

Confidence ratings in recall paired-associates: The RTT paradigm¹

MILTON D. SUBOSKI, BRUCE A. PAPPAS AND D. J. MURRAY

QUEEN'S UNIVERSITY, KINGSTON, ONTARIO

Having Ss rate their confidence in the correctness of their responses in the RTT recall paired-associates paradigm produced proportions of correct and incorrect responses in substantial agreement with all-or-none predictions. The confidence rating results, however, are inconsistent with all-or-none theory, suggesting instead a continuum of S-R associative strength.

In a previous study in which Ss rated their confidence in the correctness of their responses in the RTRT paired-associates paradigm (Suboski, Pappas, & Murray, 1966), results were obtained which suggested that confidence ratings reflect the strength of the underlying stimulus-response association. Also indicated was a continuum of states of S-R association rather than the two states, learned and unlearned, required by all-or-none theory. The present study extends the use of confidence ratings to the RTT paradigm, from which Estes and associates (Estes, 1960; Estes, Hopkins, & Crothers, 1960) obtained strong evidence for all-or-none learning.

Method

The Ss, 13 paid undergraduates, learned 10 10-item lists. Each list consisted of a random pairing of the letters A through J as stimulus items and the numbers 0 through 9 as responses. The S-R pairs were placed on 3 x 5 in. cards and presented simultaneously by hand at a 4-sec. exposure rate by a "flash-card" technique to all Ss. Presentation of each list was followed after 15 sec. by two tests, 15 sec. apart. In the tests, Ss worked through ten-page response booklets at a 7-sec. per page rate, cued by E to turn pages. Each booklet page contained a randomly selected stimulus letter typed at the top, a space in which to write 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 write in the correct response number and to draw a vertical line across the confidence line to indicate their degree of confidence in the correctness of their written response. The instructions emphasized answering and rating every item, and using as many positions on the confidence line as possible. Confidence ratings were categorized by measuring the distance from the left (guess) end of the confidence line to the S drawn line. The five categories were successive 1 cm lengths of the line.

Results and Discussion

Part A of Fig. 1 shows proportion of pooled correct (C) and incorrect (N) responses for T₁ and conditional proportions for T₂. Overall proportions of correct

responses showed virtually no change from T₁ (.551) to T₂ (.550). These data differ from those presented by Estes et al primarily in the higher proportion of N-C shifts. By all-or-none theory, proportion of N-C shifts estimates the guessing probability. This possibility could account for the present data. The stimulus and response item sets were completely known to the Ss initially. Thus, the formation of an S-R association could be expected to preclude the use of the response involved for any other stimulus item. This would reduce the size of the guessing subset of responses and thereby raise the guessing probability and consequently increase the proportion of N-C shifts. Since the overall proportion of responses reused on the same test trial for other stimulus items was .259 for correct responses and .602 for incorrect, a reduced guessing subset seems a reasonable explanation for the high proportion of N-C shifts, and in accord with all-or-none predictions.

Part B of Fig. 1 shows mean confidence ratings, with T₁ means shown separately for items correct and incorrect on T₂. Items correct on both tests have high, and virtually identical mean confidence on both tests. This contrasts with our previously reported RTRT results where a significant ($p < .01$ by sign test) increase in confidence from T₁ (39.2) to T₂ (45.4) occurred. Mean ratings for N-N items show low confidence and only a slight change across tests, although ten of the 13 Ss showed a decrease. Items on which a C-N shift occurred had lower T₁ confidence than C-C items ($p < .01$, all significance levels reported are for sign tests on individual S data) and significantly ($p < .01$) lower confidence on T₂ than on T₁.

All of the results so far reported are consistent with a combination of all-or-none theory and forgetting.

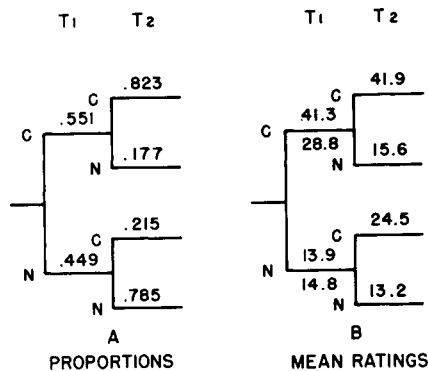


Fig. 1. Proportions (A) and mean confidence ratings (B) of correct and incorrect responses for the two tests.

Mean confidence for N-C items, however, is not. Confidence increased from T_1 to T_2 ($p < .05$) and further, confidence is significantly ($p < .05$) higher on T_2 for N-C items than for C-N items. These results are not in agreement with the hypothesis that N-C items are merely guessed correctly, but suggest instead that associative strength is somehow altered for these items from T_1 to T_2 .

Figure 2 displays proportions for several aspects of the data as a function of confidence category. As in the RTRT case, proportion of responses correct is an increasing function of confidence and shows only slight changes from T_1 to T_2 . Comparison of individual S proportions between adjacent confidence categories revealed a significantly ($p < .05$) higher proportion correct for the higher category of each pair for all except the comparison between categories 2 and 3.

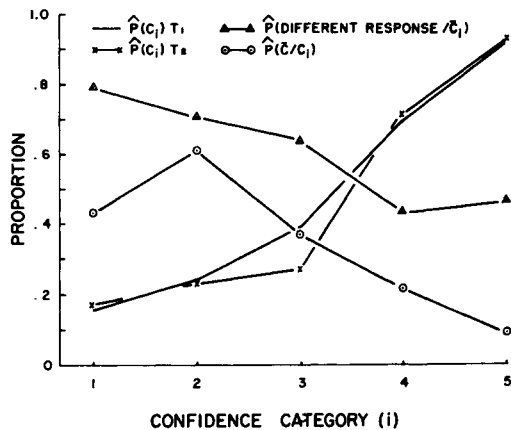


Fig. 2. Proportion correct shown separately for the two tests as a function of confidence, and proportion of shifts to a different response from first to second test shown separately for correct and incorrect on the first test.

The two other curves of Fig. 2 show the proportion of shifts to a new response from T_1 to T_2 , i.e., a C-N shift for items correct on T_1 , and simply the use of a different response, whether correct or not, for items incorrect on T_1 . Although both curves show decreases across categories, no significant sign test comparisons between adjacent or extreme categories resulted for the curve for incorrect items. Similar tests for the C-N curve showed that categories 1 and 2 differed ($p < .01$) from categories 4 and 5 respectively, with no adjacent categories significantly different. The overall higher proportion of response shifts from incorrect items than from correct items (.707 to .177) is the single result of this study which is not completely consistent with the hypothesis that the confidence ratings reflect a continuum of strengths of S-R association.

When an S gives a response to a paired-associate stimulus item, he quite obviously can do more than indicate whether the response is likely to be right or wrong, the two states predicted by all-or-none theory. Rather, he can provide what amounts to a good estimate of the probability that the response is correct. This would seem to indicate a continuum of states of associative strength rather than just the learned or unlearned of all-or-none theory.

References

- Estes, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, 67, 207-223.
- Estes, W. K., Hopkins, B. L., & Crothers, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, 60, 329-339.
- Suboski, M. D., Pappas, B. A., & Murray, D. J. Confidence ratings in recall paired-associates learning. *Psychon. Sci.*, 1966, 5, 147-148.

Note

- Support for this research was provided by N.R.C. Grants Nos. APA 160 and APT 126. The facilities of the Queen's University Computer Center were used for data analysis.