88

6.

UMEÅ PSYCHOLOGICAL REPORTS

No. 101 1976

Department of Psychology
University of Umeå

COMBINATION RULES IN MULTIPLE-CUE PROBABILITY LEARNING
II. PERFORMANCE, CONFIDENCE AND DEVELOPMENT OF RULES

Bengt-Åke Armelius

Kerstin Armelius

COMBINATION RULES IN MULTIPLE-CUE PROBABILITY LEARNING
II. PERFORMANCE, CONFIDENCE AND DEVELOPMENT OF RULES

Armeliuſ, B-Å., and Armeliuſ, K. Combination rules in multiple-cue probability learning. II. Performance, confidence and development of rules. Umeå Psychological Reports No. 101, 1976. - Subjects' use of combination rules was ſtudied in five two-cue multiple-cue probability learning tasks with varying degrees of predictability. Subjects were asked to ſtate how they made their predictions at different times during the experiment. 70 % of the subjects formulated systematic and consistent combination rules, while the remaining 30 % formulated rules that were incomplete or inconsistent. The verbal ſtatements were found to account for the subject's actual judgments in 86 % of the caſes. About 50 % of the rules were ſingle rules, with one rule covering the complete cue matrix and the reſt were multiple rules with different rules being uſed for different parts of the cue matrix. Performance and confidence were higher for subjects who had formulated systematic combination rules. The reſults of the experiment were analyzed in terms of a two-stage model for inference behavior. According to this model subjects ſample their firſt combination rule from a hierarchy of hypotheses about relations between cues and criterion. Frequent hypotheses in that hierarchy ſeem to be average, ſum and difference of the two cue values. In the ſecond stage subjects teſt their hypotheses or develop them through their experience with the task. In the preſent experiment there was ſome evidence that subjects uſing multiple rules conſtruct their rules on the baſis of their experience with the task, eſpecially in tasks with high predictability.

Recently, the interest in process studies of inference behavior has increased. Brehmer (1974) has developed a hypotheses sampling model for subjects use of rules in single cue probability learning (SCPL). Armelius and Armelius (1975, 1976a) have adopted basically the same approach to the study of combination rules in multiple-cue probability learning (MCPL).

According to the model presented by Brehmer (1974) learning to make inferences is a two-stage process. First, subjects have to find the correct rule, which is assumed to take place through sampling of hypotheses about the rule relating cue and criterion. The hypotheses have different sampling probabilities in order to account for the regularities in the order of sampling of different hypotheses. The sampling probabilities reflect the relative strengths of the hypotheses, and the strength of a hypothesis is dependent on earlier experience. In the second stage subjects learn to use the rule they have sampled. This is where the test of the correctness of the hypotheses occur.

The results so far in SCPL indicate that there exists a hierarchy of hypotheses about the functional relation between cue and criterion and that the hypotheses in this hierarchy have different sampling probabilities. The selection of hypotheses seems to be independent of the input characteristics of the task; the same hypotheses are tried regardless of the task (Brehmer, et al., 1974). This is consistent with the sampling conception of hypotheses selection. The results concerning the second stage indicate that subjects have difficulties with their tests of hypotheses, especially when the task is ambiguous due to low predictability or little diagnostic value of the task. Brehmer et al., (1974) found that subjects often reject the correct hypothesis on the basis of a single trial.

The hypotheses in SCPL refer to the functional relation between cue and criterion. In MCPL the hypotheses may also refer to the rule for combining the cue values to predict the criterion. The two stage model presented by Brehmer seems adequate to describe how subjects select

and test their hypotheses in MCPL as well as in SCPL. There are, however, no results that describe any hierarchy of possible hypotheses in MCPL. The results in MCPL so far indicate that a) the rules given by subjects at the end of learning account for a large amount of systematic variance in subjects' responses b) about 50 % of the subjects formulate systematic rules c) the rules are either single rules that cover the complete cue matrix or multiple rules that cover different parts of the cue matrix d) subjects who formulate combination rules reach a higher level of performance than other subjects e) the frequency of rules is not related to characteristics of the tasks.

In the study by Armelius and Armelius (1976a) a hypothesis about the development of rules was suggested. According to this hypothesis subjects start with a relatively unspecific rule, e.g., an average of the two cue values. When they find that this rule is incorrect they may follow one of two strategies; one alternative is to search for another single rule that covers the complete cue matrix. The other alternative is to develop a number of rules that each allow prediction of the criterion for a limited set of cue value combinations.

The present study is an attempt to study the development of rules in MCPL. The development of rules will be studied by asking subjects to state their rules at different times during the experiment. This may also give some ideas about the characteristics of the set of preexperimental hypotheses about combination rules that subjects bring to the experiment.

A second purpose of the experiment is to study the relation between confidence and the formulation of combination rules. In another report on this experiment Armelius and Armelius (1976b) found that subjects' confidence in the present tasks was related to how well subjects believed that they performed rather than to how well they actually performed. This result was the consequence of the fact that subjects knew very little about their actual performance. The hypothesis to be tested in the present study is that subjects' confidence is dependent on whether they have formulated a combination rule or not.

Method

Subjects. Fifty undergraduate students at the University of Umeå participated in the experiment to fulfill a course requirement. The subjects were randomly assigned to experimental treatments.

Experimental tasks and design. Five different two-cue MCPL-tasks were constructed. The tasks differed with respect to the intercorrelation (r_{ij}) between the two cues and total task predictability (R_e^2). The tasks were so constructed that, compared to an orthogonal task with the same cue-criterion correlations, the cue intercorrelation had one of three different effects on R_e^2 , (a) to increase R_e^2 , (b) to decrease R_e^2 and (c) to leave R_e^2 unchanged. There were two tasks where R_e^2 was higher when the cues were intercorrelated than when the cues were orthogonal. In one of these two tasks the increase in R_e^2 was accomplished by a positive cue intercorrelation and in the other by a negative cue intercorrelation (see Dudycha, Dudycha & Schmidt, 1975 for a description of the relation between r_{ij} and R_e^2). The cue-criterion correlations were the same in all conditions. Table 1 gives the task characteristics for each experimental task.

Table 1. Task characteristics for the five experimental tasks.

Experimental task	r_{e1}	r_{e2}	r_{ij}	R_e^2	β_{e1}	β_{e2}
1	.80	.40	-.23	1.00	.92	.58
2	.80	.40	.00	.80	.80	.40
3	.80	.40	.50	.64	.80	.00
4	.80	.40	.80	.80	1.33	-.67
5	.80	.40	.87	1.00	1.88	-1.25

Procedure. The cue- and criterion-values were presented in booklets. On the face of each page the two 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 150 training trials the subjects (a) observed the two cue values (b) gave their prediction of the criterion value on their answer sheets and (c) observed the correct criterion value. The subjects were allowed to work at their own pace. They were not informed of the structure of the tasks. They were told to base their predictions on the values of the two cues and it was emphasized that due to the nature of the task they should not expect to be correct on each trial.

Before the first training block a test block was presented to the subjects. They were shown the two cues and asked to make a prediction about the criterion value, but they were never shown the correct criterion value. On every fifth trial they were asked to state how they made their predictions and the confidence they had in their prediction. The confidence ratings were made on a scale ranging from -50 to +50, where -50 meant guessing and +50 meant that their prediction was within ± 2 units from the criterion.

These two questions were repeated on every fifth trial in blocks 3, 5 and 7. The cue and criterion values were identical in blocks 5 and 7 except for trial order. This was done to insure an estimate of the reliability of subjects' judgments. At the end of the experiment the subjects were asked to state as fully as possible how they had made their predictions during the experiment.

By means of the subjects' final description of their prediction rule and the earlier descriptions of how they had made their predictions it was possible to categorize the subjects into one of three categories according to what type of rule they were using. Each subject was classified as: a) not systematic b) single rule c) multiple rule by two independent judges (the authors). The authors agreed upon the classification in 86 % of the cases. The remaining 14 per cent were not related to any special category of rule or condition of the experiment. They were discussed until consensus was reached concerning their classification.

Results

Combination rules. 35 out of the 50 subjects formulated combination rules at the end of learning that could be translated into prediction equations. The remaining 15 rules were either inconsistent or too incomplete to allow prediction of the criterion. Only one subject never gave any description of how he made his predictions at any time during the experiment. The distribution of the rule categories over the different groups in the experiment is shown in Table 2.

Table 2. Frequency of different rule categories in the five experimental groups.

Rule category	Experimental group					Total	Chisquare
	1	2	3	4	5		
No rule	2	5	4	3	1	15	n.s.
Single rule	4	1	4	6	3	18	n.s.
Multiple rule	4	4	2	1	6	17	n.s.
Number of rules	8	5	6	7	9	35	

As can be seen from Table 2 there is a tendency towards higher frequency of rules in groups with high predictability (1 and 5) than in the remaining groups, although the difference is not significant.

Goodness of fit. The amount of systematic variance in subjects' judgments at the end of learning was estimated as the correlation between judgments in blocks 5 and 7 for each subject. This correlation was higher for subjects with rules than for subjects without rules. The average correlations were .55, .89 and .90 for no rule, single rule and multiple rule respectively.

In order to decide whether the combination rules were adequate descriptions of subjects behavior, a F-test was made for each subject. The F-test was designed to test whether the combination rules accounted for the systematic variance in the judgments. Two estimates of error variance were computed for each subject. The first estimate was based on the subject's judgments in blocks 5 and 7 and is an expression of deviations from systematic variance. The second estimate was an expression of the difference between the subject's actual judgments at block 5 and the judgments generated by his combination rule. The rationale of the F-test is described in detail in Armelius and Armelius (1976a).

The F-test was significant, for 5 subjects, (out of 35), 2 single rule and 3 multiple rule, i.e., for these subjects the combination rules did not account for the systematic variance in the responses. 4 of these subjects had very high reliability, that is, they responded consistently but did not express accurately what combination rule they used. One subject changed his combination rule from block 5 to block 7.

Development of rules. The frequency of different types of rule in block 1 may be seen as a first index of the sampling probabilities of hypotheses of combination rules in MCPL. In Table 3 the first rule given by each subject is shown. The rules are crosstabulated with the final classification in rule category in order to see whether there are any systematic relations.

Table 3. First rule given by each subject in the different rule categories.

Rule category \ First rule	No rule	Single rule	Multiple rule	Total
Average	1	11	4	16
Difference	5	2	3	10
Sum	3	4	3	10
Miscellaneous	3	1	6	10
Ratio	2	0	0	2
One cue	0	0	1	1
No rule	1	0	0	1
Sum	15	18	17	50

As can be seen from the table it is obvious that an average of the two cue values is the most frequent first rule. The difference or sum of the cue values are also very frequent rules. 50 % of the subjects who started with an average of the two cue values never changed their rules. Subjects who start with a miscellaneous rule have a relatively great chance of ending up in the multiple rule category.

Since the criterion value was never shown in block 1 subjects had to employ whatever rule was available and there was no chance to test the correctness of the rule. Thus, there should be no reason for subjects to try different rules in block 1. However, 37 subjects out of the 50 stated at least one rule during block 1. The average number of rules tried on the five possible occasions in blocks 1 and 7 respectively are shown in Table 4.

Table 4. Number of different rules given in blocks 1 and 7 for the different rule categories.

Rule category	Block 1	Block 7
No rule	1.60	1.47
Single rule	1.39	1.06
Multiple rule	1.71	2.06

There is no difference in the number of rules given in block 1 for the different rule categories. In block 7 the difference is significant ($p < .05$). As expected more rules are given by the multiple rule category than by the other categories in block 7.

Performance, confidence and rule categories. For each subject and block the squared multiple correlation between cues and judgments, R_s^2 , the correlation between the linearly predictable variance in the task system and that in the subject's system, G , and the correlation between each subject's judgments and the correct criterion values, r_a , were computed. The correlation measures r_a and G were transformed to

Fishers' Z-values before the statistical analysis. The average performance measures and confidence ratings at the last block of learning for the different rule categories are shown in Table 5.

Table 5. Average achievement, consistency, matching of regression weights, relative achievement and confidence at the last block of learning for the different rule categories.

	No rule	Single rule	Multiple rule
Achievement, r_a	.52	.72	.83
Consistency, R_s^2	.50	.76	.78
Matching, G	.95	.93	.97
Relative achievement			
r_a/R_e	.51	.71	.82
Confidence	-27.4	-7.7	3.2

Subjects using multiple rules reached a significantly higher level of achievement than subjects with no rules ($p < .01$). Consistency, relative achievement and confidence were higher for subjects with rules than for subjects with no rules ($p < .01$). The difference between multiple and single rules never reached significance although multiple rules are higher on all performance measures. The main reason for the higher performance for subjects with rules is that they are more consistent than other subjects.

Discussion

The results of the present experiment, in general are consistent with the results of previous studies on combination rules in MCPL. The rules given at the end of learning account for the systematic variance in subjects' judgments in 86 % of the cases who formulated a systematic combination rules. In the present study 70 % of the subjects formulated such rules. This is a somewhat larger percentage than in previous studies

(50 %), which may be due to the fact that subjects were asked to state their rule several times during the experiment. The results of the present study also show that performance is higher for subjects who have formulated systematic rules than for other subjects. This is true both for the single rule and multiple rule category. The increase in performance seems to be due mainly to a higher level of consistency among subjects who formulate systematic combination rules than for other subjects. In previous studies (Armeliu s & Armeliu s, 1975, 1976a) both consistency and matching of regression weights were higher for subjects with systematic rules.

The rules in the present experiment were classified either as single rule or multiple rule with acceptable interjudge reliability. The classification received support from the estimates of the number of rules given by subjects in the different categories at the first and last block of the experiment. The frequency of the two rule categories were about equal in the present experiment, while there were 87 % multiple rules in the experiment by Armeliu s and Armeliu s (1976a). The high percentage of multiple rules in that experiment may have been caused by the questions given to the subjects. A first question asked subjects to describe which cue value combinations they found especially easy to make predictions from, and how they made their predictions in those cases. A second question was concerned with difficult combinations of cue values. These are questions that favor multiple cues. The classification in single and multiple rules seems to be a good description of an important difference between the rules subjects use. The same difference was found in SCPL by Brehmer et al., (1974) who called them "functional" and "classification" rules respectively.

A new result in the present experiment is that subjects who had formulated systematic combination rules also had a higher level of confidence in their judgments. As shown in a previous report on this experiment (Armeliu s & Armeliu s 1976b), we found no relation between actual performance and confidence. There was also evidence that subjects know very little about their performance in these tasks. Together these results suggest that subjects' confidence is more strongly related to

how subjects make their predictions than to how accurate the predictions are. The reason for this may be that subjects cannot evaluate their performance in probabilistic tasks.

The results of the present experiment also have some implications for the two-stage model of inference behavior in MCPL. In block one no feedback was given which means that there was no opportunity for subjects to construct a rule on the basis of the feedback from the task. The fact that 86 % of the subjects formulated at least one combination rule under these circumstances is consistent with a sampling conception for the first stage of the model. The most common first rule given by the subjects is an average of the two cue values. The sum and difference of the cue values, are also frequent rules, while other mathematical expressions are relatively infrequent. An interesting feature of the first rule is that with one exception, it includes both cues.

While a sampling model seems adequate for the first of the two stages, a construction model may be adequate for the second stage where subjects test their rules and learn to utilize them. When feedback is given subjects have the possibility to change their rules on the basis of their experience with the task. According to the sampling conception subjects change their rules by rejecting the inadequate rule and sampling a new rule from their pool of preexperimental hypotheses. This procedure continues until they find a rule that is not rejected by the outcome of the test. In MCPL this may possible be the case for subjects who follow the single rule strategy. 10 out of the 18 subjects who were classified in the single rule category never changed their rule, however. These subjects therefore must have improved their performance by learning to apply the rule first selected for the task.

The tendency towards more multiple rules in the tasks with high predictability indicates that a construction model may be more adequate for describing the second stage for subjects in the multiple rule category. The feedback given to subjects in tasks with low predictability will be inconsistent with any deterministic rule since the criterion values contain random error variance. Therefore, the cues

in such tasks contain little or no diagnostic value and there is no information that subjects can use to verify, or modify their combination rule. In tasks with high predictability, on the other hand, subjects have a greater possibility to profit from the feedback given to them. Thus, they may learn the systematic relationship between cues and criterion for at least a part of the cue matrix. Through the multiple rule strategy they may successively develop rules for the complete cue matrix. In this way subjects can utilize anchoring effects and other heuristics known from studies of decision making (e.g., Slovic, 1972). The hypothesis advocated in the present paper therefore, is that the multiple rule strategy should be more frequent in tasks with high diagnostic value. Neither the present experiment nor the previous experiment (Armeliu & Armeliu, 1976a) contradict this hypothesis, but provide some support.

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 this paper.

References

- Armeliuſ, B-Å., & Armeliuſ, K. Integration rules in a multiple-cue probability learning task with intercorrelated cues. Umeå Psychological Reports No. 80, 1975.
- Armeliuſ, B-Å., & Armeliuſ, K. Combination rules in multiple-cue probability learning. I. The effect of task characteristics and performance. Umeå Psychological Reports No. 99, 1976(a).
- Armeliuſ, B-Å., & Armeliuſ, K. Confidence and performance in probabilistic inference tasks with intercorrelated cues. Umeå Psychological Reports No. 96, 1976(b).
- Brehmer, B. Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. Organizational Behavior and Human Performance, 1974, 11, 1-27.
- Brehmer, B., Kuylienſtierna, J., & Liljergren, J-E. Effects of function form and cue validity on the subjects' hypotheses in probabilistic inference tasks. Organizational Behavior and Human Performance, 1974, 11, 338-354.
- Dudycha, A., Dudycha, L., & Schmitt, N. Cue redundancy: Some overlooked relationships in MCPL. Organizational Behavior and Human Performance, 1974, 11, 222-234.
- Slovic, P. From Shakespeare to Simon: Speculations- and some evidence about man's ability to process information. Oregon Research Institute: ORI Research Monograph, 1972, 12, No. 12.