F U N C T I O N A L   L E A R N I N G

L.F.W. de Klerk

J. de Leeuw

S. Oppe

may 1968

PSYCHOLOGICAL INSTITUTE / UNIVERSITY OF LEIDEN / THE NETHERLANDS

1. introduction

By functional learning is meant the learning of pairs of stimuli which vary quantitatively and which are related by means of a function.

In two recent papers on predictive behavior, Björkman (1966, 1967) has defined functional learning in terms of what he called the S-E scheme. This scheme refers to situations where stimuli (S) can function as cues for events (E). In Brunswik's lens model (1952), S would correspond to the cues and E would correspond to the distal object. In paired associate learning, S would correspond to the stimulus that is presented and E would correspond to the response that is to be learned. In an S-E relationship, S and E can "vary quantitatively, in discrete steps or continuously, and thus be represented by points on continua" (Björkman, 1966, p. 44). In situations like this, the assignment of E to S can be based on the condition that S and E differ within themselves, but E can also be a function of S. Corresponding learning is called functional learning.

Due to the associative character of the 'environment', subjects are supposed to learn the specific assignment of E to S. This assumption can be tested directly by inserting new stimuli. When the subjects have learned a function, then the responses to the interpolated (or extrapolated) stimuli are completely predictable.

Carroll (1963) has found evidence that subjects establish continuous relations between stimuli and responses, using a material in which S and E were lines, varying in length. He tested the assumption that the extent to which acquisition of a function takes place is dependent on the complexity of the function. Here, complexity of a function is defined by the number of its parameters. Carroll suggested that "a linear function is the best single function with which the organism can be equipped since other functions can be approximated by a linear function".

With the same type of material, De Klerk, De Leeuw & Oppe (1966) found, indeed, that the effectiveness of learning is different for different functions. A linear function with a positive slope turned out to be the easiest function to learn. This study confirmed the results which were obtained by Björkman (1965). He has used the same type of material to compare a linear function with a positive slope with a linear function with a negative slope. Analysis of the results revealed that the former type was more assessable and more effectively learned than the latter type.

Peterson, Ulehla & Lehman (1965) have studied functional learning with a material where S and E were numbers. They have found for 17 different S-E relations that subjects learn to use order embodied in S-E relationships to aid them in responding to new stimuli. Error in initial responses to new stimuli decreased as the number of S-E pairings increased.

From the above sketched picture of the research on functional learning it can be said that the individual subject has to learn the function E = f(S). Information about the degree to which the subject has learned this function, can be obtained by comparing the individual response (R) with E. Björkman, for instance, has measured the effectiveness of learning a linear function by taking the squared deviation of the reproduced line length (RLL) from the correct line length (CLL). A disadvantage of this measure is that, when errors are made by the subject, it gives no information about the function he has used to relate E to S. It can be said that errors may be due to (see also section 2):

(1) random scatter;

(2) psychophysical effects (as, f.i., end-effects);

(3) the fact that a wrong function has been used by the subject.

An alternative procedure could have been to measure cue-response correlations. The idea is that when the subject learns the specific S-E pairings, his response, R, should be in fair correspondence with E.

Cue-response correlations have been studied, for instance, by Summers (1962). He used tasks in which the stimuli contained three visual cues. Each cue varied over eight values, which were designated 1 through 8. Correlations were imposed between each of the cues and a variable CLL, whose magnitude varied with the magnitude of the cues. On each trial, the subject had to indicate how long he thought the line should be. The correlation between CLL and RLL was measured and compared with the validities of the cues. It was found that on the average subjects weighted the cues roughly proportional to their objective validities. Similar results were obtained by Uhl (1963).

Cue-response correlations are also used in situations where E is probabilistically related with S. Smedslund (1955), in the spirit of Brunswik's conception of ecological validity, used tasks where the single response dimension was probabilistically related with the numerical values of the stimulus-dimensions. The relation between the 'correct response' and the cues was obtained by taking the average of the numerical values of the cues, and adding a random 'correction score', denoted by e. Thus, here the subjects had to learn the function: $E = f(S) + e$.

The results, that were obtained by means of analyses of multiple cue-response correlations, clearly demonstrate the existence of multiple probability learning (in the sense of learning to utilize many probabilistic cues simultaneously).

In a more recent paper of Smedslund (1963), the results of two experiments on the concept of correlation in adult subjects were discussed. In these experiments the subject's frequency estimates and inferences of relationship were studied, relative to five different 2 x 2 distributions. It was concluded that adult subjects with no statistical training apparently have no adequate concept of correlation. The data on the subject's inferences reveal that normal adults "do not have a cognitive structure isomorphic with the concept of correlation" (op. cit., p. 172).

Azuma & Cronbach (1966) have carried out a multiple probability learning experiment in which the results of a multiple correlation technique were compared with verbal reports. They found that multiple regression analysis does not reflect and may even obscure the subject's spontaneous strategies in subdividing and analyzing the stimulus material. The analysis of cue-response correlations showed that the successful subjects progressed towards proper weighting of the cues, whereas the introspective reports led to the conclusion that subjects give rather good responses without necessarily attaining the rule used by the experimenter. According to Azuma & Cronbach, cues are used twice, first in a rule that dictates sub-class assignment, then in a second rule that dictates the response within the subclass.

In the next section, two alternative models for analysis will be discussed, by which it is possible to make accurate inferences about the function that is used by the subject to relate E to S.

2. models for analysis

Our discussion of models for analysis will be limited to multiple cue learning situations where E is related to a finite number of objectively quantifiable stimulus-dimensions. It is supposed that each stimulus is characterized by m dimensions, denoted by $X_1, \ldots X_m$. The stimuli are sampled from the cartesian product set $X = \prod_{i=1}^{m} X_i$.

The values of E are obtained by defining:

(1) a set of m real valued functions, $\emptyset_i$, which map the dimensions into the real line;

(2) a combination rule, f, which maps ordered m-tuples into the real line, i.e.,

$$E = \emptyset(x_i) = \emptyset(x_{i1}, \ldots, x_{im}) = f\left[\emptyset(x_{i1}), \emptyset(x_{i2}), \ldots, \emptyset(x_{im})\right]$$

The task of the subject is to learn the nature of the mappings $\emptyset_1, \ldots, \emptyset_m$, and the combination rule f. He is supposed to learn this because the feedback, given by the experimenter, gives information about the values $\emptyset(x_i)$.

As to the subject's responses, R, we may write:

$$R = \psi(x_i) = \psi(x_{i1}, \ldots, x_{im}) = g\left[\psi_1(x_{i1}), \psi_2(x_{i2}), \ldots, \psi_m(x_{im})\right]$$

In section 1 it was pointed out that the subject (S) can make three kinds of error. He may 'measure' one or more dimensions in the wrong way: $\psi_i \neq \emptyset_i$ for at least one i. Secondly, he may combine dimension values in the wrong way: $g \neq f$. And, in the third place, there may be random noise in the system. Because g should not include this random error, we may write for the response of the subject:

$$\psi(x_i) = g\left[\psi_1(x_{i1}), \psi_2(x_{i2}), \ldots, \psi_m(x_{im})\right] + e_i,$$

where e is a random variate.

In the experiments reported in this paper, stimuli are varied on essentially continuous dimensions, for each j, $\emptyset_j(x_{ij}) = w_j x_{ij}$, and the objective combination rule f is of the additive type: $\emptyset(x_i) = \sum_{j=1}^{m} w_j x_{ij}$. Here, x_{ij} is the value of stimulusi on dimension j. We shall assume throughout this paper that the subject also uses an additive combination rule to compute the subjective estimates of the $\emptyset(x_i)$ values.

Now, we shall discuss three essentially different models for analysis:

I. MULTIPLE LINEAR REGRESSION (MLR)

Here, the basic assumptions are:

I-1 $\psi(x_i) = \sum_{j=1}^{m} \psi_j(x_{ij}) + c$

I-2 $\psi(x_i) = k_i$, where k_i stands for the numerical response of the subject.

I-3 $\psi(x_{ij}) = w_j x_{ij}$.

From these assumptions it follows that there are real numbers $w_1, \ldots, w_m$, c, such that $k_i = \sum_{j=1}^{m} w_j x_{ij} + c$ for each $i = 1, \ldots, m$. In words, we may say that the responsesof the subject are a linear function (combination) of the dimension values.

II. NON METRIC MULTIPLE REGRESSION (NMR)

The basic assumptions are:

II-1 $\psi(x_i) = \sum_{j=1}^{m} \psi_j(x_{ij})$

II-2 $k_i \underset{o}{\geqslant} k_1 \Leftrightarrow \psi(x_i) \geqslant \psi(x_1)$

II-3 $\psi_j(x_{ij}) = w_j x_{ij}$

The binary relation $\underset{o}{\geqslant}$ in II-2 is an empirical partial order defined over the set of responses K. From the above mentioned assumptions it follows that there are real numbers $w_1, \ldots, w_m$ such that:

$$k_i \underset{o}{\geqslant} k_1 \Leftrightarrow \sum_{j=1}^{m} w_j x_{ij} \geqslant \sum_{j=1}^{m} w_j x_{1j} \Leftrightarrow \sum_{j=1}^{m} w_j (x_{ij} - x_{1j}) \geqslant 0$$

for each pair $(k_i, k_1) \in \underset{o}{\geqslant}$

III. ADDITIVE CONJOINT MEASUREMENT (ACM)

The two basic assumptions are:

III-1 $\psi(x_i) = \sum_{j=1}^{m} \psi_j(x_{ij})$

III-2 $k_i \underset{o}{\geqslant} k_l \Longleftrightarrow \psi(x_i) \geqslant \psi(x_l)$

From these assumptions it follows that there are real valued functions $\psi_1, \ldots, \psi_m$, defined on $X_1, \ldots, X_m$, respectively, in such a way that:

$$k_i \underset{o}{\geqslant} k_l \Longleftrightarrow \sum_{j=1}^{m} \psi_j(x_{ij}) \geqslant \sum_{j=1}^{m} \psi_j(x_{lj})$$

for each pair $(k_i, k_l) \varepsilon \underset{o}{\geqslant}$

The relationship between the three models can be explained in terms of the transformation they induce on the dependent variable K and the independent variables $X_i, \ldots, X_m$ (cf. table 1)

model	transformation of dependent variable	transformation of the independent variables
MLR	linear	linear
NMR	monotonic	linear
ACM	monotonic	any

Table 1. Types of transformation, induced by three models for analysis on the dependent and the independent variables.

From this table it appears that the ACM-model is much weaker than the other two, and that MLR has the strongest assumptions. The three models have in common the additive combination rule g.

3. experiment I

This experiment is an exploratory study. It has been carried out with the intention of investigating the extent to which subjects learn to combine dimension values into an estimate of the scale-values assigned by the experimenter.

3.1 stimuli

The stimuli used in this experiment were 10.5 x 14 cm cards. On the face of each 100 cards, a line of 10 cm length and a circle, the center of which was located on the line, were printed. The two dimensions were: position of the center of the circle on the line (this is called the P-dimension) and the size of the diameter of the circle (S-dimension). Each stimulus card was characterized by a specific value on each of these two quantifiable dimensions.

A major requirement in constructing the stimulus cards was that they should vary with respect to the numerical value of $\phi(x_i)$. Here, $\phi(x_i)$ was defined as the weighted addition of the numerical dimension values (with mm as the unit of measurement). Specifically, $\phi(x_i)$ was defined by:

$$\phi(x_i) = .5\, P_i + 2.5\, S_i$$

The 100 stimulus cards were constructed in such a way that $0 < \phi(x_i) < 100$. The P-values varied from 0 to 100, and the S-values from 3 to 20.

3.2 method and procedure

Prior to the experiment, the subjects were shown some of the stimulus cards. It was explained to them, that the stimuli varied with respect to a certain quantitative property $\phi(x)$, and that this measure is to be conceived of as a value on a scale, graded from 0 to 100. They were instructed to try to predict the numerical values of the various stimulus cards.

In the experiment proper, the stimulus cards were presented to the subject one at a time. After the presentation of each stimulus, the subject was required to give an estimate of the numerical value of $\phi(x_i)$. These subjective estimates or k_i-values (see section 2) were recorded on an answer sheet. After each response, the correct value of $\phi(x_i)$ was told.

The subjects were 12 volunteers from introductory psychology courses. None of them was familiar with this type of experiments.

3.3 results

The empirical data are analyzed per block of 20 consecutive trials. After the presentation of the last block, the 100 stimulus cards were presented again. The order of presentation remained the same for all subjects.

In order to investigate the extent to which subjects have learned the combination rule used by the experimenter, the deviations of k_i from $\emptyset(x_i)$ has been computed per block, according to:

$$D = \sqrt{\frac{\Sigma(\emptyset(x_i) - k_i)^2}{20}}$$

The results of this analysis are summarized in table 2.

subject	I	II	III	IV	V	VI	VII	VIII	IX	X
1	13.8	12.3	8.5	8.2	7.2	9.0	6.4	5.3	7.8	5.9
2	12.5	12.2	9.3	12.2	10.3	8.1	10.9	9.4	10.3	6.9
3	12.4	9.7	8.3	6.4	10.4	8.2	7.6	7.8	6.6	7.6
4	10.6	10.0	7.0	7.9	6,7	7.8	9.4	6.7	8.5	8.3
5	21.6	13.0	14.4	18.0	14.2	10.4	13.3	9.2	10.9	9.2
6	12.1	9.3	7.2	6.8	9.0	8.6	9.1	6.8	8.8	7.7
7	10.2	10.2	7.3	8.3	5.7	8.5	6.0	5.5	6.0	5.0
8	11.8	12.9	8.0	10.2	12.3	9.4	9.9	9.1	7.5	9.7
9	11.2	9.8	8.5	7.4	5.4	6.2	5.4	5.9	7.6	6.7
10	11.1	13.6	7.1	6.3	5.7	8.7	8.6	6.2	7.4	7.5
11	28.0	20.0	19.9	16.4	10.3	12.5	12.5	10.9	10.5	12.1
12	10.7	12.1	9.4	6.7	7.5	7.2	7.2	9.2	7.4	9.3
average	13.8	11.3	9.6	9.6	8.7	8.7	8.9	7.7	8.3	8.0

table 2. The individual D-scores, computed per block of 20 stimuli

A second analysis followed Smedslund (1955), Summers (1962) and Uhl (1963) in measuring the multiple correlation coefficient between the dimension values and k_i. The results of this MLR-analysis (see section 2) are given in table 3.

subject	I	II	III	IV	V	VI	VII	VIII	IX	X
1	.880	.983	.951	.965	.912	.935	.983	.968	.952	.951
2	.907	.954	.943	.920	.899	.940	.956	.918	.925	.919
3	.888	.943	.906	.971	.810	.938	.967	.929	.957	.924
4	.922	.921	.943	.951	.934	.934	.955	.946	.975	.885
5	.397	.926	.882	.756	.919	.945	.970	.956	.947	.924
6	.956	.977	.965	.962	.859	.932	.936	.949	.925	.941
7	.897	.961	.944	.973	.947	.939	.985	.972	.971	.957
8	.927	.910	.980	.967	.847	.961	.961	.984	.982	.891
9	.915	.974	.919	.963	.955	.958	.983	.956	.951	.926
10	.931	.939	.959	.956	.949	.934	.934	.949	.957	.849
11	.503	.821	.830	.845	.930	.909	.938	.913	.946	.851
12	.890	.912	.891	.964	.926	.960	.957	.904	.951	.908
average	.834	.935	.926	.933	.907	.944	.961	.945	.953	.911

table 3. The individual R^2-scores, computed per block of 20 stimuli

The results of the ACM-analysis (see section 2) are presented in the $\psi_j(x_{ij})$ versus x_{ij} plots (where j = 1,2), which are given in fig. 1. The dimension values, x_i, are grouped into classes with a class width of 10 for the P-dimension, and 5 for the S-dimension.

In order to investigate the extent to which the ACM-model fits to the data obtained from the two experimental runs, a goodness of fit measure μ has been computed (see appendix I). The μ-values are listed in table 4.

Subject	1st run	2nd run	subject	1st run	2nd run
1	.0363	.0125	7	.0200	.0094
2	.0382	.0111	8	.0272	.0067
3	.0422	.0223	9	.0044	.0137
4	.0392	.0079	10	.0312	.0166
5	.0534	.0168	11	.1031	.0470
6	.0107	.0077	12	.0251	.0226

The average μ-values are: for the 1st run: .0359; for the 2nd: .0162

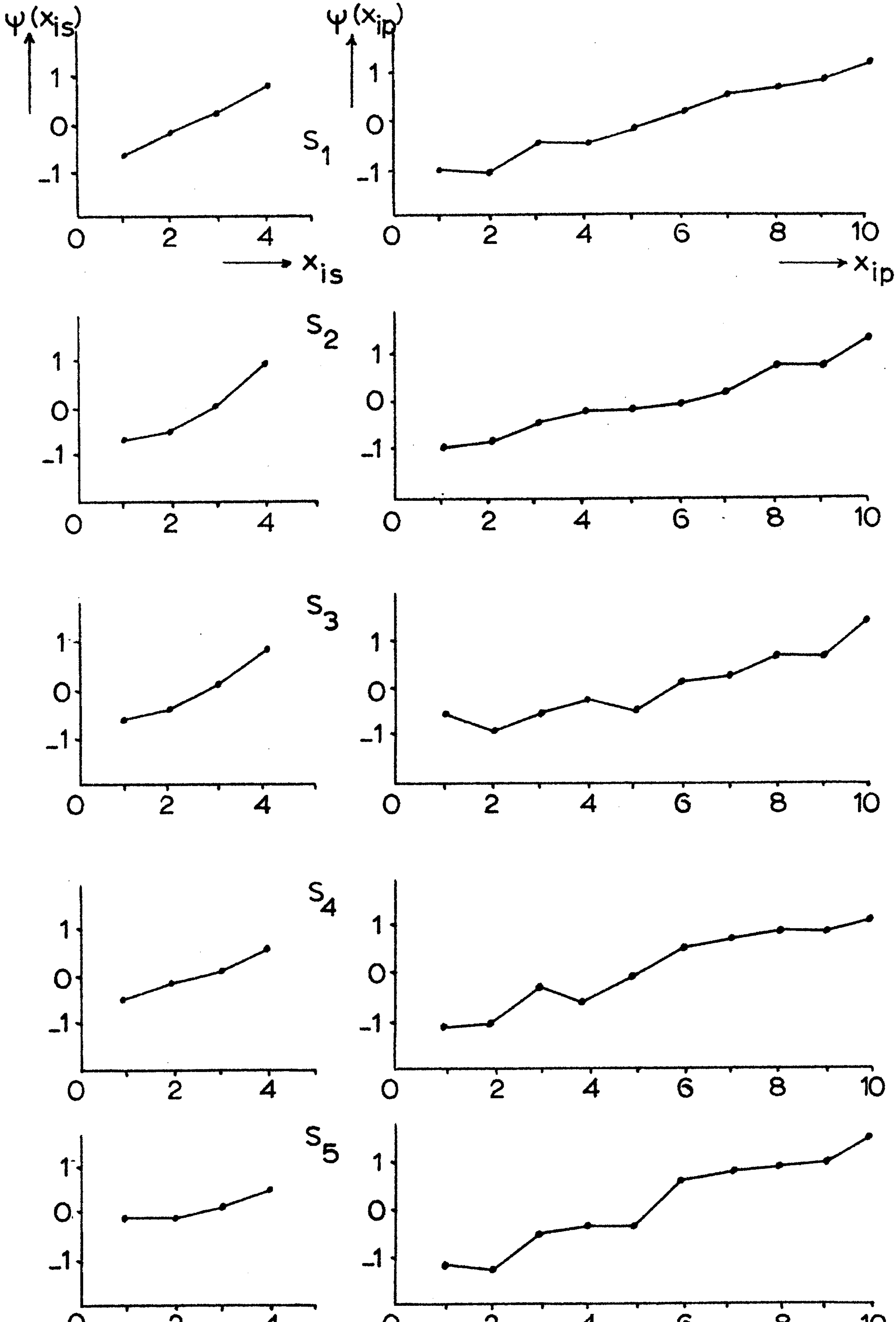
Table 4. The μ-values for each series of 100 stimuli and for each subject.

It should be noted that the μ-value, computed for the objective $\phi(x_i)$-values, is .0021. That it is not equal to zero is due to the grouping of the variates into classes.

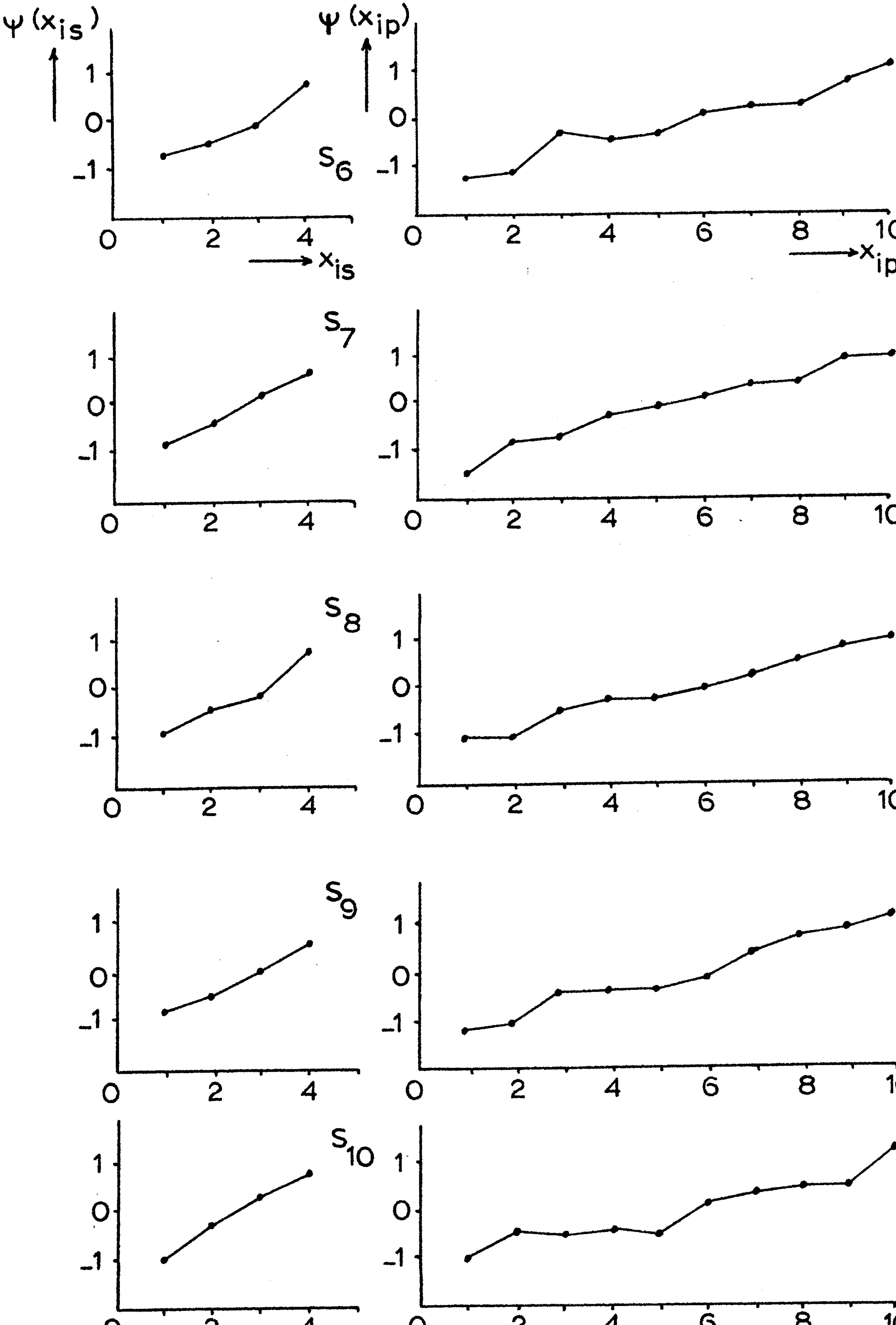
3.4 discussion

The data of table 2 reveal that the numerical values of D decrease when the number of training stimuli increases. This effect is most obvious for the first four blocks of stimuli. As to the other six blocks it can be said that - on the average - performance hardly improves. This indicates that within the ten blocks performance has become more and more asymptotic. The results also show substantial individual differences. The amount of learning as well as the rate of learning differ per subject.

Another point to be made concerns the way in which the stimuli are partitioned into blocks. It must be said that the blocks appeared to be different with respect to their degrees of difficulty. This is probably due to the fact that the range of $\phi(x_i)$-values was different for the ten blocks. For this reason it can be argued that the (completely arbitrary) way in which the stimuli are partitioned into blocks might have caused an undesired interaction between the degree of difficulty and the course of learning. At least, it can explain the fluctuations that are found during the last six blocks.



ψ(x_is)
ψ(x_ip)
x_is
x_ip
S1
S2
S3
S4
S5
1
0
-1
0
2
4
6
8
10



ψ(x_is)
ψ(x_ip)
x_is
x_ip
s6
s7
s8
s9
s10
1
0
-1
0
2
4
6
8
10

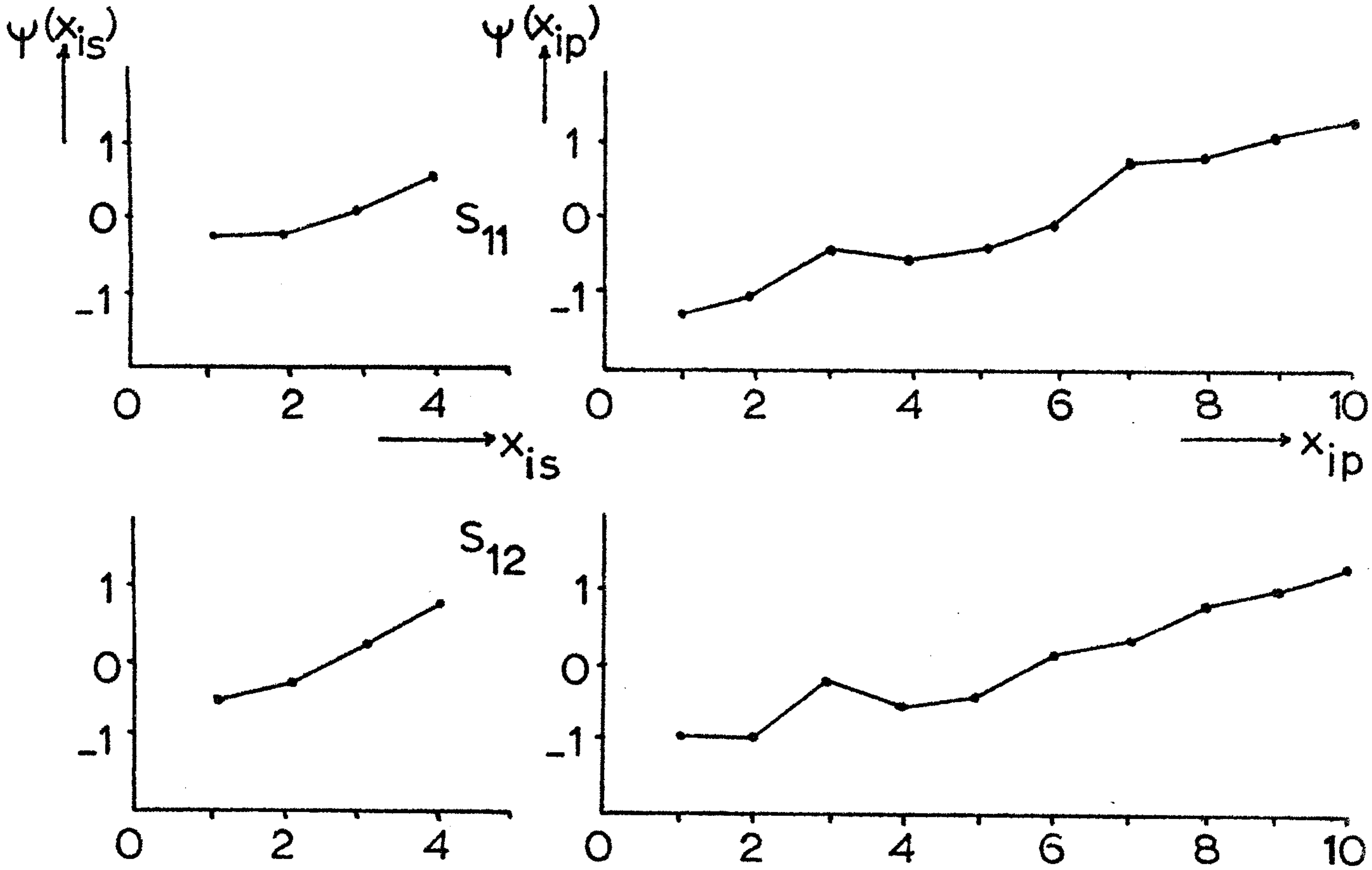


Fig 1. $\psi_j(x_{ij})$ versus x_{ij} plots, as derived from the ACM-analysis. The figures on the left refer to the S-dimension; the figures on the right refer to the P-dimension.

The results of table 3 reveal that the values of the multiple correlation coefficient between the dimension values and the subject's responses are relatively high, indicating that the MLR-model describes the data reasonably well. The data also show that after the second block, performance remains quite stable. Thus, compared with the analysis of D-values, it can be concluded that the MLR-model provides less information about the amount of learning. As said before, the D-values might deviate from zero because of three different reasons:

(a) random scatter;

(b) the type of the function (i.e., combination rule) used by the subject;

(c) the parameters of the function used by the subject.

The fact that the values of R deviate from unity, might be due to: a) random scatter, and b) non-linearity of the regression between the cues and the responses, but not to the parameters of the best fitting regression plane. Thus, when we assume that the subjects have used a linear function, it can be concluded that the learning effect, found in table 2, indicates that the subjects have learned primarily to estimate the parameters of the regression plane, rather than to improve the accuracy (i.e., consistency) with which a particular combination rule is applied to the dimension values.

From the data of table 4 it follows that in general the subjects perform better in the second run (with only one exception, i.e., S_9). Some subjects reach μ-values which are very close to the objective μ. Considering the relatively low values of μ (except for S_{11}, who is also a poor performer in terms of D and R), it can be said that the ACM-model describes the data reasonably well.

From the $\psi_j(x_{ij})$ versus x_{ij} plots, presented in fig. 1, it can be shown that the type of transformation of the dimension values which is necessary to make the system additive, is somewhat different for the two dimensions. As far as the transformation of the P-values is concerned, it can be said that there are no systematic deviations from a straight line. This is, however, not the case for all the transformations that have been carried out for the S-values. Fig. 1 reveals that the subjects S_1, S_3, S_5, S_6, S_8, S_9, S_{11}, and S_{12} show a tendency to non-linear weighting. These subjects have weighted the relatively high S-values heavier than the low S-values. This can be interpreted to mean that the subjects have paid more attention to the entire surface of the circles than to the size of their diameter

4. experiment II

This experiment is to be considered as a probablistic scalar concept learning task (De Klerk, 1968). Here, concepts are defined as probability distributions over the stimulus space. The stimuli are sampled randomly from two different concepts. The subject's task is to learn to discriminate optimally between the instances of the two concepts, called VEC and NO VEC, respectively.

4.1 stimuli

The stimuli used in this experiment were 10.5 x 14 cm white cards. On each card, an equilateral triangle (with sides of 8 cm length) was printed. The angular points of the triangles were the midpoints of three circles.

The length of the diameter of two of the three circles was the independent variable. The stimuli were constructed in such a way that there were two relevant dimensions and one constant dimension. As to the constant dimension it can be said that the length of the diameter was kept constant throughout the experiment; the values of the relevant dimensions (i.e., the length of the diameters, expressed in mm's) followed normal distributions with different means, but with identical standard deviations.

In this experiment a set of 100 stimuli has been used. One half of the entire set was sampled from the VEC population. The other 50 stimuli were sampled from the NO VEC population. The means of the two relevant value-distributions which defined the VEC stimuli, were 10.6 and 10.5. For the NO VEC stimuli, the means were 17.6 and 17.5, respectively. The length of the diameter of the constant circle was 12.0 mm. The standard deviation of the relevant value-distributions was about 3.

In short, the subjects (Ss) had to learn to discriminate between two a priori defined concepts. These concepts were defined in such a way that:

1. the two probability distributions are two-dimensional normal distributions;
2. the two variance-covariance matrices are identical scalar matrices;
3. the two vectors of means differ from each other with respect to both elements.

4.2 method and procedure

The stimuli were presented successively. As to each stimulus, S had to report the a posteriori probability that a given stimulus had been sampled from the VEC-population. The Ss were instructed that they had to express their 'subjective probabilities' by specifying a value on a 0 to 100 scale. It was explained to them that 100 means: "I am absolutely certain that this stimulus is a positive instance of the concept VEC", and that 0 means: "It is absolutely certain that this stimulus has been sampled from the NO VEC population".

After each probability statement the Ss were told from which of the two mutually exclusive populations the stimulus has been sampled.

The block of 100 stimuli was presented three times. The 300 subjective probabilities were registered on an answer sheet. Eight Ss participated. None of them was familiar with this type of experiment.

4.3 results

The three models, described in section 2, were applied to the data of this experiment. The first procedure was a multiple linear regression analysis with stimulus characteristics, $\phi(x_{ij})$, as input and with β-weights and squared multiple correlation coefficients, R^2, as output. The values of R^2 are listed in table 5.

block:	I	II	III		I	II	III
subect: 1	.053	.027	.001	5	.654	.769	.825
2	.529	.764	.747	6	.736	.838	.857
3	.717	.742	.775	7	.637	.770	.777
4	.845	.902	.936	8	.887	.843	.849
			average		.632	.707	.721

table 5. The values of R^2, computed per block and per subject.

The second procedure was a non metric regression analysis. From the results of this analysis, the Mahalanobis' Distance measure, D^2, can be derived (see appendix II). This measure gives information about the degree to which Ss have learned to discriminate between the instances of both concepts. The D^2-values are given in table 6.

The data of this table are compared with the R^2-values, given in table 5. The results are presented in figure 2.

Finally, the goodness of fit measure, α, (see appendix II) has been computed. This measure appeared to be 99.62, which means that:

block:	I	II	III		I	II	III
subject: 1	.001	.001	.000	5	.021	.026	.038
2	.017	.033	.058	6	.026	.036	.036
3	.026	.038	.041	7	.020	.022	.024
4	.027	.031	.031	8	.032	.029	.030
		average:			.021	.027	.032

table 6. The values of D^2 per block and per subject.

a) the model describes the data very well indeed;

b) there is very much consensus among the subjects with respect to the way in which they have discriminated between the instances of both concepts.

In table 7 are given the D^2-values, which are obtained by applying the ACM-model to the experimental data.

block:	I	II	III		I	II	III
subjects 1	.041	.025	.000	5	.408	.511	.708
2	.365	.744	1.181	6	.485	.637	.642
3	.509	.800	.812	7	.386	,370	.457
4	.477	.520	.467	8	.536	.445	.460
		average:			.401	.505	.591

Table 7. The values of D^2, inferred from the ACM-model, per block and per subject.

In figure 3, the D^2-values of table 7 are plotted against the R^2-values, obtained from the MLR-model. In this case, α appeared to be 96.15 which indicates that the ACM-model describes the data very well, though somewhat more variance was found as was to be expected from the NMR-analysis. Finally, the D^2-measures, inferred from the ACM-model are plotted against the D^2-values, inferred from the NMR-model (see fig 4).

4.4 discussion.

From the results of the tables 5, 6 and 7, it appears that on the average performance improves as a function of the amount of training.

As to the values of R^2, it can be said that they are highly significant, except for those of S_1 which are not significant at all. Evidently, the MLR-model did a very good job.

The dependent variable in MLR-analysis was not the subjective probability (SP) estimates of the subjects, but a transformation of these SPs, according to: $k_i = \ln SP_i/(1-SP_i)$. This transformation is

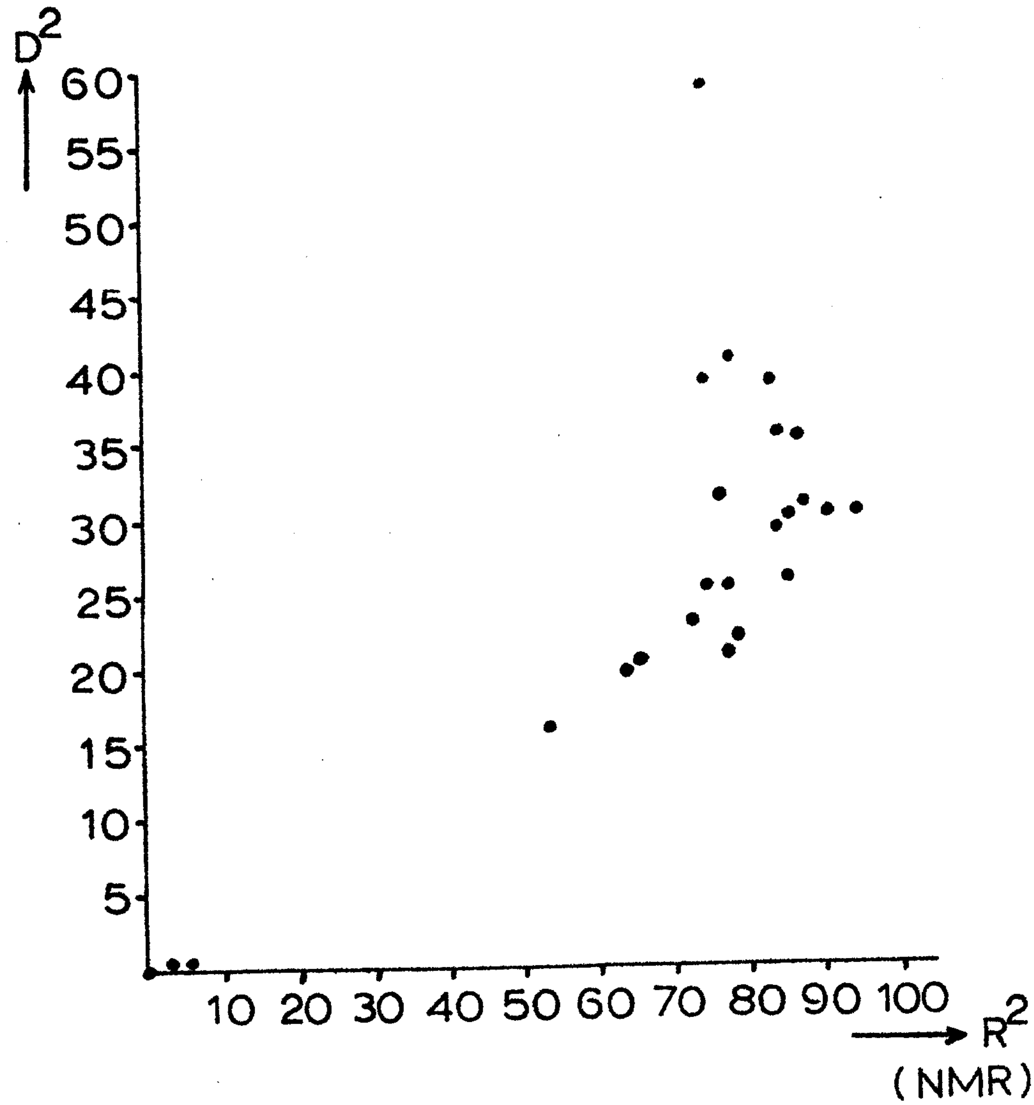


FIG.2

The values of the Mahalanobis' D^2, derived from the NMR-analysis versus the R^2-values.

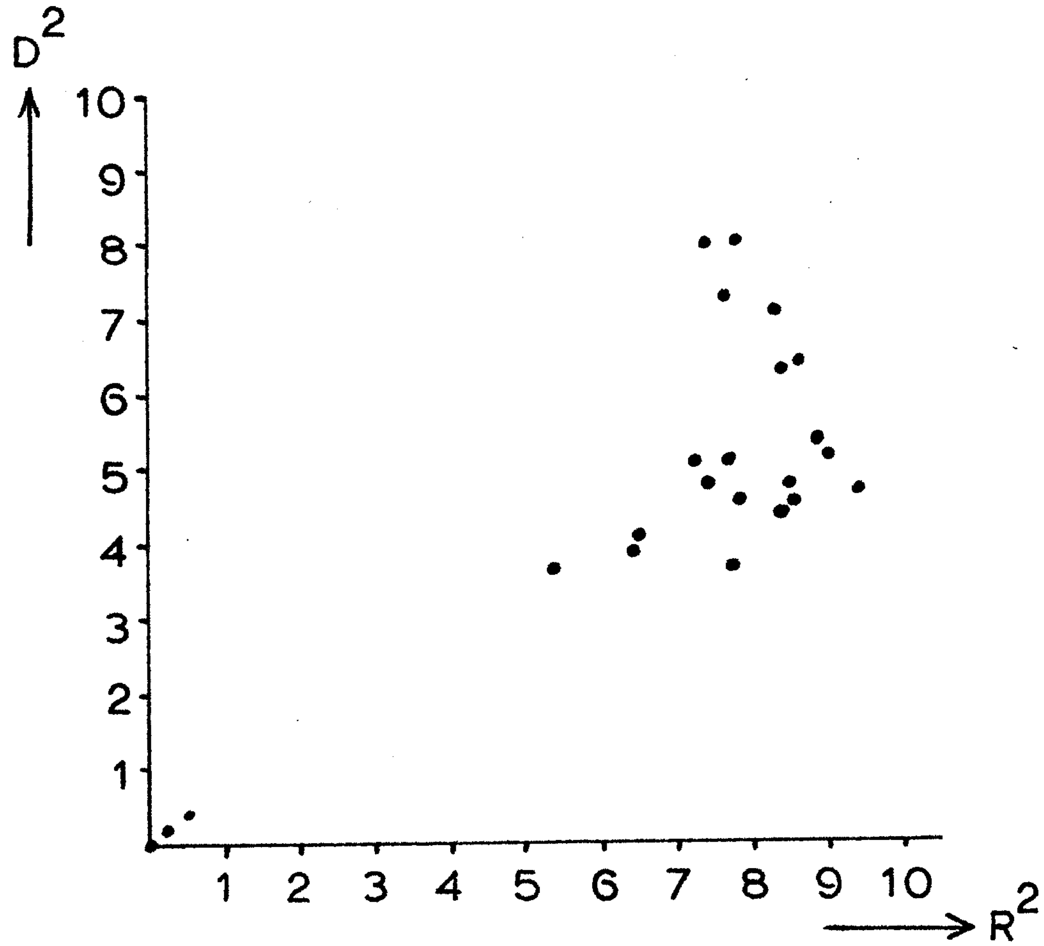


FIG.3

(ACM)

The values of the Mahalanobis' D^2, derived from the ACM-analysis, versus the R^2-values.

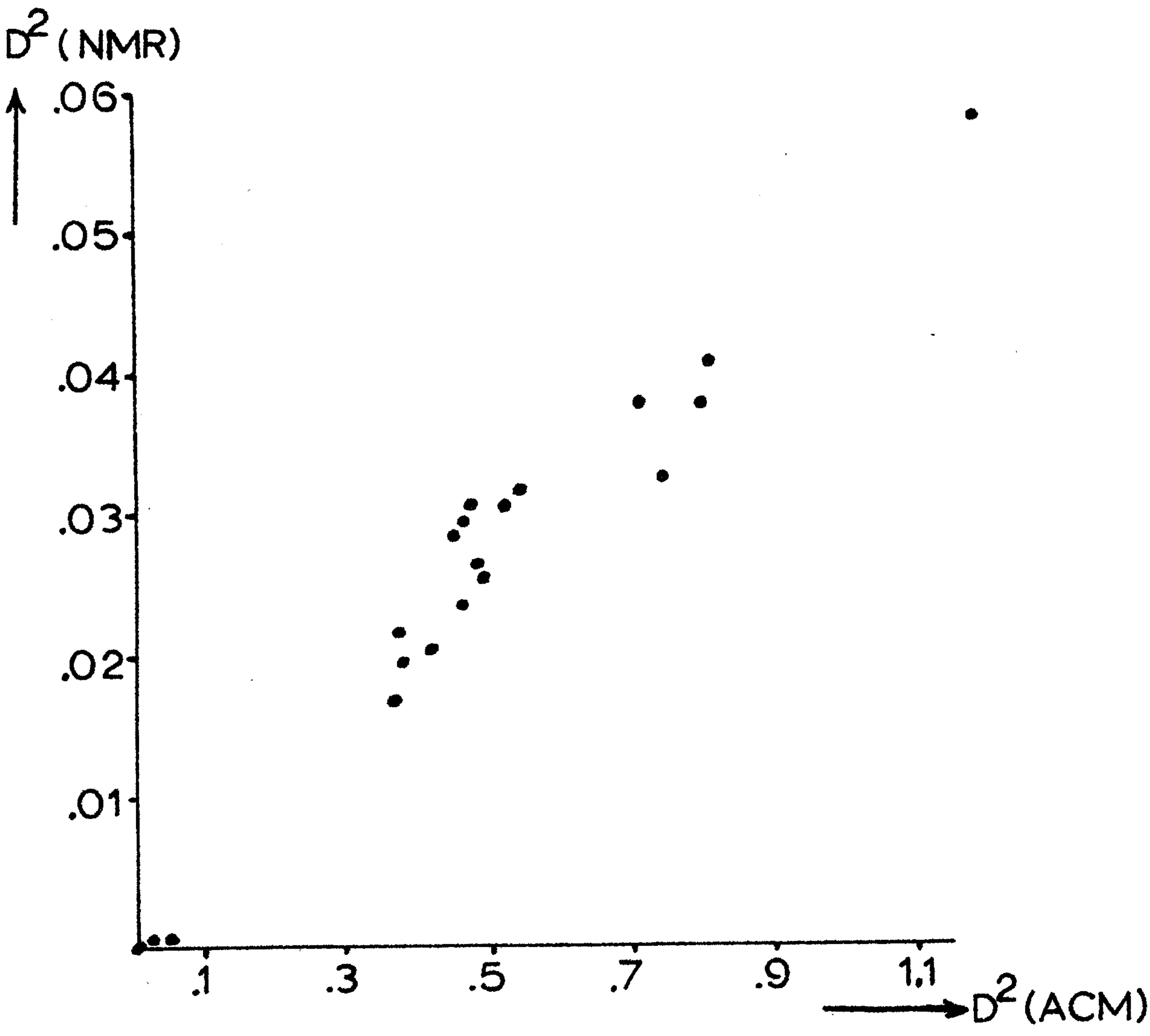


FIG. 4

In this figure, the Mahalanobis' D^2-values, which are derived by applying the NMR-model to the data, are plotted against the D^2-values, obtained from the ACM-model.

suggested by applying Bayes' theorem to situations like this (De Klerk 1968). The optimal strategy defined by Bayes' theorem prescribes an additive combination of linearily transformed stimulus characteristics. In other words, an MLR-model, provided that the a posteriori probabilities are transformed as described above. As a matter of fact, this is the principal theoretical justification for applying additive models to these data.

The Mahalanobis' D^2-measures from table 6 can be transformed into F-ratios and tested for significance (Porebski, 1966). The conclusion is the same as for the R^2-values: all F-ratios are very significant, except for S_1. A similar conclusion is valid for the D^2-values of table 7.

From figures 2, 3, and 4, it is apparent that when the Ss show good performance according to one of the three models, they have a good performance on the others as well. The relationship between the two different D^2-measures and the R^2-measure is complicated by the fact that the latter cannot exceed unity. This explains the increasing variation in D^2 for the higher values of R^2. That the relationship between both measures is not monotonic, can be explained by the fact that NMR and ACM models induce a monotonic transformation of the independent variable K (see page 5). This means that the group of Ss that deviates clearly from the linear relationship between D^2 and R^2 performs rather satisfactory, but only after a non-linear transformation of their $\ln \frac{SP_i}{1-SP_i}$ values.

In the ACM-analysis, optimal transformations of the stimulus-dimensions are constructed. These optimal transformations are given in figure 5. In the middle regions of the scales, where almost all stimuli are located, the relationship is linear, indicating that Ss transform the dimensions in a Bayesian way. At the ends of the scales the transformation deviates considerably from linearity, which may be due to:

a) the response scale being bounded by the anchors 0 and 100;

b) the relative scarceness of stimuli at the end-points;

c) the fact that Ss tend to be more conservative in extremely unlikely situations, as has been found by Vlek & Van der Heyden (1967)

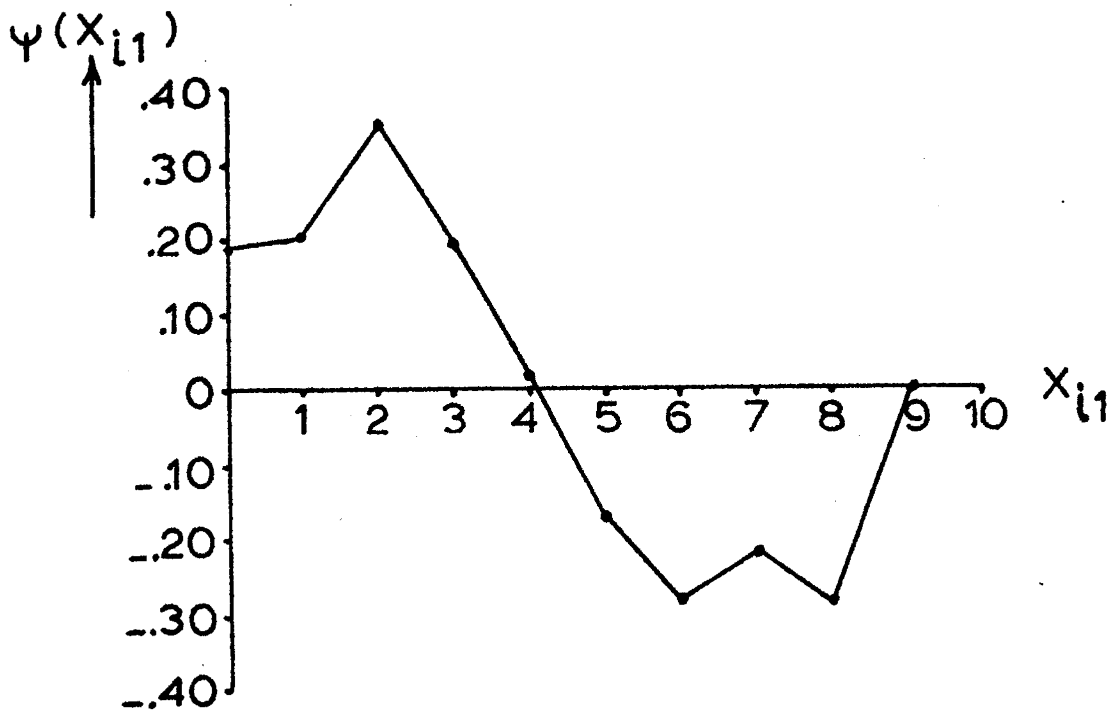


FIG.5[a]

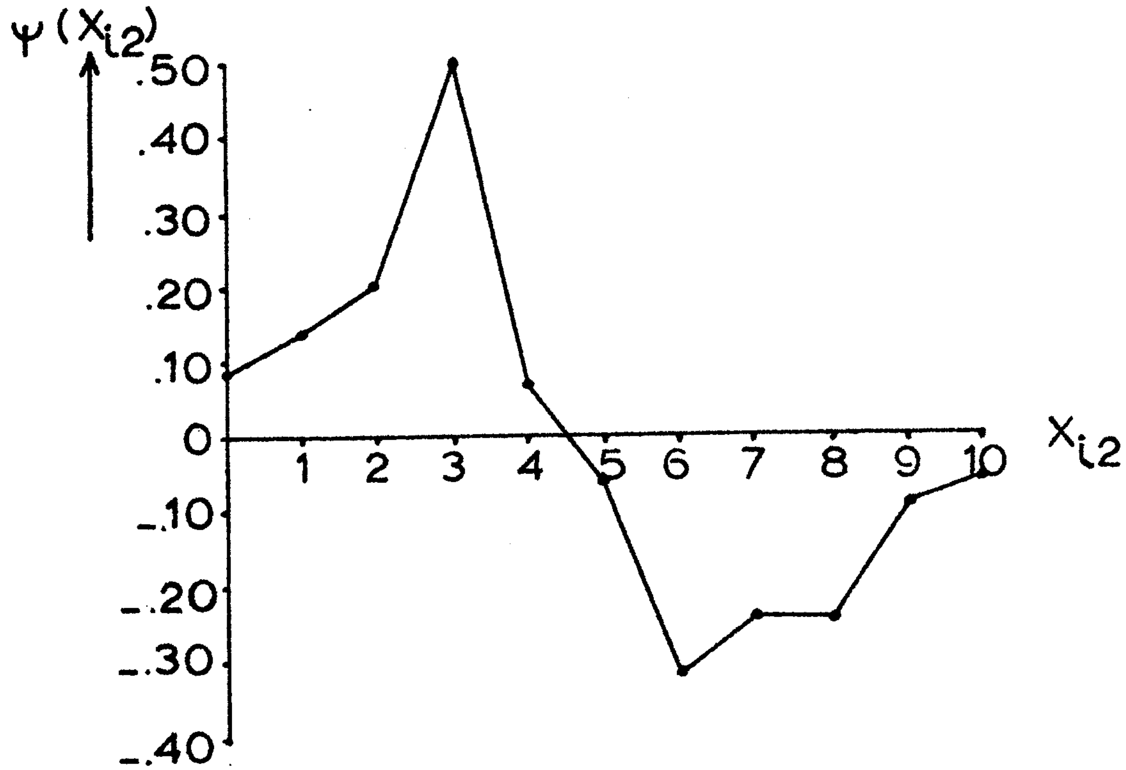


FIG.5[b]

The values of $\psi(x_{ij})$, as obtained from the ACM-model, versus the values of x_{ij}, for both relevant dimensions, respectively.

APPENDIX I

The ACM-soultions discussed in section 3.3 were found by a technique originally due to Kruskal (1964a and b). The coefficient μ is defined as:

$$\mu = \frac{\Sigma(\psi(x_i) - \widehat{\psi(x_i)})^2}{\Sigma(\psi(x_i))^2}$$

where summation extends over all x in the data structure. We have pointed out that in the ACM-model:

$$\psi(x_i) = \Sigma\, \phi_j(x_{ij})$$

The $\widehat{\psi(x_i)}$ are defined as those real numbers that maximally resemble $\psi(x_i)$ in the least square sense, under the condition that they satisfy the ordinal restraints in the data structure. By the constant gradient method we try to find numerical values for $\phi_j(x_{ij})$ in such a way that μ is minimized. The $\psi(x_i)$-values are scaled after iteration in such a way that μ can be interpreted as a ratio of variances: the residual variance divided by the total variance. Evidently, this algorithm tries to find a 'best' solution to an inconsistent system of linear inequalities. A complete description can be found in De Leeuw (1968a).

APPENDIX II

The NMR and ACM solutions discussed in section 4.3 are found with programs in the CDARD-series (De Leeuw, 1968b). A generalized correlation coefficient, obtained by Kendall-weighting of differences, is defines as: $d = \sum_{ij}\sum \text{sign}(k_i - k_j)(\psi(x_i) - \psi(x_j))$

Such a coefficient can be computed for each of the p subjects, and in CDARD-approach we choose the $\phi_j(x_{ij})$ values in such a way that $\rho = \Sigma d_l^2$ is maximized. The coefficients d are signed distances between two distributions of differences. If appropriately scaled and if the usual distributional assumptions are made, they can be transformed into Mahalanobis' Distance measures. In NMR, we have:

$$d_l = \sum_i \sum_j \text{sign}(k_{il} - k_{jl})\left(\sum_q w_q x_{iq} - \sum_q w_q x_{jq} \right) =$$

$$= \sum_i \sum_j \text{sign}(k_{il} - k_{jl})\left(\sum_q (x_{iq} - x_{jq}) w_q \right)$$

For ACM, the relevant quantity is:

$$d_l = \sum_i \sum_j \text{sign}(k_{il} - k_{jl})\left(\Sigma\, \phi(x_{iq}) - \Sigma\, \phi(x_{jq}) \right)$$

and again: $\rho = \Sigma\, d_l^2$ is maximized.

In De Leeuw (1968b) these problems are translated into matrix algebra, and the solutions are shown to be the first eigenvector of a particular symmetric Gramian matrix. The percentage of the variance explained by the first eigenvector will be called α.

S U M M A R Y

Functional learning refers to situations where the subject has to learn the function that relates E to S.

According to Smedslund, Summers, and Uhl, the performance of the individual subject can be measured by the method of cue-response correlations. In this paper, three alternative models for analysis are discussed, that make it possible to make accurate inferences about the function, used by the subject. These models are:

(1) multiple linear regression model (MLR)

(2) non-metric multiple regression model (NMR)

(3) additive conjoint measurement model (ACM)

In experiment I, E was a linear function of two variables: P (i.e., the position of the midpoint of a circle on a line), and S (i.e., the size of the diameter of the circle).

Individual D-scores are compared with MLR-data. It was concluded that the latter provide less information about the amount of learning. Ss have learned primarily to estimate the parameters the regression plane, rather than to improve the accuracy with which a particular combination rule is applied to the dimension values.

In experiment II - a probabilistic scalar concept learning task - E is a function of the probability with which the stimuli are sampled from one of two mutually exclusive populations. The three above-mentioned models are applied to the data of this experiment. The transformation of the dimension-values, induced by ACM, shows a deviation from linearity at the end-points of the scale. Some factors are mentioned which might have caused these deviations

R E F E R E N C E S

Azuma, H., and Cronbach, L.J., (1966) Cue-response correlations in the attainment of a scalar concept. Am. J. Psychol., 79, 38-49.

Björkman, M., (1965) Learning of linear functions: a comparison between a positive and a negative slope. Rep. Psychol. Lab., Univer. Stockholm, No 183.

Björkman, M., (1966) Predictive behavior; some aspects based on an ecological orientation. Scand. J. Psychol., Vol7, 43-57

Björkman, M., (1967) Stimulus- Event learning and Event learning as concurrent processes. Organizational behavior and human performance, 2, 3, 219-236.

Carroll, J.D., (1963) Functional learning: the learning of continuous functional mappings relating stimulus and response continua. Educational Testing Service. Research Bulletin (RB-63-26), Princeton.

De Klerk, L.F.W., De Leeuw, J., and Oppe, S., (1966). Funktioneel leren. Hypothese, 11, 1, 10-19.

De Klerk, L.F.W., (1968) Probabilistic Concept Learning: a study of choice behavior with multinormally distributed stimuli. Voorschoten: VAM (Netherl.)

De Leeuw, J., (1968a) Algorithms for the numerical analysis of ordinal data structures (forth-coming)

De Leeuw, J., (1968b) Canonical discriminant analysis of relational data. Research Note RN 007-68. Dept. of Data Theory for the Social Sciences, Univer. Leiden.

Kruskal, J.B., (1964a) Multidimensional Scaling by optimizing goodness of fit to nonmetric hypothesis. Psychometrika, Vol 29, 1-27.

Kruskal, J.B., (1964b) Nonmetric multidimensional scaling: a numerical method. Psychometrika, Vol 29, 115-129.

Peterson, C.R., Ulehla, Z.J., and Lehman, R.S., (1965) Function order and paired associate learning. J. Exp. Psychol., 69, 2, 119-120

Porebski, O.R., (1966) On the interrelated nature of the multivariate statistics used in discriminatory analysis. Brit. J. Math. Stat. Psychol., 19, 2, 197-214.

Smedslund, J., (1955) Multiple-probability learning. Oslo: Akademisk Forlag

Smedslund, J., (1963) The concept of correlation in adults. Scand. J. Psychol., 4, 165-173.

Summers, S.A., (1962) The learning of responses to multiple weighted cues. J. Exp. Psychol., 64, 29-34.

Uhl, C.N., (1963) Learning of interval concepts: I. Effects of differences in stimulus weights. J. Exp. Psychol., 66, 264-273

Vlek, C.A.J., and Van der Heyden, L.H.C., (1967) Subjective likelihood functions and variations in the accuracy of probabilistic information processing. Rep. E 017 67, Psychol. Inst. Univer. Leiden.