

The solution of anagrams: A reexamination of the effects of transition letter probabilities, letter moves, and word frequency on anagram difficulty*

GERALD A. MENDELSON and ANNE T. O'BRIEN
*Institute of Personality Assessment and Research
University of California, Berkeley, California 94720*

Word frequency (WF), number of letter moves, and solution word transition letter probabilities (TP) were related to anagram solution. The solution word TP measure was based on the relative frequencies of correct to incorrect bigrams within the pool of bigrams defined by the letters of the anagram rather than on the absolute frequencies of the correct bigrams. This bigram rank measure, which also took word length and letter position into account, was a powerful predictor of anagram difficulty ($p < .001$). Likewise, number of letter moves predicted anagram solution strongly ($p < .001$), but WF was only a marginal predictor ($.05 < p < .10$). In addition, there were no significant interactions among the three variables, nor was *anagram* TP consistently related to anagram difficulty. The results were interpreted in terms of an approach which combined elements of an hypothesis and an S-R mediational theory.

Mayzner and Tresselt (1959, 1962) have proposed, as part of their S-R mediational theory approach to anagram solution, that transition letter probabilities (TP) in both the anagram and the solution word are related to problem difficulty. They predicted and, in these two studies, found that *anagrams* composed of high-frequency bigrams were relatively difficult to solve but that, conversely, *solution words* composed of high-frequency bigrams were relatively easy to solve. They reasoned that high-frequency bigrams in the anagram retain their unit quality and, thus, are more difficult to permute in the process of rearranging the letters of the anagram. In regard to the solution word, they assumed that Ss initially rearrange the letters of the anagram into letter units which are relatively probable in terms of empirical frequencies. These initial rearrangements match the solution word if it is composed of frequent bigrams, but correct anagram rearrangements are likely to occur only late in the problem-solving sequence for low-TP words. Two other predictions derived from S-R mediational theory were also formulated and confirmed, namely, that word frequency is negatively related to anagram difficulty and that the number of letter moves required to go from the anagram to the solution word is positively related to anagram difficulty (Mayzner & Tresselt, 1958).

The findings in regard to both word frequency and letter moves have been consistently replicated (for the former, Mayzner & Tresselt, 1958, 1966; Dominowski,

1965, 1967; Warren & Thompson, 1969, inter alia; for the latter, Hunter, 1959, 1961; Dominowski, 1966), but neither of the predictions concerning TP have fared well. Mayzner and Tresselt (1966) and Warren and Thompson (1969) found that high-TP anagrams had longer solution latencies than low-TP anagrams, but Erlebacher (1962), Stachnik (1963), Mayzner and Tresselt (1963), and Dominowski and Duncan (1964) failed to confirm these findings. Variations in experimental procedures may account for the conflicting results, but it is likely that anagram TP does not have an important influence on Ss' responses (cf. Johnson, 1966). Research on solution word TP has likewise produced an inconsistent set of findings: Erlebacher (1962), Dominowski and Duncan (1964), and Warren and Thompson (1969) failed to replicate Mayzner and Tresselt's (1962) finding. However, subsequent to their initial study, Mayzner and Tresselt (1963, 1966) introduced a new and, it seems, more effective measure of solution word TP. Their original measure, the one usually employed in anagram research, consists of the sum of the frequencies of the successive bigrams which comprise the solution word. For the word APRON, the four bigram frequencies obtained from Underwood and Schulz (1960) are AP 102, PR 181, RO 433, and ON 786, and the word TP is 1502. The new measure takes into account the fact that bigram frequencies vary as a function of word length (WL) and letter position (LP). The bigram ON, for example, is a relatively common beginning or ending of three- and four-letter words but rarely begins or ends five-letter words. Using a sample of 20,000 words, Mayzner and Tresselt (1965) obtained frequency counts of the occurrence of bigrams in words of three to seven letters in each of the possible positions. Their new TP measure consists of the sum of bigram frequencies by WL and LP. By this method, APRON, though composed

*Work on this research was begun by the senior author at the Center for Advanced Study in the Behavioral Sciences, Stanford, California and completed at the Institute of Personality Assessment and Research, University of California, Berkeley. We wish to thank colleagues and staff of both institutions and Professor Geoffrey Keppel for their help and encouragement. Thanks are due, too, to Christopher Flindt and Barbara Gruen for their help in collecting the data. Ms. O'Brien is now at the University of Oregon.

of bigrams common in the language as a whole, has a relatively low TP score, i.e., in five-letter words, AP in Position 1-2, PR in Position 2-3, etc., are all rather uncommon. The two methods, then, may yield different predictions for a given solution word. Mayzner and Tresselt (1963) found the WL and LP measure to be clearly superior to their original measure.

Certain assumptions about the relationship between word TP and problem difficulty are shared by both measures. First, it is assumed that the bigram is the fundamental unit on which rearrangements of the anagram are based, an assumption which has been supported directly by Mayzner, Tresselt, and Helbock (1964) and indirectly by Dominowski (1968) and Tresselt and Mayzner (1965). Second, it is assumed that, in generating rearrangements of the anagrams, Ss will sample bigrams roughly in the order of their empirical frequencies. This assumption is consistent with the more general "spew" hypothesis of Underwood and Schulz (1960). They do differ, however, in regard to the specification of the pool of bigram frequencies from which S is assumed to be responding. For the original measure it consists of the general frequencies of bigrams in the language, while for the WL and LP measure it consists of the frequencies of bigrams in all possible letter positions in words of a specified length. The third shared assumption is that, however the bigram pool is defined, Ss respond from the pool as a whole. In both measures it is the frequency values of the bigrams in the solution word *relative to all other bigrams in the total pool* which is assumed to determine problem difficulty. This assumption seems questionable, particularly in light of the findings of Underwood and Schulz (1960), which suggest that the hypothesized "spew" process operates in conjunction with a selector mechanism (cf. Dominowski, 1967; Warren & Thompson, 1969). That is, S produces verbal units in the order of their frequency within a pool constrained by the characteristics of the stimulus input. This line of argument leads to the basic assumption of the present research that the pool of bigrams from which S responds in anagram solution is limited to those bigrams that can be formed from the letters of the particular anagram presented. Thus, it is predicted that anagram difficulty is a function of the frequency values of the bigrams in the solution word *relative to all other bigrams in the pool defined by the letters of the anagram*.

The manner in which solution word TP influences the process of anagram solution is the major, but not the sole, concern of the research to be reported. In addition, the relationships of word frequency and letter moves to anagram difficulty will be examined, not only to obtain further information about these relationships but also to assess the possibility of interactions between and among the three variables which have been most consistently associated with anagram solution. (To the best of our knowledge, no previous study has manipulated all three variables simultaneously.) The design of the study is

such that practice effects and the effect of anagram TP on solution can be investigated as well. Finally, an effort was made in this study to develop representative samples of solution words and the anagrams formed from solution words, so that previously neglected problems of the generalizability of findings could be avoided. In summary, then, the objective of the present research is to assess the validity of our reanalysis of the operation of solution word TP, but to do so within the context of a comprehensive look at the process of anagram solution generally.

METHOD

TP Measure

The primary methodological problem in this research was to devise a TP measure which indexes the comparative frequencies of the correct and incorrect bigrams that can be formed from the anagram. (By "correct" is meant those bigrams which do, in fact, comprise the solution word.) Given the assumptions outlined above, it seemed essential that this index should reflect the ordering of bigram frequencies, so a bigram rank measure was employed. The clearest and most direct way to describe this measure is to illustrate its derivation for a particular solution word, BEACH. First, a matrix of bigram frequencies was generated, as shown in Table 1. The rows are the 20 bigrams that can be formed from the five letters of the words, and columns represent each of the four positions in which a bigram can occur. The entries, taken from Mayzner and Tresselt's (1965) tables, are the frequencies with which each bigram appears in a given letter position in five-letter words.

Three bigram rank measures were obtained from this matrix, one based on a comparison of the correct bigrams to all others in the matrix, a second on the comparison of a given correct bigram to others in the same position, and the third on a comparison of a given correct bigram to others beginning with the same letter. In all three cases, the bigram rank score consisted of the sum, across all four correct bigrams, of the number of incorrect bigrams of greater frequency. By reference to Table 1, it can be seen that five incorrect bigrams exceed the value of BE₁₋₂, one exceeds the value of EA₂₋₃, five exceed the value of AC₃₋₄, and one exceeds the value of CH₄₋₅. The bigram rank score (total), then, is 12. The bigram rank score by position is 4: 1 in Column 1-2, 1 in Column 2-3, 2 in Column 3-4, and 0 in Column 4-5. Finally, the bigram rank score by letter is 1: The correct bigram has a higher frequency value than all other bigrams for the letters B, E, and C, but for AC₃₋₄ one bigram (AB₁₋₂) has a higher value. It was found in preliminary work, as well as in the present study, that the three bigram rank measures are very highly intercorrelated ($r > .925$), so in the present research only the first, the bigram rank (total), was utilized.

Three important features of these bigram rank measures should be noted. First, they are based only on the relative magnitude of bigram frequencies within the anagram; the absolute frequency values do not influence the measure. Second, no differential weighting procedure is employed. The TP score consists of the simple sum of the bigram rank scores for each of the four correct bigrams. Third, in line with the findings of Mayzner and Tresselt (1963, 1966), it is assumed that Ss base their implicit rearrangements on both relative frequency and letter position probabilities. We will consider each of these points further during the presentation of results.

Selection of Solution Words

Schwartz and Olson (1968) have pointed out that many allegedly single-solution anagrams can in fact yield two or more solutions. As an aid to Es, they subsequently published a list of all five-letter words in the *Webster's Third International*

Dictionary (1964) that cannot be rearranged to form another word. From this list, all words which might reasonably be expected to be in the vocabulary of an average college student were extracted. Three- or four-letter words which become five-letter words by the addition of a suffix or a final "s" without otherwise changing form and words in which a given letter appeared more than once were dropped. The frequency values of the remaining words were then obtained from the Kucera and Francis (1967) count. The words were next sorted into frequency categories, and the bigram rank scores described above were calculated for all words in the 6 to 10 (low) and 35 to 100 (high) categories. Finally, the words within each frequency category were sorted into three groups on the basis of their bigram rank scores: low (5-30), middle (50-70), and high (90 and above). Thus, six groups of potential solution words (high-low frequency by high-middle-low bigram rank) were established, from each of which five words were randomly selected. The logic of the experiment demands that two conditions be met in the selection of solution words: (1) that within the high (or low) frequency conditions, the low, middle, and high bigram rank cells have frequency means which are approximately equal and (2) that within the low (or middle or high) bigram rank conditions the low- and high-frequency cells have bigram rank means which are approximately equal. If either condition was not satisfied, a new random sample of five words was drawn in a cell and this procedure was followed until the necessary equalities were obtained. For the final selection, the frequency means for the high and low conditions were 61.1 and 8.0, respectively, and for the low, middle, and high bigram rank conditions, the mean scores were 19.4, 60.5, and 109.0, respectively.

We have gone into some detail about the process by which solution words were selected because, as Johnson (1966) points out, one rarely knows in anagram research how words were, in fact, chosen for an experiment. Consequently, it is usually impossible to ascertain the nature of the population from which the sample of words was drawn or to detect the possible biases in the sampling itself. This, of course, raises questions about the generality of findings. The procedure adopted in the present experiment was designed to meet these problems by specifying precisely the population of words to which results can be generalized. There is, however, no way of determining whether this population is representative of words in general; it can only be stated that it is a broad and functionally significant population.

Construction of Anagrams

It is evident that the various anagrams that can be formed from a given solution word are not of equal difficulty. Thus, in order to test hypotheses about the effects of frequency and TP, several anagrams should be formed from each solution word. In the present experiment, the anagrams were constructed by selecting 15 of the 119 permutations of the letters of a five-letter word; five of them were one move, five were two move, and five were three move permutations. The selection was based on the procedure employed by Dominowski (1966): Of the one-move anagrams, one began with the first, one with the second, one with the third, one with the fourth, and one with the fifth letter of the solution word. The same principle was followed for two- and three-move anagrams. Further, over the entire set of 15 anagram orders, each letter of the solution word appeared with approximately equal frequency in each letter position of the anagram. All 30 solution words were presented in each of the 15 anagram orders thus established. It should be noted that this procedure makes it impossible for Ss to develop a letter order bias which could aid them in the solution process or which could have differential effects on particular solution words.

Construction of Anagram Lists

Fifteen sets of 30 anagrams were constructed such that, across

Table 1
Bigram Frequency Matrix for BEACH

		1-2	2-3	3-4	4-5
B	E	41*	0	6	10
	A	15	0	3	0
	C	0	0	0	0
E	H	0	0	0	0
	B	0	1	0	0
	A	21	134*	75	0
A	C	0	4	8	0
	H	0	0	0	0
	B	93	6	0	0
C	E	0	0	0	0
	C	1	14	51*	0
	H	6	0	0	0
H	B	0	0	0	0
	E	24	263	55	1
	A	22	33	0	0
	C	0	0	0	0

*Asterisked frequencies are those for the bigram by position entries which comprise the solution word.

the sets, each solution word appeared once in all of its 15 anagram orders. Within sets there were five blocks of anagrams, each block consisting of six anagrams drawn randomly from each of the six frequency by bigram rank cells. Further, each of the 15 anagram orders appeared once in the first and once in the second half of a given list, and across sets each anagram order appeared in each of the 30 possible positions. This method of composing the lists was intended both to prevent any confounding of possible practice effects with the primary variables of study, TP, frequency, and letter moves, and to permit a careful examination of practice effects in anagram solution.

Subjects and Procedure

The Ss were 30 male and 30 female undergraduates who participated in the experiment as part of a course requirement. Four Ss, two male and two female, responded to each of the 15 sets.

The anagrams were typed on 3 x 5 in. index cards in oversized capital letters separated by two spaces. Ss were tested individually by three different Es, two female and one male undergraduate. In all, 33 anagrams were presented to each S, the first three serving as a warm-up and as a check on the S's comprehension of the instructions. (There was only one S who failed to understand the task.) A maximum time of 60 sec was allowed for each anagram, and Ss were asked to give their responses verbally. Following a response, E noted the latency but waited approximately 5 sec before recording the solution and going on to the next anagram. This delay was intended both to provide a short break between anagrams and to allow Ss to correct an incorrect response if they wished to do so. Other than the pause between anagrams, the entire set was presented continuously. For the average S the entire procedure, including instructions, took about 30 min.

Anagram Solution Scores

Three different solution scores were obtained from these data. The first was the number of Ss within each of the letter move conditions who attained a correct solution for a given word. The second was the median latency of solution for a given word, again within each of the letter move conditions. It should be noted that, for both of these measures, a score is based on the

responses of 20 Ss, i.e., 4 Ss in each of the five anagram forms within a letter move condition, and that for a given word 20 different Ss are in each of the letter move conditions. The third score is also a latency measure, but one which was devised to reduce the effects of the problem caused by the distribution of anagram latency scores when a time limit has been imposed. The problem is threefold: (1) The distribution is bimodal, for the anagrams are most often solved rapidly or not at all. In the present data, 36% of the anagrams were solved within 10 sec and 38% were not solved. (2) A failure to attain a solution has conventionally been given a latency score equivalent to the time limit, in the present study 60 sec. This means that quantitatively a solution attained in, for example, 56 sec is differentiated very little from one not attained at all. Likewise, the difference between solution times of 2 and 6 sec and between a time of 56 sec and no solution is quantitatively the same. This makes little sense psychologically and, since both problems are created by the frequent occurrence of failures to obtain a solution, a conventional (e.g., log or square root) transformation of the raw latency scores is of very limited utility. A final aspect of this problem concerns the insensitivity of the median latency measure in some cases. For example, if all 20 Ss attempting two-move anagrams formed from a given word failed to attain a correct solution, the median latency score would be 60 sec. If, in contrast, nine Ss had attained correct solutions, the median latency would still be 60 sec. In either case, one would conclude that the anagram was difficult, but the evident difference in difficulty would be completely obscured by the measure.

The most direct way to minimize these problems is to devise a solution score which transforms the latency data in terms of the obtained distribution of solution times. The transformation employed here was a very simple one: A failure to obtain a solution received a score of 2, a solution time of 8-60 sec received a score of 1, and a solution time of 7 or fewer sec received a score of 0. As noted before, 38% of the anagrams were not solved; of the remaining 62%, 31% were solved by the end of 7 sec. Thus, the three scoring categories correspond as closely as possible to equal frequency intervals. In effect, this measure is a compromise between a raw latency and a number of solutions measure. The latter, which is essentially a 0-1 score, is, of course, insensitive to latency differences among solutions but does discriminate clearly between a successful solution and a failure to attain a solution. The transformed score just described retains this feature but also reflects latency differences. One final point in regard to this score is that it provides a measure suitable for the analysis of the performance of individual Ss.

It was predicted that anagram difficulty would be directly related to the bigram rank measure and to number of letter moves but inversely related to word frequency.

RESULTS

Because of the design of the experiment, it is possible to explore many hypotheses about the process of anagram solution. Consequently, the results will be presented in four sections, the first of which concerns the hypotheses stated above.

TP, Frequency, and Letter Moves

The effects of these three variables on the probability and latency of solution were evaluated initially by a three-way repeated-measures analysis of variance, with bigram rank and frequency level serving as the between effects and number of letter moves as the within effect. In this design, the "subjects" are the 30 solution words, each of which yields three observations, one each for one-, two-, and three-move anagrams.

Three different analyses of variance of this form, each using a different method of scoring the solution data, were performed. The results of the three analyses are quite similar, but strongest for the transformed scores. Since it is a priori the most meaningful and flexible of the three and, it appears, the most sensitive as well, we will limit our consideration to this measure. The hypotheses concerning the bigram rank and letter move variables were confirmed by the data. The mean solution scores for the low, middle, and high bigram rank words were .74, 1.13, and 1.38, respectively, with $F(2,24) = 13.32$, $p < .001$, and MS error = .233. In the low condition, 46% of the anagrams were solved within 7 sec and 80% within the time limit. The corresponding percentages for the middle condition were 27% and 60% and for the high condition were 18% and 44%. A trend analysis performed on the transformed score data yielded an $F(1,24) = 26.22$ for the linear component ($p < .001$) and < 1 for the quadratic component. A bit more than 98% of the variation among the bigram rank conditions can be predicted from linear regression.

The mean solution scores of the one-, two-, and three-move anagrams were .69, 1.20, and 1.37, respectively, with $F(2,48) = 58.35$, $p < .001$, and MS error = .064. The percentage figures were 79% (52% within 7 sec), 57% (23% within 7 sec), and 48% (16% within 7 sec), respectively. The trend analysis for these data yielded $F(1,48) = 108.04$ ($p < .001$) for the linear component and $F(1,48) = 8.65$ ($p < .01$) for the quadratic component. A bit more than 93% of the variation among the letter move conditions is attributable to linear regression.

The difference between high- and low-frequency word means was in the predicted direction; the means were .99 and 1.18 and the percentages were 66% (35% within 7 sec) and 56% (25% within 7 sec). However, the difference is of only marginal significance [$F(1,24) = 3.53$, $.05 < p < .10$, MS error = .233]. None of the interactions attained or even approached significance.

In order to assess the effects of the bigram rank and frequency variables in greater detail and to ascertain whether letter order in the anagram or anagram TP affected solution scores, the data for one-, two-, and three-move anagrams were analyzed separately. In these analyses of variance, bigram rank and frequency were the between effects and order or anagram bigram total were repeated effects. The results can be presented briefly. The bigram rank effect was significant in all three analyses—at the .025 level for one-, the .001 level for two-, and the .01 level for three-move anagrams. The frequency effect was significant for two-move anagrams at the .025 level, of marginal significance for three-move anagrams ($.05 < p < .10$), and insignificant for one-move anagrams. Anagram order, i.e., whether the anagram began with the first, second, etc., letter of the solution word, had no discernible effect on solution scores. Finally, Mayzner and Tresselt's prediction that high-TP anagrams would be more difficult to solve than low-TP

anagrams was not confirmed. In fact, the evidence of this study strongly suggests that, aside from number of letter moves, no anagram variable has an appreciable effect on problem difficulty.

To summarize the results of these analyses: (1) anagram difficulty is a direct linear function of the bigram rank measure. The relationship is strongest for two-move anagrams and least strong for one-move anagrams, but the interaction between bigram rank and number of moves is trivial; (2) anagram difficulty increases as a function of the number of letter moves required to go from the anagram to the solution word. The evidence indicates that the relationship has a quadratic as well as a linear component, in that three-move anagrams are only slightly more difficult than one-move anagrams; (3) word frequency is inversely related to anagram difficulty, but the relationship is only of marginal significance; (4) the three variables (bigram rank, frequency, and number of letter moves) combine additively rather than interactively in their effect on anagram difficulty; (5) letter order in the anagram, i.e., whether the anagram begins with the first or second, etc., letter of the solution word, and anagram bigram total do not have a significant effect on anagram difficulty.

In the analyses reported above, the data of the male and female Ss were combined, but subsequent analyses showed no significant differences between the sexes in overall solution scores or in the pattern of relationships summarized above.

Analyses of Individual Protocols

As indicated previously, the "subjects" in the analyses just reported were, in effect, the solution words and, thus, the results obtained pertain to grouped data. It is important to ask, consequently, whether the same pattern of relationships holds for individual problem solvers as well as for group data. To answer this question, product-moment correlations were obtained for each S between the three independent variables and his or her 30 solution scores. For the bigram rank measure, the correlation was in the predicted direction for 57 of the 60 Ss and in 39 cases ($p < .10$). For the number of moves variable, the correlation was in the predicted direction for 59 of the Ss and for 34 of them $p < .10$. Frequency, however, was a predictor of marginal significance: The correlation was in the predicted direction for 44 of the Ss, but in only nine cases was $p < .10$. Finally, multiple correlation coefficients were obtained for each S and 50 out of 60 were significant at beyond the .05 level. Thus, the analyses of the grouped data and the individual protocols lead to essentially similar conclusions about the relationships between the three independent variables and anagram difficulty.

Further Analyses of the Bigram Rank Measure

The distinctive feature of the present research is the

reconceptualization of the manner in which word TP is related to anagram solution. It is clear that the bigram rank measure based on this reconceptualization is a powerful predictor of anagram difficulty, but it remains possible that, for these data, other measures of word TP would predict as well, i.e., that the bigram rank measure is redundant. Consequently, further analyses were undertaken to ascertain whether the bigram rank measure accounts for variation in solution scores which is not accounted for by other measures of word TP. In doing this, it was also possible to examine three assumptions that underlie the bigram rank measure: (1) that it is the relative frequencies of the bigrams which comprise the word rather than their absolute frequencies which are critical; (2) that all four bigrams contribute to prediction; and (3) that TP measures which use bigram frequencies based on word length and letter position are more strongly related to anagram solution than those which use bigram frequencies in the language as a whole.

These issues were examined by obtaining product-moment correlations between the solution scores of the 30 words and several different measures of word TP. Among these were three used in previous research: (1) the sum of the frequencies of the four bigrams that comprise the solution word [using the Underwood and Schulz (1960) frequency count]; (2) the sum of the four bigram frequencies taking word length and letter position into account (designated Sum WL and LP); (3) a bigram rank measure devised by Dominowski (1965) which had a moderate but significant relationship to anagram difficulty in his study. This last measure is obtained by counting for each correct bigram the number of other bigrams beginning with the same letter that have a higher frequency in the language as a whole and then summing over the four bigrams. (This measure is designated BR-Dom.) An additional bigram rank measure, also based on frequencies in the language as a whole, was devised by the present authors. First the Underwood and Schulz frequencies for each of the 20 bigrams that can be formed from the five letters of the solution word were obtained. Next, for each correct bigram, the number of incorrect bigrams of greater frequency were counted. Finally, the rank counts for the four correct bigrams were summed. This measure, it should be noted, assumes that the S is responding from the pool of bigrams defined by the letters of the anagram, but the frequencies on which it is based do not take WL and LP into account. It is designated BR-total. The last measure was the bigram rank score used in the previous analyses and it is designated BR-WL and LP. Of the five measures, two (the first and second) are based on absolute frequencies and three on relative frequencies. Further, the first, third, and fourth are based on frequencies in the language as a whole, while the second and fifth are based on WL and LP frequencies. The correlations between the TP measures and solution

Table 2
Correlations Between Various Measures of Solution
Word TP and Solution Scores

Measure	Word Frequency		Average r
	High	Low	
Sum Bigram Frequencies	.330	-.324	-.003
Sum WL and LP	-.413	-.585*	-.505
BR-Dom	.026	.318	.176
BR-Total	.173	.583*	.398
BR-WL and LP	.671†	.846†	.773

* $p < .05$

† $p < .01$

scores are shown in Table 2 for high- and low-frequency words separately and, in addition, the average r over the two frequency conditions is included.

The results in Table 2 permit clear answers to the questions raised at the beginning of this section. First, the bigram rank measure based on WL and LP frequencies accounted for at least twice as much of the variance of solution scores as any other of the word TP measures. Second, the two measures (Sum WL and LP and BR-WL and LP) based on word length and letter position frequencies yield higher correlations than those based on frequencies in the language as a whole. Third, of the latter measures, the one that indexes the relative frequency of the correct and incorrect bigrams within the word (BR-total) is the best predictor. Likewise, of the two measures based on word length and letter position frequencies, the better predictor is, of course, the one that is based on relative rather than absolute frequencies. In the light of the previous literature, it should also be noted that the sum of solution word bigram frequencies is not significantly related to solution scores and that, for high-frequency words, the relationship is opposite that predicted originally by Mayzner and Tresselt. However, the present results do replicate their later findings in regard to the WL and LP measure. These findings certainly can go a long way in resolving inconsistencies of the previous literature on solution word TP.

The last aspect of these analyses to be reported concerns the assumption that all four bigrams of the solution word contribute to the prediction of anagram scores. For the first bigram, the average r between the bigram rank scores and the solution scores was .551, for the second it was .613, for the third .170, and for the fourth .686. The finding that all bigrams but the third were significantly correlated with solution scores is contrary to the assumption that the first bigram should be particularly important given our left to right reading and writing habit (Johnson, 1966; Warren & Thompson, 1969). It is also contrary to the results obtained by Mayzner and Tresselt (1966). They found that only the first two bigrams predicted solution scores and that for the last two the direction of correlation tended to be opposite to expectation. That result was not replicated in the present study using their Sum WL and LP measure

either; that is, the correlations between bigram frequencies and solution scores were .324 for the first, .342 for the second, -.035 for the third, and .604 for the fourth bigram. It is difficult to account for the discrepancy but, since Mayzner and Tresselt's study was concerned with testing hypotheses about multiple-solution anagrams, their sample of solution words and anagrams can in no sense be considered representative. Whatever the source of the discrepancy, though, two conclusions seem warranted: First, the issue of the degree to which each of the bigrams predicts solution scores must be considered unresolved; but second, the procedure of taking the unweighted sum of the four bigram rank scores as the measure of word TP is justified in the present study.

Practice Effects

The presence of practice effects was evaluated by examining the performance of Ss in each of the five blocks of six anagrams within a list. The main effect for blocks was not significant, a result which is consistent with Johnson's (1966) conclusion that "no important general practice effect has been demonstrated [p. 379]."

DISCUSSION

The most important finding of this experiment is that solution word TP, as indexed by the bigram rank measure, is a powerful predictor of anagram difficulty. The clarity of this finding contrasts sharply with the confusing and contradictory nature of previous data concerned with solution word TP. It is possible, of course, that this contrast is attributable to some peculiarity in the sample of solution words employed. In order to evaluate this possibility, the data of three previous studies were reanalyzed by correlating the bigram rank measure with the median latency scores provided in the text. The correlations for the data presented by Mayzner and Tresselt (1959) were .621 for low and .632 for high TP anagrams; when word frequency was partialled out, the correlations became .657 and .571. The correlation for the data presented by Mayzner and Tresselt (1962) was .806; with word frequency partialled out, it was .807. Finally, for the data presented by Warren and Thompson (1969), the correlation was .519 and the partial correlation was .396. In none of these studies was there any attempt to sample a number of anagrams for each solution word, but, nevertheless, the correlation coefficients are rather high. It should also be noted that the range of frequency values in each of the three studies exceeded by far that of the present study. It appears, then, that our results are generalizable across experiments and solution words, and we can turn to the questions of what accounts for the predictive success of the bigram rank measure and what it implies about the process of anagram solution.

Clearly, the bigram rank measure used in this study bears a close resemblance to other measures of solution

word TP. The one critical regard in which it differs from its predecessors, however, is the assumption that, in reorganizing the letters of the anagram, S restricts the sampling of bigrams to those contained in the anagram. This assumption means that, rather than the absolute probabilities of the occurrence of bigrams in the language as a whole, it is the *relative probabilities of the bigrams in the restricted set* which determine solution difficulty. Further, it is assumed that the relevant probabilities are those which take word length and letter position into account. The findings of this study, particularly those concerned with the comparisons of the various solution word TP measures (Table 2), support both assumptions in that (a) measures based on relative probabilities are better predictors than those based on absolute probabilities and (b) measures based on WL and LP bigram frequencies are better predictors than those based on bigram frequencies in the language as a whole. It seems reasonable to conclude that the bigram rank measure attains its predictive power by combining these two assumptions.

Implicit in the discussion above, particularly in such terms as "rearranging the letters of the anagram" and "sampling of bigrams," is a far more active and goal-directed problem solver than S-R mediational theory suggests. Indeed, in light of the data of this study, the process of anagram solution seems best described in terms of an hypothesis conception. Like previous investigators, we assume that the basic unit with which S works in reorganizing the anagram is the bigram but, in the present formulation, the pool of bigrams by position which can be formed from the anagram is construed as constituting a limited set of potential hypotheses. The first hypothesis formulated will, in general, be the most probable of the set. If the hypothesis is not confirmed by the subsequent testing, the next most probable hypothesis will be formulated and so on, i.e., it is assumed that the order in which the successive hypotheses are formulated will roughly correspond to the relative probabilities of the bigrams in the set. It follows, then, that when the bigrams of the solution word are relatively improbable, they will appear late in the sequence of hypothesis formation and solution will consequently be delayed.

Thus far we have discussed only one aspect of the anagram solving process, the formulation of hypotheses. The second aspect of the process involves the question of how Ss move from partial reorganizations of the anagram to the production of whole words, that is, solutions. What follows, though consistent with the data, must be considered tentative until more direct evidence is available. After the bigram hypothesis is formulated, Ss may test it in one or both of two ways: by attempting to match a whole word to the partial reorganization or by attempting to arrange the remaining letters about the hypothesized pair. The first process is construed as essentially associative in nature—the bigram serves as a stimulus which evokes whole word responses. Here, too,

a selector mechanism is undoubtedly operative, for the range of possible word responses is quite constrained: The word must have the hypothesized bigram in the correct position, must contain a specific number of letters, and must include the remaining letters of the anagram. It is at this point in the process that solution word frequency would influence anagram solving, for frequent words would be more readily evoked than infrequent ones.

The second process of hypothesis testing, arranging the remaining letters about the hypothesized pair, presumably occurs in a manner which parallels the formation of the initial hypothesis. That is, once the initial bigram has been selected, the pool of possibilities would be consequently reduced. From this pool, Ss would then select the most probable letter or letters of those remaining to form a subsidiary hypothesis and so on. The process just described is consistent with the reports of many Ss and with an important empirical observation. Solution words with only one highly probable bigram are not necessarily easy to solve; if the remaining bigrams are improbable, solution is nevertheless difficult to attain. This may also explain, in part, why the bigram rank scores of all four bigrams are correlated in the same direction with solution scores. Once all or most of the letters have been thus arranged, the letter sequence must still be matched to a whole word. In this sense, the two processes of hypothesis testing are interdependent, but it is sometimes the case that a S will arrange all the letters of the anagram and then realize, to his surprise, that he has produced the correct word.

In presenting this hypothesis theory description of the anagram solving process, we have focused entirely on the role of solution word variables, relative bigram frequencies within the solution word, and word frequency. It is clear from the data, however, that one anagram variable (number of letter moves) is also a powerful determinant of problem difficulty. The present results replicated previous findings that one-move anagrams are much easier to solve than those that require multiple moves. It was also found that two-move anagrams were a bit easier to solve than three-move anagrams. Mayzner and Tresselt's explanation of the effect—that, as the number of letter moves increases, the number of implicit reorganizations required also increases—is not very convincing. Not only does it fail to account for the finding that the relationship is negatively accelerated, but also it fails to consider that, from the S's standpoint, all anagrams contain precisely the same number of possible permutations of the letters. If S knew that an anagram required but one move for solution, the number of reorganizations could be greatly reduced, of course, but it is not until after the anagram has been solved that he has that information.

A more likely approach is to consider that the anagram is a visual stimulus which may have greater or lesser similarity to the solution word. A rough index of

similarity is the number of bigrams in the anagram which are also bigrams in the solution word. For one-move anagrams the number is two or three, for two-move anagrams it is zero or one, and for three-move anagrams it is always zero. Moreover, in one-move anagrams the correct bigrams will also be in the correct order with respect to each other. Thus, one-move anagrams will on the whole look considerably more like the solution word than will two- or three-move anagrams. The visual resemblance between two- or three-move anagrams and the solution word will be at best slight, but a bit greater for the former than for the latter. If it is assumed that the greater the similarity between the anagram and the solution word, the greater is the likelihood that the anagram will evoke the solution word associatively and directly, it becomes clear why one-move anagrams are relatively easy. Support for this assumption comes from a comparison of the distributions of the solution latencies for one-, two-, and three-move anagrams: 35% of the one-move anagrams were solved within 3 sec of the presentation of the stimulus, but only 11% of the two-move and 7% of the three-move anagrams were solved so rapidly. Now it is important to note that Wallenhorst (1965) reports that the mean latency for free associations to 100 words from the Kent-Rosanoff list was 1.60 sec (SD = .88). Thus, solution latencies of 0 to 3 sec are comparable to the latencies obtained in free associations to word stimuli. After 3 sec, 44% of the one-move, 47% of the two-move, and 42% of the three-move anagrams were solved. It appears, then, that the obtained difference in difficulty is a function of what occurs immediately after the presentation of the anagram stimulus; if the solution does not occur at once, then the number of required letter moves seems to have little effect on solution latency. This perceptual-associationistic interpretation of the effect of the letter move variable is consistent with the report of Ss that they sometimes simply "see" the solution; indeed, some Ss adopt the strategy of first "looking for" a solution before trying a more analytic and systematic approach.

One final point requires discussion before concluding this paper. As in other studies, more frequent words proved easier to solve, but, in contrast to other studies, the relationship was only of borderline significance. This difference may be attributable to the fact that the range of word frequencies sampled in the present study is narrower than that in previous studies. In particular, very low-frequency words, i.e., those whose frequency of occurrence in the Thorndike-Lorge count is less than once per million, were not used in this study. This suggests that in anagram solution the word frequency effect may be pronounced for extreme values but only minor for midrange values. In fact, the data of Dominowski (1967) and Mayzner and Tresselt (1958) indicate that solution latencies are not a monotonic function of word frequency, but a precise delineation of

the relationship must await more systematic and carefully controlled studies (cf. Eriksen, 1963) on the relationship of word frequency to recognition thresholds.

It should be apparent from this discussion that no single theoretical position is adequate to account for the anagram solving process. The view taken here has been that the sequential rearrangement of the anagram stimulus is best described in terms of hypothesis formation. The sample of hypotheses is limited by the letters of the anagram, but the order in which hypotheses are formulated is determined by response hierarchies which conform roughly to the probability of occurrence of verbal units, particularly bigrams, in the language. The process of going from the anagram itself or from partial reorganizations of the anagram to the solution word was seen as conforming to an S-R model. Thus, the process of anagram solution may differ from anagram to anagram and may vary, as well, at different stages of the solution sequence. In short, S-R mediational and hypothesis theory approaches to the analysis of anagram solution are complementary; together they seem to possess considerable heuristic and explanatory power.

REFERENCES

- Dominowski, R. L. Problem difficulty as a function of relative frequency of correct responses. *Psychonomic Science*, 1965, 3, 317-318.
- Dominowski, R. L. Anagram solving as a function of letter moves. *Journal of Verbal Learning & Verbal Behavior*, 1966, 5, 107-111.
- Dominowski, R. L. Anagram solving as a function of bigram rank and word frequency. *Journal of Experimental Psychology*, 1967, 75, 299-306.
- Dominowski, R. L. Anagram solving as a function of letter-sequence information. *Journal of Experimental Psychology*, 1968, 76, 78-83.
- Dominowski, R. L., & Duncan, C. P. Anagram solving as a function of bigram frequency. *Journal of Verbal Learning & Verbal Behavior*, 1964, 3, 321-325.
- Eriksen, C. W. Perception and personality. In J. M. Wepman and R. W. Heine (Eds.), *Concepts of personality*. Chicago: Aldine, 1963. Pp. 31-60.
- Erlebacher, A. *Parameters of anagram tasks*. (Doctoral dissertation, University of Wisconsin) Ann Arbor: University Microfilms, 1962. No. 63-591.
- Hunter, I. M. L. The solving of five-letter anagram problems. *British Journal of Psychology*, 1959, 50, 193-206.
- Hunter, I. M. L. Further studies on anagram solving. *British Journal of Psychology*, 1961, 52, 161-165.
- Kucera, H., & Francis, W. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.
- Johnson, D. M. Solution of anagrams. *Psychological Bulletin*, 1966, 66, 371-384.
- Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of letter order and word frequency. *Journal of Experimental Psychology*, 1958, 56, 376-379.
- Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of transition probabilities. *Journal of Psychology*, 1959, 47, 117-125.
- Mayzner, M. S., & Tresselt, M. E. A function of word transition probabilities. *Journal of Experimental Psychology*, 1962, 63, 510-513.
- Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of word length and letter position variables. *Journal of Psychology*, 1963, 55, 469-475.
- 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, 13-31.

- 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.
- Schwartz, R., & Olson, R. Are alleged single-solution anagrams really single-solution? *Journal of Verbal Learning & Verbal Behavior*, 1968, 1, 567-568.
- Stachnik, T. Transitional probability in anagram solution in a group setting. *Journal of Psychology*, 1963, 55, 259-261.
- Tresselt, M. E., & Mayzner, M. S. Anagram solution times: A function of individual differences in stored bigram frequencies. *Journal of Experimental Psychology*, 1965, 70, 606-610.
- Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning*. New York: Lippincott, 1960.
- Wallenhorst, R. Some relations between reaction time and choice of response in word association. *Psychological Reports*, 1965, 17, 619-626.
- Warren, M. W., & Thomson, W. J. Anagram solution as a function of transition probabilities and solution word frequency. *Psychonomic Science*, 1969, 17, 333-334.

(Received for publication April 9, 1973;
revision received January 28, 1974.)