the number of associations, then it is first necessary to specify the relative amounts of interference and facilitation resulting from these associations. For a given stimulus with its associative distribution such a specification would be quite specific to the to-be-learned response. In asserting that a positive correlation exists between number of associations and learning rate, APT assumes either that

[1] The collection of the language norms and construction of the computer algorithm was done by R. Lachman, partially supported by NSF Grant GB-6530 and MH-14153. The informational statistics and data analysis were supported by the Computing Center at the State University of New York at Buffalo. The suggestions and criticisms of G. A. Bruder, B. R. Bugelski, C. Perrino, and E. M. Segal are gratefully acknowledged.

interference does not occur among associates, or else that the interference effects are less than facilitation effects. Moreover, APT assumes that interference effects increase less rapidly than facilitation effects as the number of associates is increased. Third, a difficulty of APT is that in considering only the *number* of associations it ignores the fact that associates differ in strength. It is not clear from APT how to compare a stimulus with a few very strong associates with a stimulus with many relatively weak associates.

The information measure of uncertainty (UN) provides a means of quantifying both the number and differences in strength of distributions of associates (Berlyne, 1959; Garner, 1962; Glanzer, 1962). The measure is obtained by calculating

$$\mathrm{UN} = -\sum_{i=1}^{N} p(i)\log_2 p(i) \qquad (1)$$

where $p(i)$ refers to the probability of associating response i to a particular stimulus and N refers to the number of associates to that stimulus. The greater the number of associates and the more equiprobable their occurrence, the higher the uncertainty. An information-scanning or response-conflict model would predict that increases in UN should be interfering while APT predicts that increases in UN should be facilitating. Glanzer (1962) computed UN from distributions of associates to words and showed that words with higher levels of UN were more difficult to learn, as both stimuli and responses, in paired-associate lists. If the words which follow a specified word in the language are regarded as potential mediators or sources of interference, then the predictions as to the effect of UN on learning from APT and considerations of informational scanning (or response conflict) may be extended to the domain of natural language. Such an extension of APT has been made by Underwood and Ekstrand (1967).

The present experiment seeks to explore this relationship between UN and learning, but with UN computed from samples of natural language (student essays) rather than from association norms. For each word in the sample, the transitional probability (TP) for each specific word which followed it was computed. For example, the word PEOPLE occurred 644 times and was followed by WHO 45 times. TP for PEOPLE-WHO was thus 45/644 or .073. UN was computed for the distribution of words following a specified word by substituting TP for $p(i)$ in Eq. 1. Therefore, the greater the number of different words that follow a word and the more equivalent their TP's, the higher the value of UN.

In addition to UN the present experiment was designed to investigate the direct influence of TP on paired-associate learning. If the transitional probabilities, as measured by these norms, can be taken as measures of the pre-experimental language-sequence habits, then the prediction of their effect in PA learning is obvious: the higher the transitional probability between the stimulus and response, the more rapid should be the acquistion of that pair.

METHOD

Subjects. Lists 1, 2, and 3 were administered to groups of 28, 45, and 15 students, respectively, from the introductory and experimental psychology classes at the State University of New York at Buffalo.

Verbal Materials. The language samples, essays under 200 words in length, were written by 985 introductory psychology students at Buffalo in 1965; the samples yielded 114, 169 words. Free choice was given as to topic but the vocabulary was limited to a list of the 500 most frequent (AA) words from the Thorndike-Lorge Count (1944). This list, numbered for subsequent coding, was provided to each essayist. Modifications in number, gender and tense were permitted, yielding a total vocabulary of 767 words. Except for periods, all punctuation marks were ignored.

The essayists number-coded all the words in their essays and the numbers were card-punched for computer analysis. For each word with a frequency greater than ten, the TP's of the distribution of words (and periods) following it was computed and the value of UN for this distribution calculated as described above. These uncertainties ranged from 0.00 bits for

TABLE 1

LISTS, UNCERTAINTY VALUES, AND TRANSITIONAL PROBABILITY VALUES

Condition	List 1	UN	TP	List 2	UN	TP	List 3	UN	TP
	Big-ship	5.82	.015	Both-loved	5.55	.045	Were-three	6.07	.023
	Only-human	6.10	.026	Never-seen	6.10	.058	Good-friends	5.96	.028
HUN-HTP	Great-need	5.84	.034	First-thing	5.55	.022	All-over	6.02	.023
	Often-told	5.70	.026	Just-like	6.15	.037	One-morning	5.64	.031
	Your-child	5.56	.033						
	Mean	5.80	.027		5.85	.040		5.92	.026
	Must-act	5.79	.000	Being-long	5.49	.000	Other-must	5.76	.000
	Always-best	6.18	.000	Child-most	5.43	.000	Boy-always	5.70	.000
HUN-LTP	Small-play	5.62	.000	System-came	5.66	.000	People-lost	5.88	.000
	New-garden	5.79	.000	Then-fear	5.55	.000	This-walk	6.45	.000
	Still-make	5.72	.000						
	Mean	5.82	.000		5.53	.000		5.95	.000
	Into-our	3.10	.030	Such-places	3.19	.035	Across-town	1.72	.054
	Lived-near	3.50	.018	Far-out	3.35	.054	Used-street	2.80	.023
LUN-HTP	Part-time	2.09	.034	Kind-hearts	3.50	.022	Full-person	2.33	.018
	Try-hard	2.23	.022	Told-them	3.60	.045	During-our	2.30	.017
	Liked-each	3.52	.023						
	Mean	2.88	.025		3.41	.039		2.29	.028
	Again-built	2.92	.000	Tried-food	1.59	.000	Begin-anything	2.76	.000
	Care-under	3.30	.000	Door-also	3.35	.000	Front-child	3.02	.000
LUN-LTP	Began-tree	1.55	.000	Rest-again	2.83	.000	Chance-without	2.70	.000
	Upon-land	1.11	.000	Once-found	3.06	.000	Tried-each	1.58	.000
	Wanted-money	2.28	.000						
	Mean	2.23	.000		2.69	.000		2.50	.000

KINDS (followed by OF all 17 times it occurred) to 7.37 bits for AND (followed by 453 different words on its 2,606 occurrences).

From these norms, three lists of paired associates were generated as shown in Table 1. There were 20 pairs in List 1 and 16 pairs in Lists 2 and 3. Half of the pairs in each list were high TP (HTP) in that the response was a word that had followed the stimulus in the norms—the HTP's ranged from .015 to .058. In the other half the response word had never followed the stimulus but (with the exception of BEGAN-TREE in List 1) the pair was semantically and grammatically possible. The low TP (LTP) pairs, therefore, all had values of 0.00. Half of the pairs in both TP levels in each list had stimuli with relatively high uncertainties (HUN) (5.43–6.45 bits). The other half had stimuli with relatively low uncertainties (LUN) (1.11–3.83 bits). The responses in the pairs from all four treatment combinations were equated for frequency and uncertainty. The UN and TP values for all pairs are shown in Table 1. To control for serial-position effects, pairs of all four types were represented in each fifth of List 1 and each quarter of Lists 2 and 3.

Procedure. The block anticipation-method was used with the lists recorded on tape. The pairs were presented at a 2-sec rate followed by testing with the stimuli alone, in a different order, at a 3-sec rate. Responses were written in answer booklets. To reduce the chance of responding in an inappropriate space, the answer spaces were numbered according to the stimulus order during the test. These numbers immediately preceded each stimulus on the tape. List 1 was presented for 10 trials and Lists 2 and 3 were presented for 8 trials. A different sequence was used on each trial. Each trial was recorded on a different page in the test booklet. The intertrial interval before both presentation and test was 5-sec. Following the last test trial, answer sheets were distributed and *S*s passed their booklets to someone else for scoring for errors (omissions and intrusions).

Results

Table 2 shows the mean number of correct responses for each treatment combination. The effect of UN was significant in all three lists in that the LUN pairs were associated with fewer errors: $F(1, 27) = 15.13$, $p < .001$, for List 1; $F(1, 44) = 14.54$, $p < .001$, for List 2; and $F(1, 14) = 11.59$, $p < .01$ for List 3. The error mean squares (UN × *S*s w/gps.) were 3.91, 6.41, and 2.97 for Lists 1, 2, and 3 respectively. The effect of TP was also significant in all three lists, with the HTP conditions being associated with fewer errors: $F(1, 27) = 32.17$, $p < .001$, for List 1; $F(1, 44) = 9.18$, $p < .001$ in List 2; and $F(1, 14) = 4.95$, $p < .05$ in List 3. The error mean squares (TP × *S*s w/gps.) were 4.35, 6.74, and 3.56 for Lists 1, 2, and 3 respectively. The UN × TP interaction was significant at the .001 level in Lists 2 and 3 in that the effects of TP were greater at LUN than at HUN. However, this interaction was in the opposite direction (not significant) in List 1 and therefore must be cautiously interpreted. Analyses showed that neither stimulus nor response frequency (in our norms) could consistently account for these data.

TABLE 2

Mean Number of Correct Responses Per Item as a Function of Lists and Conditions

Condition	List 1	List 2	List 3	Mean
HUN-HTP	6.67	5.85	5.12	5.88
HUN-LTP	5.46	5.82	5.28	5.52
LUN-HTP	7.11	7.13	6.58	6.94
LUN-LTP	6.32	5.98	5.33	5.88

The overall intrusion rate was virtually identical across the four conditions as shown in Table 3. In all but condition HUN-LTP, intralist intrusions occurred twice as frequently as extralist intrusions. In condition HUN-LTP, the intralist intrusion rate was only slightly, and nonsignificantly, greater than the extralist intrusion rate.

TABLE 3

Proportion of Total Errors That Were Intrusions in the Three Lists Combined

Condition	Type of Intrusion: Intralist	Type of Intrusion: Extralist	Total	*t* diff. (*df* = 12)
HUN-HTP	.186	.092	.278	2.77*
HUN-LTP	.152	.141	.293	<1.00
LUN-HTP	.186	.091	.277	3.58**
LUN-LTP	.199	.093	.292	3.49**

* $p < .05$.
** $p < .01$

Discussion

Two implications of APT in accounting for natural-language sequence learning are under scrutiny. The first is that such learning is dependent upon *associations* to the stimuli. The second is that the greater the uncertainty of the distribution of words following a stimulus, the easier it should be to find a word employable for mediation of a given response.

With respect to the first, Rosenberg (1966) has raised the issue as to the role played by associative factors (as measured by free-association norms) in the relationship between sequential language statistics and learning. Rosenberg maintains that what is responsible for the facilitation of recall as the order of approximation to English is increased (Miller and Selfridge, 1950) is the compatibility of these higher orders with *S*'s guessing habits. These guessing habits are presumably governed by associative factors and it is thus these factors, rather than the statistical constraint per se, which determine the effect of approximation to English. In the present experiment Rosenberg's (1966) position would imply that the responses in the HTP pairs are higher associates to their stimuli than responses in the LTP pairs. However, preliminary association norms show that the responses in both the HTP as well as the LTP pairs are extremely low associates to their stimuli. Moreover, Rosenberg's (1966) interpretation has been called into question by Lachman, Dumas, and Guzy (1966) who demonstrated that at least part of the effect of approximation to English is independent of that resulting from associative factors but is rather directly attributable to the statistical constraint. Consequently, the view that associative factors account for the effects of UN and TP in the present study must be rejected.

Processes operating along dimensions of statistical constraint are generally modeled in a form proposed by Field and Lachman (1966), and Garner (1966), where central importance is given to the internal scanning of sets of alternatives. Increasing the UN value of a set of alternatives would increase scanning difficulty (Hyman, 1953). The effect of UN on learning rates found in the present experiment is consistent with this interpretation and thus contrary to what would be expected from the second implication of APT, that high UN facilitates learning. That this scanning was efficient; i.e., the more probable alternatives were processed first, is evidenced by the facilitative effect found for HTP. Theoretically, the manner in which UN and TP interact is dependent upon specification of the statistical properties of the scanning distributions. Experimentally, however, such an interaction was found to be quite dependent on the particular selection of pairs as evidenced by the inconsistency between List 1 and Lists 2 and 3 in the form of this interaction. Any discussion of the determinants of this interaction must, therefore, be postponed.

That the HUN-LTP condition had, relatively, the highest rates of extralist intrusions serves as further documentation of the relevance of the distribution of TPs. The difficulty of scanning these distributions would be related to UN and TP. Given that the words comprising these distributions did not often appear as items on the lists, HUN-LTP items would be most susceptible to extralist intrusions.

References

Berlyne, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

Field, W. H., and Lachman, R. Information transmission (I) in recognition and recall as a function of alternatives (k). *J. exp. Psychol.*, 1966, **72**, 785–791.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Garner, W. R. To perceive is to know. *Amer. Psychologists*, 1966, **21**, 11–19.

Glanzer, M. Grammatical category: A rote learning and word association analysis. *J. verb. Learn. verb. Behav.*, 1962, **1**, 31–41.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Lachman, R., Dumas, J. S., and Guzy, L. T. Joint effects of word association and approximation to English. *Psychon. Sci.*, 1966, **5**, 145–146.

Mandler, G. Verbal learning. In *New directions in psychology III.* New York: Holt, Rinehart and Winston, 1967. Pp. 1–50.

Mandler, G., and Huttenlocher, J. The relationship between associative frequency, associative ability and paired-associate learning. *Amer. J. Psychol.*, 1956, **69**, 424–428.

Miller, G. A., and Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.

Noble, C. E. Meaningfulness and familiarity. In Cofer, C. N., and Musgrave, B. S. (Eds.), *Verbal behavior and learning.* New York: McGraw-Hill, 1963.

Rosenberg, S. Associative factors in the recall of connected discourse. *Psychon. Sci.*, 1966, **4**, 53–54.

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

Triesman, A. M. Effect of verbal context on latency of word selection. *Nature*, 1965, **206**, 218–219.

Underwood, B. J., and Ekstrand, B. R. Word frequency and accumulative proactive inhibition. *J. exp. Psychol.*, 1967, **74**, 193–198.

Underwood, B. J., and Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

Yntema, D. B., and Trask, F. P. Recall as a search process. *J. verb. Learn. verb. Behav.*, 1963, **2**, 65–74.

(Received: October 2, 1967)