NOTES AND DISCUSSIONS

FAMILIARITY OF LETTER-SEQUENCES AND TACHISTOSCOPIC IDENTIFICATION

An observer (O) can identify fewer letters of an unfamiliar sequence than of a familiar sequence when such sequences are presented tachistoscopically for exposures too brief for complete identification. The information O receives from an exposure, however, is approximately constant when redundancy of letter sequences is taken into account. One of the clearest demonstrations of this fact has been produced by Miller, Bruner, and Postman.[1] The purpose of this note is to suggest a slight modification in the calculation of informational content of their stimulus-material, which will be shown to lead to an improvement in the orderliness of their results.

Miller *et al.* constructed pseudo words (sequences of eight letters) at zero, first, second, and fourth orders of approximation to English according to the procedure described by Shannon.[2] These pseudo words were presented to *O*s tachistoscopically for durations ranging from 10 to 500 m.sec. The *O*s identified and reported as many letters as they could.

When computing the information Miller *et al.* used both Shannon's and their own estimates of the redundancy of such sequences and argued as follows.[3] Letters chosen at random are all equally likely and the redundancy of the zero-order sequence is 0%. Letters in the first-order approximation are selected with frequencies approximately equal to those found in printed English. These frequencies so vary from letter to letter that the average amount of information per letter is less by about 15%. The information in the first-order sequence is thus 85%. The estimate of the redundancy of the second-order approximation is based on the distribution of frequencies of letter pairs as they occur in printed English. This reduces the information to 71%. The estimate of the redundancy of the fourth order is 43%, and hence the relative amount of information per letter here is 57%.

Two different measures of *O*s' performances in identifying letters were used by Miller *et al.*: *a placement-score* and a *letter-score*. To test the hypothesis that the amount of *information* received by O per exposure is con-

[1] G. A. Miller, J. S. Bruner, and Leo Postman, Familiarity of letter sequences and tachistoscopic identification, *J. gen. Psychol.*, 50, 1954, 129-139.

[2] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, 27, 1948, 379-423, 623-656.

[3] Shannon, Prediction and entropy of printed English, *Bell Syst. Tech. J.*, 30, 1951, 50-64.

stant, both scores were corrected for the average amount of information per letter by multiplying the scores by appropriate values of relative information. If the hypothesis is true and the increase in the scores is proportional to the decrease in the relative amount of information in higher-order approximations, then such a correction should eliminate variance in *O*s' performances attributable to familiarity of letter sequences.

The corrected letter scores of the *O*s in Miller *et al.* are plotted here in Fig. 1.[4] Although the curves are closer together than would be comparable

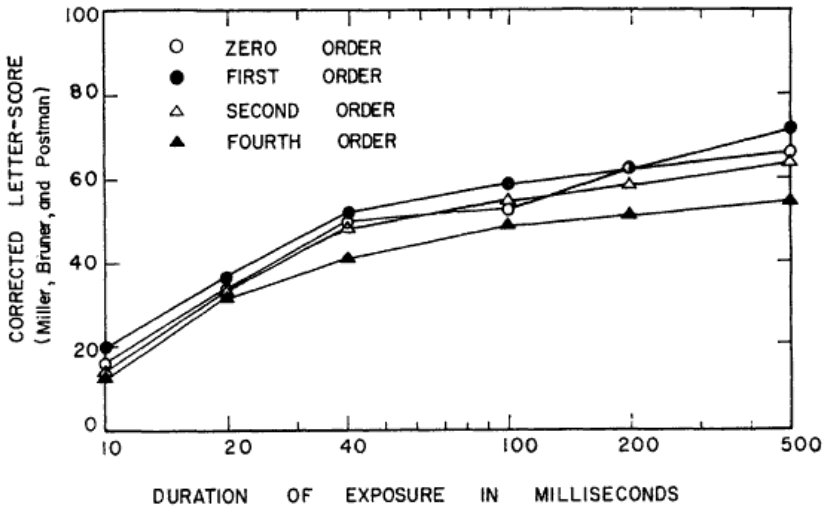

FIG. 1. CORRECTED LETTER-SCORES AS A FUNCTION OF DURATION OF EXPOSURE FOR SEQUENCES OF EIGHT LETTERS AT FOUR ORDERS OF APPROXIMATION TO ENGLISH

curves for untreated letter scores, there is still some variability left. There is too great compensation for the redundancy at the second and fourth order. This led Miller *et al.* to suggest that "the letter scores do not provide a fair test of the hypothesis."[5] Better support for the hypothesis, however, was provided by the corrected placement scores.

An examination of the corrected letter-scores in Table 3 of Miller, *et al.* makes clear that this residual variability is indeed orderly.[6] In that table all corrected letter scores for the fourth-order approximation are lower than

[4] This graph is based on data in Table 3 in Miller, Bruner, and Postman, *op. cit.*, 137. An error in that table has been corrected here. The *letter-score* for the fourth order at 40 m.sec. exposure should be 41.6, not 46.6 as shown in the original table.
[5] Miller, Bruner, and Postman, *op. cit.*, 138.
[6] *Ibid.*, 137.

those for the second order, and all corrected letter-scores for the second-order approximation are lower than those for the first order. Although this might logically be expected at relatively long durations of exposure, as Miller *et al.* point out, the reason for such an apparently systematic relation at short durations of exposure may lie in the underestimation of the relative amounts of information in the second- and fourth-order sequences.

The estimates of redundancy in various sequences used by Miller *et al.* are based on the assumption of infinitely long sequences of letters. Such estimates are also correct for short messages of zero order and first-order approximations to English, but overestimate redundancy, and thus underestimate relative information, for higher orders. Take, for instance, a fourth-order sequence. Each letter in such a sequence is limited by the preceding three letters. This means that one can predict a given letter in such a sequence with maximal efficiency if he knows the preceding three letters. Suppose, however, that a fourth-order sequence contains only two letters. The first letter in the sequence of two letters is not preceded by any other letters and hence its redundancy and relative information value are equal to the average redundancy and relative information per letter in the first-order sequence. The second letter is preceded by only one letter and its informational content is, therefore, equivalent to the average of the second-order sequence. In such a case, of course, the fact that the sequence is a part of a longer fourth-order message is irrelevant as far as redundancy and predictability are concerned.

On the basis of this general line of argument it may be appropriate to revise the amounts of relative information in Miller *et al.*'s fourth-order stimulus-words. The first letter in the sequence is constrained by its individual frequency of occurrence in printed English, and contains, on the average, 85% relative information. The second letter is preceded by just one letter, its informational content being thus equivalent to the average of letters in the second-order sequence, viz. 71%. The third letter, by the same argument, can be estimated to contain, on the average, 63% relative information. For the remaining five letters the estimates based on infinitely long sequences, 57% relative information per letter, are quite appropriate. The average information per letter in a fourth-order sequence of eight letters, on the basis of the above reasoning, is thus estimated at 63%, instead of 57% as used by Miller *et al.* The same analysis applied to the second-order words yields an average relative information value of 73%. Zero order and first order are not affected.

Correcting the letter-scores in the Miller *et al.* experiment by multiplying

them with these new values of relative information yields the functions plotted in Fig. 2.

Although there is still some variability left, the fit is decidedly better than in Fig. 1. Thus it seems that when more appropriate estimates of redundancy in higher-order approximations are used, Miller *et al.*'s corrected

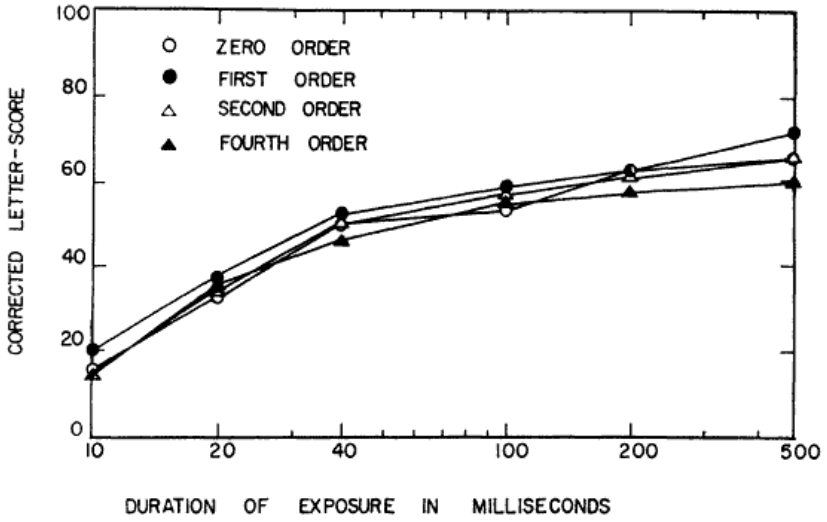

FIG. 2. DATA IN FIG. 1 REPLOTTED ON THE BASIS OF THE REVISED ANALYSIS

letter-scores also lend good support to their hypothesis. The amount of information received by an *O* familiar with the structure of the language is constant per tachistoscopic exposure.

It should be mentioned that the revised estimates of the amount of relative information do not improve the orderliness of corrected placement-scores, but the fit for those scores is no worse than that for the corrected letter-scores. Furthermore, Miller *et al.* would probably agree that other variables, beside the information content of letter-sequences, are more likely to operate when *O*'s task involves correct identification of the position of letters than when the correctness of position is disregarded.

University of Toronto ENDEL TULVING

LOCALIZED NORMALIZATION OF TILTED LINES

Gibson, and Gibson and Radner reported that a line tilted by a small angle either from the vertical (defined as a plumb line) or from the hori-

* This research was supported by a grant from the National Science Foundation.