Notes and Comment

Multidimensional letter similarity: A confound with brightness?

D. J. K. MEWHORT and M. L. DOW
Queen's University
Kingston, Ontario K7L 3N6, Canada

In a recent paper, Gilmore, Hersh, Caramazza, and Griffin (1979) have reported both a confusion matrix for uppercase letters and indices of letter similarity derived from the matrix. The matrix was based on an experiment of heroic proportions: Using stimuli composed of bright dots presented on a computer-controlled cathode-ray tube (CRT), each letter was shown 60 times to each of 20 subjects. The exposure duration, calibrated separately for each subject every 72 trials, was set to limit performance at 50% correct. Similar work has been reported by Townsend (1971) using conventional tachistoscopic procedures.

Gilmore et al. (1979) suggest, however, that the previous work, particularly the confusion matrices offered by Townsend (1971), has limited value. Two main reasons were offered. First, Townsend's experiment may have been too small to provide reliable off-diagonal estimates of probability. Reliable estimates are required, of course, for subsequent calculations of letter similarity. Second, Townsend used conventional tachistoscopic techniques involving stimuli prepared by typing black letters on a white background. A number of investigators prefer the convenience of computer-controlled experiments and, as a result, use stimuli displayed on a CRT. Because of differences in the type font associated with the two techniques, Gilmore et al. suggest that indices of letter similarity taken from Townsend's (1971) work are unlikely to apply to the CRT stimuli.

While Gilmore et al. (1979) justify their experiment in terms of an implicit contrast to Townsend's (1971) work, they do not provide explicit comparisons. Fortunately, both experiments involved a similar procedure. Both, for example, presented a single letter on each trial and adjusted the exposure duration to limit overall performance to about 50% correct. Thus, in both experiments, the diagonal of the confusion matrix indicates the difficulty of letters relative to each other. Curiously, the two experiments do not agree; there was little correlation between the measures of relative difficulty obtained in the two tasks. Treating each letter as the unit of analysis, the correlation of the proportion correct in the two studies was only .212. (With $df = 24$, a correlation

of .388 is required for significance at the 5% level.)

The low correlation confirms that there is a substantial difference between conventional and computer-controlled techniques. Presumably, although they do not comment on the difference, Gilmore et al. (1979) would attribute the difference to the change in type font. Alternately, the difference could reflect other procedural changes. Unfortunately, Gilmore et al. do not provide a number of important details concerning their procedure. They do not elaborate, for example, how their computer program wrote successive dots on the CRT. The omission is particularly unfortunate: Their procedure involved a 50% performance criterion, a value achieved by varying the exposure duration. The range for individual subjects was 10 to 70 msec, with a mean exposure of 33 msec. Further, the stimuli were presented as bright dots on a dark screen, the screen was housed in a dark room, and the subjects were given 5 min for dark adaptation. Under such conditions, the exposure times are remarkably long. Indeed, our experience with (apparently) similar equipment conflicts with their report. Using the display algorithm described by Mewhort (1978), for example, we cannot induce dark-adapted subjects to produce errors when reading single letters presented for 5 msec, a value half of the Gilmore et al. minimum and one-sixth of their mean. To obtain errors at such an exposure, we have to decrease the intensity of the display monitor to a point at which we question the stability of the equipment for displaying dots. As a second illustration of the point, consider a recent paper by Mewhort, Hearty, and Powell (1978) on the sequential-blanking effect. They used three exposure durations (all less than 33 msec) and a five-letter sequential display. Even averaging across conditions which illustrate the blanking illusion, performance was about 80% (see their Table 1). If performance on five-letter displays—including cases of blanking—is so high, how could Gilmore et al. fix performance for one-letter displays at 50% with a mean exposure greater than the longest for the five-letter case?

There are two main possibilities which might explain the low performance given relatively long exposures. Gilmore et al. (1979) may have set the intensity of the display fairly low. They do not mention setting a threshold with a luminance manipulation before starting the experiment, however. Alternately, one can disrupt performance by using a very slow dot-brightening algorithm; Gilmore et al.

may have used such an algorithm. Again, details of their technique were not provided.

The intensity values and the display algorithm jointly determine the relative brightness of each character. If, for example, one were to refresh the dots continuously, a character formed from a small number of dots would be brighter than one formed with a large number of dots.

To check on the potential relation between a character's brightness and its size, we correlated accuracy (the scores on the diagonal of the confusion matrix) with the number of dots in each character. The number of dots was obtained from the representation of the characters provided by Gilmore et al. (1979, Figure 1). The Pearson correlation coefficient was highly significant [$r(24) = -.873$, $p < .01$]. Thus, characters involving a small number of dots were reported more accurately than those involving a large number of dots. Such a relation is exactly what one would expect if brightness controlled relative per-

formance, as would be the case using a display algorithm which refreshed dots without regard for the size of the character. Because of the apparent confounding with brightness, we recommend caution in interpretation and use of their data.

REFERENCES

- GILMORE, G. C., HERSH, H., CARAMAZZA, A., & GRIFFIN, J. Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics*, 1979, **25**, 425-431.
- MEWHORT, D. J. K. DIS: An n-channel tachistoscope algorithm. *Behavior Research Methods & Instrumentation*, 1978, **10**, 756-760.
- MEWHORT, D. J. K., HEARTY, P. J., & POWELL, J. E. A note on sequential blanking. *Perception & Psychophysics*, 1978, **23**, 132-136.
- TOWNSEND, J. T. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 1971, **9**, 40-50.

(Received for publication June 29, 1979;
accepted August 22, 1979.)