UMEÅ PSYCHOLOGICAL REPORTS

No 82 1975

Department of Psychology

University of Umeå

CONFIDENCE IN MULTIPLE-CUE JUDGMENTS AS A FUNCTION OF CUE
INTERCORRELATION AND TASK PREDICTABILITY

Kerstin Armelius

Bengt-Åke Armelius

CONFIDENCE IN MULTIPLE-CUE JUDGMENTS AS A FUNCTION OF CUE INTERCORRELATION AND TASK PREDICTABILITY

Armeliuș, K., & Armeliuș, B-Å. The hypothesis that the subjects' confidence is a direct function of the cue intercorrelation, r_{ij} , in a pure judgmental task, was tested in a two-cue MCPL experiment where the cue intercorrelation and total task predictability, R_e^2 , were inversely related. The hypothesis was supported. When the subjects received no feedback confidence was determined by r_{ij} . However, when the subjects received feedback the effects of r_{ij} on confidence was, as predicted, reduced in the direction of R_e^2 .

In their paper on the psychology of prediction, Kahneman and Tversky (1973) proposed that "the degree of confidence one has in prediction reflects the degree to which the selected outcome is more representative of the input than other outcomes are." (p. 249). According to Kahneman and Tversky, different outcomes may be considered as more or less representative of the evidence on which the judgment is based. This means that in multiple cue judgment tasks, internal variability or inconsistency among the cues will determine how representative the predicted outcome appears. The more consistent the cue variables are, the more representative the predicted outcome appears. For example, if grades in a subject is predicted from the scores on two tests, a prediction of an average grade will appear more representative if the prediction is based on an intermediate score on each of the two tests than if the same prediction is based on a high score on one test and a low score on the other test. In multiple cue judgment tasks, the cue consistency will be high when cues are intercorrelated and the correlation between the cue variables will therefore be a determinant of how representative the predicted criterion value appears. According to Kahneman and Tversky, the cue intercorrelation, r_{ij} , should therefore determine the confidence the subjects have in their judgments. The subjects' confidence is expected to be a direct function of the cue inter-

correlation. For a given set of cue validities, r_{ei} , total task predictability, R_e^2 , varies quadratically with the cue intercorrelation (Dudycha, Dudycha and Schmitt, 1974). In most tasks R_e^2 will be lower when the cues are positively correlated than when the cues are orthogonal. The intuition that correlated cue variables allow greater predictability than orthogonal cues, while the task predictability is in fact lowered by the cue intercorrelation, is called "the illusion of validity" by Kahneman and Tversky.

Kahneman and Tversky discuss only the case where r_{ij} is positive, but the hypothesis that confidence is a direct function of r_{ij} implies that the subjects will feel less confident in a task where r_{ij} is negative than in an orthogonal task. As shown by Dudycha et al. (1974) total task predictability will be higher when cues are negatively correlated than in an orthogonal task with the same cue validities. The illusion of validity should be called the illusion of invalidity if the cue intercorrelation is negative, provided that Kahneman and Tversky are correct in their proposal that confidence is a direct function of r_{ij} . The prediction is that the subjects will be less confident in a task with negative r_{ij} than in an orthogonal task, although R_e^2 is in fact higher.

Kahneman and Tversky made an experiment to demonstrate the effect of positive intercorrelation on the subjects' confidence. The subjects predicted grade point average from two pairs of aptitude tests. They were told that one pair of tests was highly correlated while the other was not. The result was that the subjects were more confident in predicting from the correlated tests. This indicates that the cue intercorrelation is an important determinant of the confidence the subjects have in their judgments, at least in a pure judgmental task, where the subjects have no knowledge of their performance.

In a learning task where the subjects are informed about the correct criterion value at each trial, it is reasonable to expect that the subjects' confidence is dependent on the subjects' performance. Armelius and Armelius (1975b) studied the subjects' confidence in an experiment where the cue criterion correlations, r_{ei} , the cue intercorrelation, r_{ij} ,

and the sign of the cue intercorrelation were systematically varied. Confidence was not influenced by any of the task parameters, although performance was influenced by all of them. The authors suggested the explanation that due to the difficulty of the tasks most subjects felt uncertain in all tasks. The rather small differences that existed in performance did not show up in differences in the subjects' confidence. In summary, the illusions of validity and invalidity have not been found in tasks where the subjects are informed about their performance. This is not so surprising since the subjects should be expected to learn something about the predictability of the task when they receive feedback. Therefore, the illusions should be expected to disappear or be reduced in learning tasks.

The purpose of the present study is to test the hypothesis that the illusions of validity and invalidity exist in judgmental tasks and that they are reduced when the subjects are informed about their performance. In order to test the hypothesis four two-cue MCPL-tasks with an inverse relation between R_e^2 and r_{ij} will be used.

Method

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

Experimental tasks and design. Four different experimental two-cue MCPL-tasks were constructed. The tasks differed with respect to the intercorrelation between the two cues and total task predictability. The tasks were constructed so that r_{ij} and R_e^2 were inversely related. Each task consisted of 25 trials.

Labelled cues and criterions were used, since the subjects' beliefs about the tasks rather than their learning was of primary interest in the present study. The tasks required the subjects to judge how suitable a fictitious pupil was for further studies to a given profession (the criterion variable) when his grade points in two different subjects were

known (the cue variables). The subjects were also required to state how confident they were that their judgment was correct at each trial. Another group of psychology students rated the correlation between grades in different subjects. The four pairs of subjects with the required intercorrelations were chosen for the experimental tasks (see Table 1). The profession was chosen to make it seem probable with a moderate relation ($r_{ei} = .45/.40$) between grades in the two subjects and suitability for further studies to that profession. Table 1 gives the statistical characteristics, the labels for the two cues and the criterion and the rated correlations between the cues for each experimental task.

Table 1. Statistical characteristics, labels of the cues and the criterion and rated cue intercorrelation for the four experimental tasks.

Experimental task	R_e^2	r_{e1}	r_{e2}	r_{ij}	Rated r_{ij}	Cue labels	Criterion labels
1	1.00	.45	.40	-.63	-.40	woodcraft-mathem.	dentist
2	.36	.45	.40	.00	-.07	mathem.-drawing	architect
3	.23	.45	.40	.63	.55	swedish-french	interpreter
4	.19	.45	.40	.90	.76	mathem.-physics	civil engineer

Half of the subjects received information about the correct judgment at each trial while the other half did not. The design was a 4 (Experimental tasks) x 2 (Feedback - No feedback) factorial with repeated measures on the first factor. Confidence was expected to increase with the cue intercorrelation when no feedback was given. The effect of feedback was supposed to decrease the effect of the cue intercorrelation on confidence, i.e. a significant interaction between experimental tasks and feedback was expected on confidence.

Procedure. The experimental tasks were presented in booklets. On the face of each page the two cues were presented as a number between .5 and 5.0.

The judgments of suitability for further studies were made as a number between 1 to 20. The confidence ratings were made on a percent scale, where 0 % meant that the subjects were guessing and 100 % meant that they were completely sure that their judgment was correct. In the feedback condition the subjects were told that they should not expect to be correct at each trial due to the nature of the task. Learning meant that their answers would be closer to the correct values, rather than perfectly correct at each trial. For the feedback condition the correct criterion value was presented at the back of each page. The subjects were instructed that for each task there was a moderate relationship between grades in the two subjects and suitability for further studies to the given profession. The subjects were allowed to work at their own pace. After the subjects had completed their prediction tasks, they were asked to rate the correlation between the grades in the four pairs of subjects.

Results

The rated correlations between the grades after completion of prediction were .14, .23, .47 and .53 for each experimental task respectively. These correlations are perfectly correlated with the desired ones at rank order level and most important the rated correlations are inversely related to R_e^2 .

Learning. The correlation between the subject's judgments and the correct criterion values, r_a , the squared multiple correlation between cues and judgments, R_g^2 and the correlation between the lineary predictable variance in the task system and that in the subject system, G , were computed for each of the experimental tasks for each subject. The correlation measure G was transformed to Fisher's Z-values before statistical analysis. The post hoc tests were made according to the Neuman-Keuls' method. All performance measures were subjected to a 4 (Experimental tasks) x 2 (Feedback - No-feedback) analysis of variance (ANOVA) with repeated measures on the first factor.

Significant main effects for experimental tasks ($F_{3/54} = 17.08, p < .01$) and for feedback - no-feedback ($F_{1/19} = 24.30, p < .01$) were obtained

on subject consistency. Consistency was higher when no feedback was given. Post hoc tests showed that consistency was significantly lower in the task with negative cue intercorrelation than in the other three tasks. No other effects were significant. Subject consistency in the different conditions and tasks is shown in Figure 1.

Figure 1. Average subject consistency as a function of cue intercorrelation in the feedback and no feedback conditions.

A significant main effect for experimental tasks was obtained on G ($F_{3/54} = 7.74$, $p < .01$), see Figure 2.

Figure 2. Average matching as a function of cue intercorrelation in the feedback and no feedback conditions.

In Figure 2 it can be seen that matching of the regression weights is lower in the task with negative cue intercorrelation and the task with the highest positive cue intercorrelation than in the other two tasks ($p < .01$). No other differences were significant.

Significant main effects for experimental tasks were obtained on achievement ($F_{3/54} = 18.18$, $p < .01$). Achievement was directly related to total task predictability. r_a was highest in task 1, ($r_{ij} = -.63$, $R_e^2 = 1.00$) and lowest in task 4 ($r_{ij} = .90$, $R_e^2 = .19$).

The post hoc tests showed that the difference in r_a was significant between task 1 ($r_{ij} = -.63$, $R_e^2 = 1.00$) and both task 3 ($r_{ij} = .63$, $R_e^2 = .23$) and task 4 ($r_{ij} = .90$, $R_e^2 = .19$). The difference between task 2 ($r_{ij} = .00$, $R_e^2 = .36$) and task 4 ($r_{ij} = .90$, $R_e^2 = .19$) was also significant. No other effects reached significance. The results for r_a are shown in Figure 3.

Figure 3. Average achievement as a function of cue intercorrelation in the feedback and no feedback conditions.

In summary, the main learning results are that achievement is positively related to task predictability and negatively related to the cue intercorrelation. For all tasks subjects consistency is higher when the subjects receive no feedback.

Confidence ratings. The average confidence for each experimental task was computed for each subject and subjected to a 4 (Experimental tasks) x 2 (Feedback - No feedback) ANOVA with repeated measures on the first factor.

Significant effects for experimental tasks ($F_{3/54} = 10.31, p < .01$) and a significant interaction between experimental tasks and feedback - no feedback ($F_{3/54} = 3.52, p < .05$) was obtained. The results are shown in Figure 4.

Figure 4. Average confidence as a function of cue intercorrelation in the feedback and no feedback conditions.

As can be seen in Figure 4, confidence is a direct function of the cue intercorrelation only in the no feedback condition. This conclusion was supported by the results of a trend analysis, which showed a significant linear trend in the no feedback condition, but not in the feedback condition. When feedback is given confidence is significantly lower in task 1 ($r_{ij} = -.63$) than in all three other tasks. These differences in the effect of r_{ij} on confidence, dependent on whether or not the subjects receive feedback, has resulted in the significant interaction.

Discussion

The present results clearly show that the illusions of validity and invalidity exist in a pure judgmental task and that the illusions are reduced when the subjects are informed about the criterion values. The proposal made by Kahneman and Tversky, that the subjects confidence is determined by r_{ij} seems to be true when no information is given to the subjects about their performance. In the present study the subjects' confidence was a direct function of r_{ij} only in the no feedback condition.

Feedback was expected to correct for the illusions of validity and invalidity and to make the subjects' confidence more dependent on R_e^2 . As expected the effect of feedback in the present study was to increase confidence in the task with negative r_{ij} and high R_e^2 and to decrease confidence in the tasks with positive r_{ij} and low R_e^2 . Feedback had no effect on the subjects' confidence in the orthogonal task. The illusion of invalidity was, however, not eliminated when the subjects received feedback. An explanation to the finding that the illusion of invalidity still exists in the feedback condition is that the 25 trials of feedback in the present study might not have been enough to correct for the illusion of invalidity. As shown by Armelius and Armelius (1975c) MCPL-tasks with high negative r_{ij} are very difficult to learn. Therefore, more trials might be needed to make the subjects learn the predictability of the task and to eliminate the illusion of invalidity.

Subject consistency and confidence were both higher when the subjects received no feedback in the present study. Consistency and confidence were directly related in the no feedback condition. These similarity of results may be related to how the subjects learn MCPL-tasks. As shown by Brehmer (1974) probabilistic inference tasks may be seen as a hypotheses testing activity. In MCPL-tasks the subjects try to find the correct rule relating cues and criterion (Armeliu & Armeliu, 1975a). If the subjects change rules frequently, consistency will be low. When the subjects are confident in their judgments there is no reason to test different hypotheses, which results in a positive relation between confidence and consistency. In addition, it follows that the hypothesis that two valid cues should be correlated to allow the greatest predictability, is one of the more dominant in the subjects' hierarchy of hypotheses about relations in MCPL-tasks. In other words, one of the first hypotheses that the subjects try in MCPL-tasks is what in the present study has been called the illusions of validity and invalidity.

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

References

- Armeliuș, K., & Armeliuș, B-Å. Integration rules in a multiple-cue probability learning task with intercorrelated cues. Umeå Psychological Reports, No. 80, 1975 (a).
- Armeliuș, K., & Armeliuș, B-Å. Note on the effects of cue validities, cue intercorrelation and the sign of the cue intercorrelation on confidence in multiple-cue probability learning. Umeå Psychological Reports, No. 83, 1975 (b).
- Armeliuș, K., & Armeliuș, B-Å. The effect of cue criterion correlations, cue intercorrelations and the sign of the cue intercorrelation on performance in suppressor variable tasks. Umeå Psychological Reports, No. 81, 1975 (c).
- Brehmer, B. Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. Organizational Behavior and Human Performance, 1974, 11, 1-27.
- Dudycha, A., Dudycha, L., & Schmitt, N. Cue redundancy: some overlooked relationships in MCPL. Organizational Behavior and Human Performance, 1974, 11, 222-234.
- Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.