

A non-parametric analysis of recognition experiments¹

Irwin Pollack, UNIVERSITY OF MICHIGAN
Donald A. Norman, HARVARD UNIVERSITY

Abstract

A non-parametric method for evaluating the results of recognition memory experiments and psychophysical detection experiments is presented. The method is based upon an ordinal analysis of recognition performance, which transforms the results of recognition tests into equivalent results for a forced-choice experiment.

Problem

The standard paradigm in recognition memory experiments is that a set of stimulus items is first presented to S; these items are then intermixed with a new set of items and the combined set of items is then presented for recognition. S must identify each item as either old (o) or new (n). In a parallel manner, in psychophysical detection experiments, S's task is to identify each trial as a signal trial (s) or as a non-signal trial (ns). Suppose that under one experimental condition, S responds o to 90% of the old items and 1% of the new items (or responds s on 90% of signal trials and on 1% of non-signal trials). Under a different condition, S responds o to 99% of the old and 10% of the new items (or responds s on 99% of signal trials and 10% of non-signal trials). Which condition yields the better performance? What statistic of the data would allow us to assess changes in performance resulting from changes in experimental conditions? Conventional recognition measures, which ignore the role of n or ns responses, are inadequate.

The questions are easily answered when detailed quantitative theories of behavior exist. Within the context of psychophysical detection experiments, for example, there are numerous analyses that might be used (Luce, 1963; Swets et al., 1961). Since recognition memory experiments are formally similar to the psychophysical experiments, it is tempting to suggest that the same theories may be applied to both areas (Egan, 1958). Strong assumptions, however, must be made about the underlying mechanisms in order to make these analyses. Can meaningful quantitative specifications of performance be achieved with weaker assumptions?

Method

An adequate measure of recognition performance must consider two response measures: correct acceptances (o responses to old stimuli or s responses on signal trials) and incorrect acceptances (o responses to new stimuli or s responses on non-signal trials). Any assumed particular combination of the two measures implies a particular theory of behavior. Until more is known about recognition, we wish to avoid specific theories.

There is a class of experiments, however, which minimizes the effects of S's arbitrary utilization of o or s responses. The basic experiment is a two-alternative forced-choice experiment: e. g., S is presented with an old item and a new item and is instructed to select the old item; or S is presented with a signal trial and a non-signal trial and is instructed to select the signal trial. The statistic of percent correct is sufficient to summarize the results of forced-choice tests. Thus, we would clearly have a solution to our original problem if we could convert the results of a recognition experiment into an equivalent forced-choice score.

Consider any arbitrary performance by S as a point on the unit square of Fig. 1. For any particular experi-

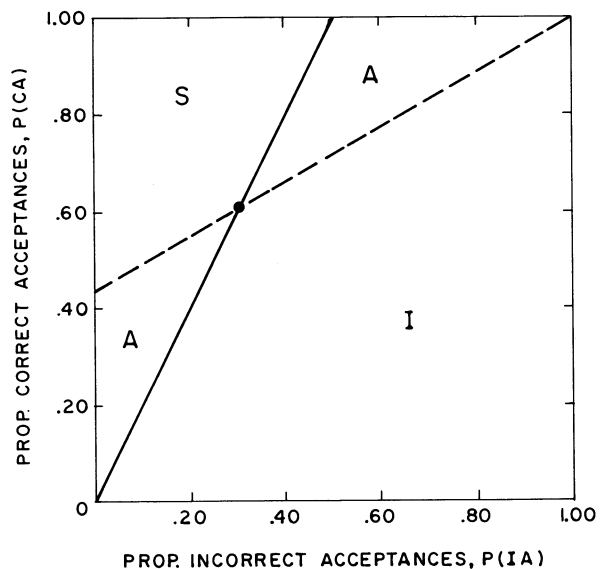


Fig. 1. An ordinal analysis of recognition experiments. Any arbitrary performance level is represented by a point in the unit square. The upper and lower bounds are defined by straight line segments drawn from (0,0) and from (1,1) through the point. Relative to the designated point, if the performance from a second condition falls in the region marked S, it is superior to the performance from the first condition; if it falls in the region marked I, it is inferior to the first condition; if it falls in the regions marked A, an ambiguous result is obtained in which, in the absence of further information, it is not possible to discriminate between the performance of the two conditions.

mental condition, there is some relation between correct and incorrect acceptances. This relation is called the operating characteristic. The actual location on the curve is determined by S's response preferences, but all points on the operating characteristic reflect equivalent recognition performance. The operating characteristic is usually unknown, for its exact shape is a function of the assumed underlying psychological mechanism and, hence, the as yet unknown psychological theory. All is not lost, however, with only one point on the curve, Norman (in press) has shown that the unit square may be divided into the four regions of Fig. 1, which relate the performance from other experimental conditions to any single point. The regions, defined in legend of Fig. 1, allow an ordinal ranking of the results of several different experiments with each other, even in the absence of information about the true operating characteristic.

D. Green (personal communication, 1963) has demonstrated that the area under the operating characteristic is theoretically equal to the proportion of correct responses attainable in a two-alternative forced-choice experiment. But if we do not know the entire operating characteristic, we can still use the results of Fig. 1 to establish upper and lower bounds on the areas and, hence, on the equivalent forced-choice results. In the absence of further assumptions, we have chosen the average of the areas subtended by the upper, and by the lower, bounds as the measure of recognition performance, i. e., the sum of the I region plus one half the A regions.

Results

Figure 2 presents the loci of points on the unit square which yield equivalent areas, as defined by our method. Except for the extreme slopes of the ascending and descending portions, these operating characteristics are similar in shape, when drawn upon the unit square, with those of constant d' , the signal detectability index of the theory of signal detection (Swets et al., 1962); with those of constant k , the ratio operating characteristic of the method of free recall (Egan et al., 1961); and with those yielded by various high-, low-, and multi-threshold theories (Luce, 1963).

The area estimation scheme proposed here thus appears to be a simple way of assessing performance in recognition experiments in terms of an equivalent forced-choice score without assuming a specific theory of the nature of the recognition process.

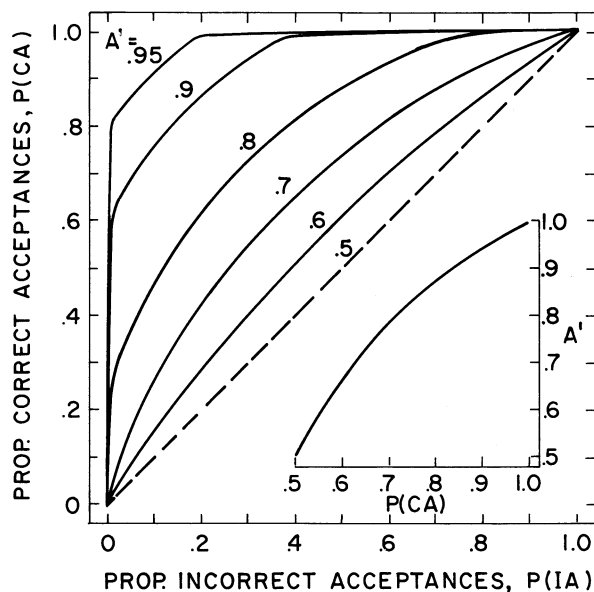


Fig. 2. Equivalent conditions for constant average area in terms of the ordinal analysis of Fig. 1. The insert in the lower right represents the variation in A' for combinations of correct and incorrect acceptances which add to 1.0, i. e., where $P(CA) + P(IA) = 1.0$.

References

- EGAN, J. P. Recognition memory and the operating characteristic. Indiana University, Hearing and Communication Laboratory, AFCRC TN 58 51, AD-152650. 15 June 1958.
- EGAN, J. P. GREENBERG, G. Z., & SCHULMAN, A. I. Operating characteristics, signal detectability, and the method of free response. *J. Acoust. Soc. Amer.*, 1961, 33, 933-1007.
- LUCE, R. D. A threshold theory for simple detection experiments. *Psychol. Rev.*, 1963, 70, 61-79.
- NORMAN, D. A comparison of data obtained with different false alarm rates. *Psychol. Rev.*, in press.
- SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, 63, 301-340.

Note

1. The second author was supported by a National Science Postdoctoral Fellowship at the Harvard Center for Cognitive Studies. The authors wish to acknowledge the assistance of H. J. LaDuke for programming the IBM 7090 computer for the results of Fig. 2.