UMEÅ PSYCHOLOGICAL REPORTS
No. 99 1976

Department of Psychology
University of Umeå

COMBINATION RULES IN MULTIPLE CUE PROBABILITY LEARNING
I. RELATION TO TASK CHARACTERISTICS AND PERFORMANCE

Bengt-Åke Armelius          Kerstin Armelius
Armelius, B-A., and Armelius, K. Combination rules in multiple cue probability learning I. Relation to task characteristics and performance. Umeå Psychological Reports No. 99, 1976. - The rules used by subjects in multiple cue probability learning (MCPL) was investigated by means of verbal reports given by subjects at the end of learning of a two cue MCPL-task. Eight tasks varied factorially with respect to task predictability, cue-criterion correlations and the sign of the cue intercorrelation. In addition there were two orthogonal tasks. 47 of the 100 subjects gave verbal descriptions that were classified as inconsistent or incomplete, 7 were classified as a single rule i.e., the same rule was used for all cue combinations, and 46 were classified as multiple rules i.e., different rules were used for different parts of the cue matrix. 77% of the 53 combination rules were found to account for the systematic variance in subjects' responses. While performance was related to the characteristics of the tasks, the frequency of combination rules was not. Subjects with multiple rules reached the highest level of performance and subjects with a single rule the lowest. It was concluded that the formulation of combination rules is important for performance in MCPL. Further developments of the method to extract combination rules were discussed.

In multiple cue probability learning (MCPL) subjects' performance is usually described in terms of the lens-model equation (Hursch, Hammond & Hursch, 1964; Tucker, 1964). The lens-model equation described the...
performance in relation to the task, but gives little or no information about the psychological processes which produce this performance. Consequently, whereas we know a lot about how subjects perform in MCPL tasks (see Slovic & Lichtenstein, 1971) progress towards a psychological theory about MCPL has been slow.

Some early studies on the processes in MCPL focused on subjects' ability to learn tasks where cues were related to the criterion in a nonadditive way (Brehmer, 1969; Summers, Summers and Karkau, 1969). The results of these studies showed that subjects are able to learn nonadditive rules relating cues and criterion. Other studies have demonstrated the usefulness of nonadditive models for description of subjects' behavior in MCPL tasks (e.g., Hoffman & Wiggins, 1968; Hoffman, Slovic & Rohrer, 1968; Einhorn, 1970). In general, however, the linear model has been found to account for as much variance in subjects' responses as any other model (Hoffman, 1960; Yntema & Torgersen, 1961; Andersson, 1968). In these latter studies the experimenter compared the amount of variance in subjects' responses accounted for by a number of models selected on theoretical grounds.

A common characteristic of most early studies on the processes underlying inference behavior is that they focus on the question whether a configural or a linear model is the better for describing the results. This is an unfortunate restriction of an important problem. This is due mainly to the controversy over the relative efficiency of clinical and statistical prediction raised by Meehl (1954). Another restriction is that all models have covered the complete cue matrix with one mathematical equation, although a number of studies (Slovic, 1966; Brehmer, 1972) have shown that subjects change their prediction rules when the cue values are inconsistent with expectations. The restriction seems to be a consequence of the fact that the lens-model equation has been the primary analytical tool. While the lens-model equation is well suited to describe subjects' performance it is not very useful for describing the processes underlying this performance, (see Lichtenstein, Earle & Slovic, 1975).
Recently, Böhmer (1974) has suggested another analytical framework for the study of inference behavior. He has shown that subjects inference behavior may be seen as a hypothesis testing activity, where subjects test different hypothesis during the learning of an inference task. Following this approach, Armelius and Armelius (1975) in a pilot study used subjects' verbal descriptions to formulate models of the inference processes. They found that 50% of the subjects were able to formulate combination rules that were sufficiently systematic and consistent to be translated into explicit prediction equations which could then be simulated by means of a computer program. These combination rules accounted for a large proportion of the systematic variation in subjects' responses in a two-cue MCPL-task. The rules were either linear combinations of the two cue-values with one prediction equation for the whole cue matrix, or configural rules with different prediction equations for different parts of the cue matrix. Subjects who had formulated combination rules reached a higher level of performance than subjects who had not.

The Armelius and Armelius (1975) study was, however, only a first step towards a more detailed analysis of the processes in MCPL. Since only one task was used in the study the influence of various task characteristics on the frequency and type of rules could not be studied. The present experiment, therefore, is an attempt to study the influence of task characteristics on the type and frequency of rules subjects use.

Method

Subjects. Eight-two undergraduate psychology students from the University of Umeå participated in the experiment to fulfill a course requirement and 18 students of education from the University of Umeå were paid to participate in the experiment.

The experiment was conducted in two stages, a Learning stage and a Test stage. Between the two stages forty trials without feedback were given.
Learning stage. The Learning tasks were two-cue MCPL-tasks. The design was a 2 (Sign of the cue intercorrelation) x 2 (Levels of cue-criterion correlation for cue 1: \( r_{el} = .60 \) and .80) x 2 (Levels of task predictability: \( R^2_e = 1.00 \) and .70) x 5 (Blocks of 20 trials) factorial design with repeated measures on the last factor. The correlation between the second cue and the criterion was \( r_{e2} = .00 \) for all tasks.

In addition to these factorially combined tasks there were two orthogonal learning tasks with the same levels of the cue-criterion correlations as the other learning tasks: \( r_{el} = .60 \) and .80, and \( r_{e2} = .00 \). Table 1 gives the task characteristics for the learning tasks. The details of the Learning stage are presented in Armelius and Armelius (1976b).

Table 1. Task characteristics for the Learning tasks.

<table>
<thead>
<tr>
<th>Learning tasks</th>
<th>( r_{ij} )</th>
<th>( r_{el} )</th>
<th>( r_{e2} )</th>
<th>beta(_{el} )</th>
<th>beta(_{e2} )</th>
<th>( R^2_e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.80</td>
<td>.60</td>
<td>.00</td>
<td>1.67</td>
<td>-1.33</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>.70</td>
<td>.60</td>
<td>.00</td>
<td>1.18</td>
<td>-.82</td>
<td>.70</td>
</tr>
<tr>
<td>3</td>
<td>.60</td>
<td>.80</td>
<td>.00</td>
<td>1.25</td>
<td>-.75</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>.30</td>
<td>.80</td>
<td>.00</td>
<td>.88</td>
<td>-.26</td>
<td>.70</td>
</tr>
<tr>
<td>5</td>
<td>-.80</td>
<td>.60</td>
<td>.00</td>
<td>1.67</td>
<td>1.33</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>-.70</td>
<td>.60</td>
<td>.00</td>
<td>1.18</td>
<td>.82</td>
<td>.70</td>
</tr>
<tr>
<td>7</td>
<td>-.60</td>
<td>.80</td>
<td>.00</td>
<td>1.25</td>
<td>.75</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>-.30</td>
<td>.80</td>
<td>.00</td>
<td>.88</td>
<td>.26</td>
<td>.70</td>
</tr>
<tr>
<td>9</td>
<td>.00</td>
<td>.60</td>
<td>.00</td>
<td>.60</td>
<td>.00</td>
<td>.36</td>
</tr>
<tr>
<td>10</td>
<td>.00</td>
<td>.80</td>
<td>.00</td>
<td>.80</td>
<td>.00</td>
<td>.64</td>
</tr>
</tbody>
</table>

The learning tasks were presented in booklets. On the face of each page in the booklet the cues were presented as two bars numbered from one through twenty. The value of each cue was represented as the shaded part of the bar. The criterion value was presented as a number between one and thirty on the other side of the page. On each of the 100 training trials, the subjects (a) observed the two cue values, (b) gave their prediction on an answer sheet and (c) observed the correct criterion.
value. Subjects were allowed to work at their own pace. They were not informed about the structure of the tasks. They were told to base their predictions on the values of the cues, and it was emphasized that due to the nature of the tasks they should not expect to be correct on each trial.

Test stage. The cue and criterion values of the Test stage were identical to those of the last block in the Learning stage except for the order of trials. For the Test stage the subjects were instructed to focus their attention on how they solved the tasks. They were told to think about how they made their predictions so that they could describe their strategy after completion of the Test stage.

Combination rules. After completing the Test stage the subjects filled in a questionnaire with the following questions:

1. Describe in your own words how you made your predictions. If you used any particular rules or methods please try to describe them as completely as possible.

2. Now describe the cases for which you considered it especially difficult to make the predictions. Also describe how you made your predictions in those cases.

3. Now describe the cases for which you found it especially easy to make predictions, and how you made the predictions for these cases.

The answers to the questions were classified in one of three categories by two judges (the authors) independently. The categories were:

No rule. All incomplete or inconsistent descriptions of combination rules were placed in this category. All parts of the cue matrix had to be covered in order to be placed in one of the other categories.

Multiple rules. Whenever the subject used different rules for different cue combinations he was placed in this category.
Single rule. When the subject used only one prediction rule for all combinations he was placed in this category.

The two judges agreed upon the categorization in 83 of the 100 cases. In order to classify the remaining 17 subjects the two judges discussed until they reached consensus. The disagreements were not systematically related to the different categories.

Results

Learning stage. The results of the Learning stage are presented in detail in Armelius and Armelius (1976). In summary, the results were that all performance measures were influenced by the characteristics of the tasks. Specifically, the correlation between the subjects' judgments and the criterion values, $r_a$, was higher for high values of both $r_{ei}$ and $R_e$. Subject consistency, $R_s^2$, was higher for high values of $r_{ei}$ and the positive values of $r_{ij}$. The matching of regression weights, $G$, was higher for high values of $r_{ei}$. The results for $r_a$ in the Learning stage and the test block are presented in Fig. 1.

![Figure 1](image.png)

Figure 1. The effects of total task predictability, cue-criterion correlations and blocks on $r_a$ in the experimental and control conditions.
Frequency of combinations rules and task characteristics. It was possible to formulate combination rules for 53 out of the 100 subjects. 46 of these rules were classified as multiple and 7 as single rules. The verbal reports of the remaining 47 subjects were too incomplete or unsystematic to allow the formulation of prediction equations.

Since there were only 7 subjects in the single rule category the two types of rule were taken together in order to analyze the influence of task characteristics on the frequency of rules. The analysis of the frequency of rules was made as a 2 (levels of $R_e$) x 2 (levels of $r_{e_1}$) x 2 (sign of $r_{ij}$) nonmetric analysis of variance (Winer, 1962). There were no significant effects for any of the task characteristics. There were 23 rules in the high $R_e$ and 18 in the low $R_e$ conditions respectively. For $r_{e_1}$ the frequencies were 17 and 24 for high and low values respectively. For the sign of $r_{ij}$ finally, there were 16 for the positive sign and 25 for the negative sign. There were 5 rules in task 9 and 7 in group 10.

Goodness of fit of the combination rules. The goodness of fit of each combination rule was tested by means of a F-test based on the assumption that the judgments in blocks 5 and 8 are two independent samples of judgments drawn from the population of judgments generated by the same judgment process. Two different independent estimates of error variance were computed for each subject. The first estimate was computed as the average squared deviation of the judgments at block five from the predicted judgments at block five computed on the basis of the judgments at block 8. This is an index of the amount of error in the judgments since the cue and criterion values were the same in blocks 5 and 8. The predicted judgments were computed by formula (1).

$$J'_{5} = r_{58} \frac{s_{8}}{s_{8}} (J_{8} - \bar{J}_{8}) + \bar{J}_{5} \tag{1}$$
where

\[ J'_5 = \text{predicted judgment at block 5} \]
\[ r_{58} = \text{correlation between judgments at blocks 5 and 8} \]
\[ s_5 = \text{standard deviation of judgments at block 5} \]
\[ s_8 = \text{standard deviation of judgments at block 8} \]
\[ J_8 = \text{judgment at block 8} \]
\[ J_5 = \text{Average of judgments at block 5} \]
\[ J_8 = \text{Average of judgments at block 8} \]

The error variance of the judgments was computed by formula (2).

\[
\sigma^2_{eJ} = \frac{\Sigma (J_5 - J'_5)^2}{n - 2}
\]  

(2)

where

\[ \sigma^2_{eJ} = \text{error variance of judgments} \]
\[ n = \text{number of judgments at block 5} \]

The second estimate of error variance was computed as the average squared deviation of the judgments at block 5 from the judgments at the same block predicted by the combination rule given by the subject. The predicted judgments were computed by means of the prediction equations for each subject. The error variance of the combination rule was computed by formula (3).

\[
\sigma^2_{eC} = \frac{\Sigma (J_5 - J'_C)^2}{n - 2}
\]  

(3)

where

\[ \sigma^2_{eC} = \text{error variance of the combination rule} \]
\[ J'_C = \text{judgment at block 5 predicted by the combination rule} \]

The F-test was designed to determine to what extent the combination rule accounted for the systematic variance that existed in the subject's judgments at the end of learning. The expected value of this F-test is 1.0 under the following assumptions:

a) The distributions of deviations for the two error estimates are normal.
b) The combination rules account for the systematic variance in subjects judgments.

c) Subjects use the same combination rules in blocks 5 and 8. Any violation of these assumptions may cause a significant F-value.

The results were that the combination rules accounted for the systematic variance in 77% of the cases of 41 out of the 53 subjects. In the remaining 12 cases the F-test was significant (p < .01). These 12 subjects were not systematically related to the characteristics of the tasks or to the classification in types of rule.

Performance and type of rule. In order to test the hypothesis that subjects who formulate rules perform better than those who do not, the average performance indices were computed for each type of rule. The results are presented in Table 2.

Table 2. Average performance indices for each type of rule.

<table>
<thead>
<tr>
<th>Type of rule</th>
<th>No rule</th>
<th>Multiple rule</th>
<th>Single rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average $r_a$</td>
<td>.43</td>
<td>.57</td>
<td>.26 (p &lt; .05)</td>
</tr>
<tr>
<td>Average $R^2_s$</td>
<td>.40</td>
<td>.47</td>
<td>.23 (p &lt; .09)</td>
</tr>
<tr>
<td>Average $G$</td>
<td>.87</td>
<td>.97</td>
<td>.69 (p &lt; .01)</td>
</tr>
<tr>
<td>Average $r_a/R_e$</td>
<td>.45</td>
<td>.57</td>
<td>.18 (p &lt; .01)</td>
</tr>
</tbody>
</table>

As can be seen from Table 2 single rule subjects reach a lower level of performance than both the multiple rule and no rule subjects. The multiple rules are, however, systematically higher than the other categories in all performance indices. This is somewhat unexpected in view of the results of Armelius and Armelius (1975).
Discussion

The present study demonstrates that it is possible to use subjects' verbal descriptions to study the processes underlying the behavior in MCPL. Although a rather crude method was used to find the subjective rules, the rules turned out to account for the systematic variance in the subjects' responses in 77% of the cases. In its present form the method gives a combination rule for only about 50% of the subjects, but with some improvements in the technique it should be possible to increase this number. Perhaps a carefully conducted interview using think aloud techniques (cf. Kleinmuntz, 1968), will improve both the precision and the frequency of the combination rules given by subjects.

An interesting result of the present experiment is that the characteristics of the tasks do not influence the frequency of rules. Therefore it should be as common that subjects formulate rules in tasks with only random variance as in tasks with only systematic variance. (cf. Chapman & Chapman, 1967). This result does not mean that task characteristics are without importance for how subjects formulate their rules, however. On the contrary, task characteristics are important for the formulation of the details of the rules. The subjects in the groups with negative beta-weights, for example, generally formulated rules that expressed this aspect of the task. Similar results were obtained by Brehmer et al., (1974) in single cue probability learning. They point out that these results are consistent with a hypotheses sampling model, where the probabilities of various hypotheses are dependent only on the relative strengths of the hypotheses, and not on the input conditions of the experiment. The task only influences the decision to retain or reject a given hypothesis.

The results also show that combination rules are important for subjects' performance in MCPL tasks. The performance of subjects with multiple rules was higher than for other subjects. In the present context this means that the rules the subjects have developed have some validity. This seems, however, not to be the case with the rules used by single rule subjects. An inspection of the rules used by sub-
jects in the single rule category reveals that 6 out of the 7 used the average of the two cue values as a basis for their predictions. The seventh subject used the difference between the two cue values. Neither of these rules is very efficient for the present tasks. In addition, the consistency and reliability was very low for the subjects using single rules. If they had consistently followed the rules they report, their consistency and reliability should have been 1.00. The verbal reports of some single rule subjects indicate that they have tried different rules, but they have found no better rule than the one they started with. The explanation for the poor result of the single rule subjects, therefore, may be that they do not follow the rule they have reported, but try different rules throughout the experiment.

The approach to studying the processes underlying inference behavior used in the present study has the advantage of providing some ideas about how the rules are developed in MCPL. Based on impressions from the verbal reports and the results of the experiment two different strategies for the development of rules may be considered. According to both strategies subjects start with a relatively simple and non-specific rule such as a simple or weighted average of the two cue values, and use this as a basis for their predictions of the criterion values. In most tasks they soon find that this rule does not allow them to make perfect predictions. They then proceed with one of two different strategies. The first strategy, the multiple rule strategy, is to search for a limited set of cue-value combinations which enable them to predict the criterion value with a special prediction rule for each set. The number of subrules may be quite large and the extreme would be one prediction rule for each combination of cue values. This would be an analogue to paired associates learning in verbal learning. Eventually, the set of prediction rules will cover the complete cue-value matrix. The second strategy, the single rule strategy, is to try single rules, which cover the complete cue-value matrix. If subjects follow this strategy they have to evaluate their performance in probabilistic terms since most MCPL-tasks are probabilistic. In view of some recent experiments this seems to be an unrealistic requirement (Brehmer, et al., 1974, Armelius & Armelius, 1976 a).
The results of the present experiment clearly support the multiple rule strategy: 46 of the subjects developed multiple rules and only 7 who used single rules. An early experiment by Azuma and Cronbach (1966) also gives some evidence that subjects use several independent rules in probabilistic tasks. The development of rules in MCPL is presently studied in an experiment where subjects are asked to describe their rules at different points in time.

This study was supported by a grant from the Swedish Council for Social Science Research. The authors are indebted to Dr Berndt Brehmer for valuable comments on the paper.
References


Armelius, K., & Armelius, B-Å. Integration rules in a multiple cue probability learning task with a suppressor cue. Umeå Psychological Reports No. 80, 1975.

Armelius, B-Å., & Armelius, K. Confidence and performance in probabilistic inference tasks with intercorrelated cues. Umeå Psychological Reports No. 96, 1976. (a)

Armelius, K., & Armelius, B-Å. The effects of cue-criterion correlations, cue intercorrelations and the sign of the cue intercorrelation on performance in suppressor variable tasks. Organizational Behavior and Human Performance, 1976, (in press). (b)


