

# Signal detection analysis of recall paired-associates learning<sup>1</sup>

MILTON D. SUBOSKI

QUEEN'S UNIVERSITY, KINGSTON, ONTARIO

Results replicating those of Suboski, Pappas, & Murray (1966b) on the use of confidence ratings in the RTT paradigm are presented. In addition, ROC curve analyses show that Ss' ability to discriminate between correct and incorrect responses is a function of prior confidence and correctness.

In a recent application of signal detection analysis to paired-associates learning (Pappas & Suboski, 1966), recognition accuracy was found to be an increasing function of prior recall confidence. Those results were interpreted as suggesting multiple states of recall associative strength. Although a number of dependencies have been shown between recall response probabilities and confidence ratings (Suboski, Pappas, & Murray, 1966a, b), functional relationships of the type obtainable by signal detection methods have not yet been demonstrated for the RTT paradigm with recall on both tests. Such analyses are reported in the present paper.

One of the original purposes of this experiment was to determine whether the shifts in response probabilities and confidence across tests (Jones, 1962; Suboski et al, 1966b) are a result of information gained in the tests or rather a function of time as would be expected, for example, from von Wright's (1959) response competition (D) hypothesis. However, the intertest intervals chosen, 5, 15, and 45 sec., failed to produce any differences of importance, perhaps because of the limited temporal range sampled. As a consequence, all of the analyses reported here are of data pooled across the intertest intervals.

## Method

The Ss, nine volunteer undergraduates from the Queen's University summer session, learned nine 10-item lists. Each list, presented by hand at a 4 sec. per pair exposure rate, consisted of a random pairing of the letters A through J as stimulus items with the numbers 0 through 9 as responses. After presentation of each list, Ss were given two tests in which they worked through 10-page response booklets at a 7 sec. per page rate, cued by E to turn pages. Each page contained a stimulus letter, a space for the response and a 5 cm "confidence" line with the left end labelled "guess" and the right end labelled "sure." The Ss were instructed to draw a vertical line across the confidence line to indicate their degree of confidence in the correctness of their written response. Each S was run individually and within each set of three lists presented to each S, the 5,

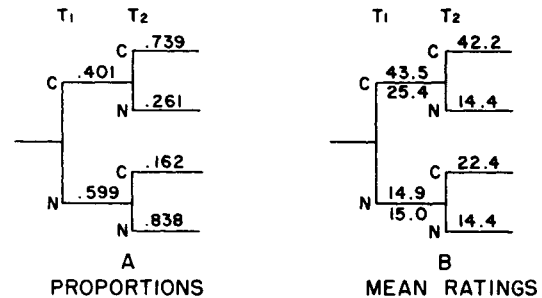


Fig. 1. Proportions (A) and pooled mean confidence ratings (B) of correct and incorrect responses for the two tests. Separate T<sub>1</sub> mean ratings are shown for items correct and incorrect on T<sub>2</sub>.

15, and 45 sec. intertest intervals (occurred in a random order.

## Results

Figure 1 displays proportions correct (C) and incorrect (N) and mean confidence ratings. Except for the overall lower proportions correct, these results essentially replicate those of Suboski et al (1966b). In particular, it may be noted that second test (T<sub>2</sub>) mean confidence for N-C items was higher than T<sub>1</sub> confidence or T<sub>2</sub> confidence for N-N items. Proportions correct and response shifts between tests for correct and incorrect items across confidence categories (successive 1 cm lengths of the confidence line), shown in Fig. 2, were also very similar to the Suboski et al results.

Figure 3 shows Type II (response conditional) oper-

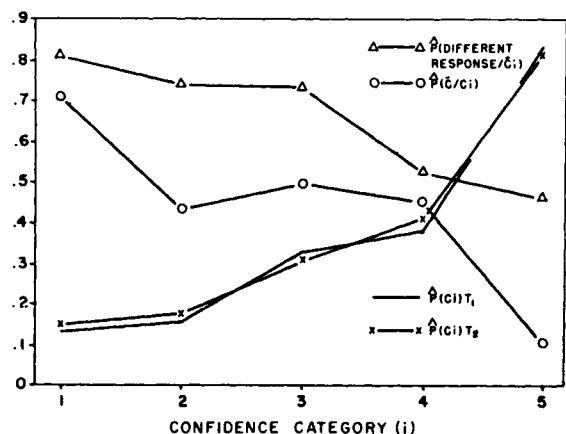


Fig. 2. Proportion correct shown separately for the two tests as a function of confidence, and proportion of shifts to a different response from T<sub>1</sub> to T<sub>2</sub> shown separately for correct and incorrect on T<sub>1</sub>.

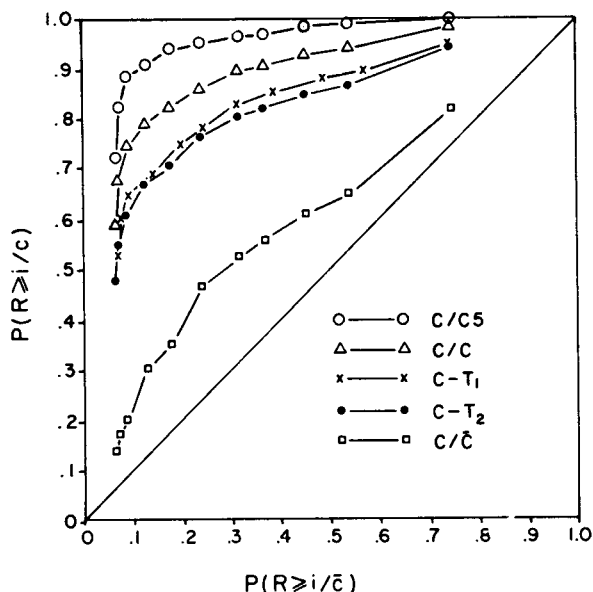


Fig. 3. ROC curves for different classifications of the data.

ating characteristic (ROC) curves (Egan & Clarke, 1966) for a number of classifications of the data. The curves were constructed by obtaining the proportions, pooled across Ss, lists, and items, of assignment of confidence category  $i$  or stricter (more confident) as the confidence category is progressively made more lax (less confident). The confidence ratings were measured to the nearest millimeter, providing 51 categories, 0-50. The 11 points plotted for each curve in Fig. 3 start with category 50 and move from category 46 to category 1 in 5 mm steps. For each curve, proportions for correct items are plotted against proportions for all incorrect items on that test trial, regardless of prior or subsequent correctness. Upward left displacement of the ROC curves from the chance line (positive diagonal) and from each other, indexes the ability of Ss to discriminate correct from incorrect responses (Egan & Clarke, 1966).

The ROC curves for the two test trials are very similar and indicate comparable ability of Ss to discriminate between correct and incorrect responses on each test. If, however, only items which were correct on  $T_1$  are selected, the ROC curve for  $T_2$  performance shows increased discriminability with greatest discriminability shown for items that were correct with high confidence on  $T_1$ . The ROC curve for items incorrect on  $T_1$  but correct on  $T_2$  shows

the lowest discriminability, but clearly well above chance. Normal-normal axes plots of the curves (not shown) appear linear through the major portion of each curve and have slopes less than unity.

#### Discussion

The present results are in substantial agreement with those of Pappas & Suboski (1966) in showing that  $T_2$  performance is a function of both correctness and confidence of  $T_1$  response in the RTT paradigm and extend the findings to the case of recall on both tests. These results further support the hypothesis that a continuum of associative strength results from paired-associates learning and is reflected by the confidence ratings.

Both mean  $T_2$  confidence and the  $T_2$  ROC curve show that items correct on  $T_2$  after having been incorrect on  $T_1$  can be discriminated from incorrect items. A similar result obtained by Pappas & Suboski (1966) could be attributed to the more sensitive recognition test on  $T_2$ , an explanation that does not apply to the present results. Another possible explanation is that information may be gained on  $T_1$  and applied on  $T_2$  (Jones, 1962). The high degree of similarity between the  $T_1$  and  $T_2$  ROC curves argues against such an explanation as does the virtually identical proportions correct across tests typically found in RTT paradigms. Such an explanation remains tenable, however, if it is assumed that  $T_1$  information gain and between test forgetting cancel.

#### References

- Egan, J. P., & Clarke, F. R. Psychophysics and signal detection. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology*. New York: McGraw-Hill, 1966. Pp. 211-246.
- Jones, J. E. All-or-none versus incremental learning. *Psychol. Rev.*, 1962, 69, 156-160.
- Pappas, B. A., & Suboski, M. D. Recognition operating characteristics as a function of prior recall confidence. *Psychon. Sci.*, 1966, 6, 83-84.
- Suboski, M. D., Pappas, B. A., & Murray, D. J. Confidence ratings in recall paired-associates learning. *Psychon. Sci.*, 1966a, 5, 147-148.
- Suboski, M. D., Pappas, B. A., & Murray, D. J. Confidence ratings in recall paired-associates: The RTT paradigm. *Psychon. Sci.*, 1966b, 5, 315-316.
- Wright, J. M. von. Forgetting and interference. *Soc. Sci. Fennica, Comm. Hum. Litt.*, XXVI, 1959, 1, 1-124.

#### Note

- I thank Anthony Gambaretto, Maryann Kernychny, and Ollan White for running Ss. This research was supported in part by grants from The Society of the Sigma Xi and the Queen's University Arts Research Committee. Facilities of the Queen's University Computer Center were used for data analysis.