

Anagram solution as a function of transition probabilities and solution word frequency

M. W. WARREN and W. J. THOMSON, Vanderbilt University, Nashville, Tenn. 37203

In a 2 by 2 by 2 repeated-measures factorial design, low anagram-transition probability and high solution-word Thorndike-Lorge frequency were found to facilitate solution of anagrams. The results are interpreted as compatible with a model of anagram solution composed of two processes: word-association emission, and initial bigram rearrangement coupled with word emission. It is additionally proposed that the sampling of bigrams for rearrangement is based on uniqueness.

Mayzner, Tresselt, & Helbock's (1964) application of an S-R mediational model to anagram solution postulates that Ss solve anagrams by successive rearrangements of letters into new combinations. The letter units involved in such rearrangements may be any size greater than two, and sampling of these units is based on their frequency of occurrence in English text.

One prediction of this model is a strong effect of solution-word transition probability, TP (Mayzner & Tresselt, 1962), which is the sum of the frequencies of the four bigrams in a solution word (or anagram). That is, solution words containing high-frequency bigrams (e.g., FRESH) should be solved more quickly than solution words with low-frequency bigrams (e.g., JUDGE), since high-frequency bigrams are more likely to be sampled for rearrangement. Data concerning this prediction have proved contradictory. For example, results supporting the S-R mediational model have been obtained by Mayzner & Tresselt (1962), while Dominowski & Duncan (1964), in two out of three experiments, failed to replicate Mayzner and Tresselt's results.

The present study was designed to further investigate the effects of solution-word TP. In addition, two other variables, solution-word frequency and anagram TP, were investigated. The significance of each of these latter two variables has been noted by Mayzner et al (1964), but the S-R mediational model would assume any effects of these variables to be strongly dependent on variations in solution-word TP.

SUBJECTS

The Ss were 36 Vanderbilt undergraduates (10 males, 26 females), who participated in partial fulfillment of

introductory psychology course requirements.

MATERIALS

The stimulus materials consisted of 32 five-letter anagrams, typed in capital letters on 3 x 4 in. cards. Sixteen solution words were selected from two frequency levels of the Thorndike-Lorge word list. Eight words were chosen from the set occurring 100 times or over per million (high T-L frequency), and eight were chosen from the set occurring less than once per million but more than once per four-million (low T-L frequency). Four of the eight high T-L frequency solution words had high TP; the remaining four had low TP. The eight low T-L frequency solution words were chosen in similar fashion. Mayzner & Tresselt's (1965) bigram frequency counts were used to calculate all TPs.

Two anagrams, a high-TP anagram and a low-TP anagram, were constructed for each solution word. All anagrams were presented in two-move letter orders (e.g., FERHS may be solved for FRESH by moving two letters, R in front of E and S in front of H). Thus, there were two lists of 16 anagrams, the high-TP anagram list and low-TP anagram list, and each list varied along the dimensions of T-L frequency and solution-word TP.

The solution words, anagrams, and median solution times of each are presented in Table 1. Bigram frequency totals for the high-TP solution words varied from 1,800 to 2,741, with a mean of 2,271, while the bigram frequency totals for the low-TP solution words varied from 412 to 901, with a mean of 746. Bigram frequency totals for the high-TP anagrams varied from 1,053 to 2,298, with a mean of 1,524, while the

bigram frequency totals for the low-TP anagrams varied from 40 to 521, with a mean of 277.

PROCEDURE

Males and females were evenly distributed in two independent groups of 18 Ss each in a 2 by 2 by 2 repeated-measures factorial design. The independent factor was anagram TP (high and low), and the repeated measures were T-L frequency (high and low) and solution-word TP (high and low). Each S was tested individually on a different random order of either the high- or low-TP anagram list. S read instructions and practiced on two examples preceding the 16-anagram list. Anagrams were presented for a maximum of 3 min (with correction), and S's solutions were given verbally.

RESULTS

Each S generated 16 latency-of-solution scores, distributed equally in the four cells. For each cell, the median of the four scores was calculated and submitted to a repeated-measures analysis of variance. This analysis revealed: (1) high T-L solution-word anagrams were solved significantly faster than low T-L solution-word anagrams [$F(1,34) = 98.19, p < .001$]; (2) low-TP anagrams were solved significantly faster than high-TP anagrams [$F(1,34) = 4.60, p < .05$]; and (3) an interaction existed between solution-word TP and anagram TP [$F(1,34) = 5.18, p < .05$]. Solution-word TP approached the .05 level of significance [$F(1,34) = 3.93, p < .07$], with low-TP solution words solved more quickly than high-TP solution words. It should be noted that the interaction between solution-word TP and anagram TP, and the near significance of solution-word TP, may have been caused by two rather idiosyncratic anagrams, namely, GABEN, which is very similar to the German verb geben (to give), and VURIS, which is nearly identical in letter arrangement and pronunciation to VIRUS.

Table 1
Median Solution Times (in Sec) for Solution Words and Anagrams

Thorndike-Lorge Frequency	Solution Word TP	Solution Words		Anagram TP			
		High	Low	High		Low	
				High	Low	High	Low
High	High	FRESH	30.6	FERHS	32.8	EHFRS	30.6
		BEGAN	31.5	GABEN	122.5	AEBGN	10.7
		GRANT	34.2	GARTN	62.5	TGRNA	11.6
		CHAIR	31.4	HICAR	22.4	CIHRA	34.8
	Low	VALUE	82.2	UVEAL	93.4	AELVU	19.6
		FIGHT	9.7	GHITF	14.0	IHTFG	6.4
		JUDGE	3.8	UJGD	4.9	UEJDG	3.1
		FRUIT	12.5	FUTRI	21.6	IUFRT	8.8
Low	High	TANGO	180.0	TONAG	180.0	TAOGN	180.0
		GHOUL	180.0	GLOUH	180.0	UOGLH	102.8
		HIKER	82.7	KERIH	98.3	EIHKR	58.3
		GROIN	46.3	GORNI	98.4	IOGRN	15.9
	Low	VIRUS	66.8	VURIS	22.9	IVSRU	165.6
		TRIAD	180.0	TARDI	180.0	RTDIA	133.3
		LOCUS	99.7	SLOUC	122.2	UOLCS	65.1
		GRAPH	75.8	GHARP	180.0	APRHG	29.6

Using the frequency tables for five-letter words compiled by Mayzner & Tresselt (1965), the frequency of each possible bigram in a solution word was tabulated, and a ratio of the frequency of the initial bigram to the frequency of all bigrams was calculated. For example, in the word FRESH, the initial bigram FR occurs 16 times. All 20 possible bigrams in FRESH occur 1,720 times, and thus, the ratio of initial bigram to all bigrams is .009. Similar initial bigram (IB) scores were calculated for each solution word. A Spearman rank correlation was calculated between the ranks of IB scores and the ranks of median solution-word latencies. Rho was found to be $-.42$ ($p \approx .05$).

A further analysis considered bigrams which are likely to begin a word, independent of the absolute frequency of the bigram. For example, the most frequent possible bigram in FRESH is ER, which occurs 416 times. However, this bigram is found at the start of a word only once. In contrast, FR occurs only 16 times, but each occurrence is in the initial position of a five-letter word, and never in any other position. Thus, for each possible bigram, the probability, P, of occurrence of a bigram at the start of a word, conditional on absolute occurrence, was calculated. For ER, this value was .002; for FR, the value was 1.00. Next, the ratio of P for the initial bigram to the sum of all Ps was calculated for each solution word. These conditional initial bigram ratios (CIB) were then ranked and compared to the rank of median solution-word latencies. Rho was calculated to be $.66$ ($p < .01$).

DISCUSSION

Mayzner et al's (1964) S-R mediational model assumes that high-frequency letter units are sampled and manipulated. Two aspects of the present data argue against such interpretation. Such a model predicts that anagrams with high-TP solution words will be solved more rapidly than anagrams with low-TP solution words. Data of the present experiment tended in the opposite direction.

Mayzner & Tresselt (1966) have also found that the initial portion of the solution word is the most critical in determining solution latencies. Therefore, a slight extension of the S-R mediational model would predict the bigram frequency for the initial bigram to be correlated positively with speed of solution. However, the ranks of IB scores were found to correlate negatively with the ranks of median solution time, implying that S may be sampling unique, rather than frequent, letter units.

An alternative letter-unit sampling assumption proposes that S selects letter units which are unique to the initial position of a word. That is, the absolute frequency of a letter unit is unimportant; only the proportion of occurrences in the initial position to other positions is critical. Ranks of the CIB scores based on this assumption were found to correlate positively with the ranks of median solution time. Unfortunately, this latter sampling assumption was conceived post hoc, and hence, experimental distinctions between these two methods of sampling were not considered in designing the experiment. In two instances of the widest divergence in predictions between the two sampling schemes, ranks of the CIB scores appeared to be better predictors of median solution latency.

Assuming that S performs some type of letter rearrangement based on novelty, the effect of anagram TP may be due to a postponement of the onset of this letter-sampling/rearrangement process. Mayzner & Tresselt (1959) suggested that high-TP anagrams are more difficult to "break up" into novel letter rearrangements. It is also possible that high-TP anagrams, since they are more word-like, are more likely to elicit a string of associations than are low-TP anagrams. Some of these associations may be correct solution words, but it is assumed that letter rearrangement in conjunction with word elicitation is a more efficient solution strategy. Hence, any variable which postpones this latter process will increase solution latency.

The effect of T-L frequency may also be related to the proposed letter-rearrangement-plus-word-elicitation process. That is, S may emit associations to anagrams from a response hierarchy that is frequency-ordered. Then, as S rearranges initial letters, he continues to emit associations from his frequency hierarchy. Such a process is continued until the solution word is found. Unfortunately, the data of the present experiment did not contain enough error data to sufficiently check this proposal. However, it was noted that several erroneous words and near-words (e.g., VISUR, JUGED) were emitted. Mayzner et al (1964) also found the intrusion of incorrect words.

In conclusion, the present results are compatible with a model which postulates two processes: first, word-association emission; second, initial letter rearrangement interspersed with word emission. It is further suggested that letter units are sampled for rearrangement on the basis of uniqueness, rather than high frequency.

REFERENCES

- DOMINOWSKI, R. L., & DUNCAN, C. P. Anagram solving as a function of bigram frequency. *Journal of Verbal Learning & Verbal Behavior*, 1964, 3, 321-325.
- MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of transition probabilities. *Journal of Experimental Psychology*, 1959, 47, 117-125.
- MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of word transition probabilities. *Journal of Experimental Psychology*, 1962, 63, 510-513.
- MAYZNER, M. S., & TRESSELT, M. E. Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, 1, No. 2.
- MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of multiple-solution anagrams. *Journal of Experimental Psychology*, 1966, 71, 66-73.
- MAYZNER, M. S., TRESSELT, M. E., & HELBOCK, H. An exploratory study of mediational responses in anagram problem solving. *Journal of Psychology*, 1964, 57, 263-274.